
Triptech: A Method for Evaluating Early Design Concepts

Julie Anne Séguin
Google Inc.
Mountain View, CA, USA
jaseguin@google.com

Alec Scharff
Google Inc.
Mountain View, CA, USA
scharff@google.com

Kyle Pedersen
Google Inc.
Mountain View, CA, USA
pedersenk@google.com

ABSTRACT

Measuring user experience (UX) is an important part of the design process, yet there are few methods to evaluate UX in the early phases of product development. We introduce Triptech, a method used to quickly explore novel product ideas. We present how it was used to gauge the frequency and importance of user needs, to assess the desirability and perceived usefulness of design concepts, and to draft UX requirements for *Now Playing*—an on-device music recognition system for the Pixel 2. We discuss the merits and limitations of the Triptech method and its applicability to tech-driven innovation practices.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

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

© 2019 Copyright is held by the author/owner(s).

ACM ISBN 978-1-4503-5971-9/19/05.

DOI: <https://doi.org/10.1145/3290607.3299061>

Tips for formulating user need statements

1. Integrate user research. Review existing user or consumer research to formulate hypotheses about needs or pain points.
2. Write short, tech-agnostic statements. Omit any mention of the technology or concept. Include an example if necessary but keep the length to a sentence or two.
3. Use the first person. Ensure respondents are rating frequency and importance from their perspective.
4. Omit elements of time. Words such as “always”, “often”, “sometimes”, “rarely” or any other reference to prevalence make it confusing for respondents to answer the frequency question and for the practitioner trying to interpret the results.
5. Stick to a single need or problem. Including more than one problem per statement makes it impossible to know which issue is driving frequency and importance ratings.

Bad example

People frequently underestimate how long it takes to get to a destination and arrive late. (See tips 3, 4, & 5)

Good example

I arrive late because I underestimate how long it will take to get to my destination.

CCS CONCEPTS

• Human-centered computing~User studies

KEYWORDS

User experience research; user experience evaluation; concept testing; surveys; interviews; focus groups; user-centered design

ACM Reference format:

Julie Anne Séguin, Alec Scharff, and Kyle Pedersen. 2019. Triptech: A Method for Evaluating Early Design Concepts. 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/10.1145/3290607.3299061>

1 INTRODUCTION

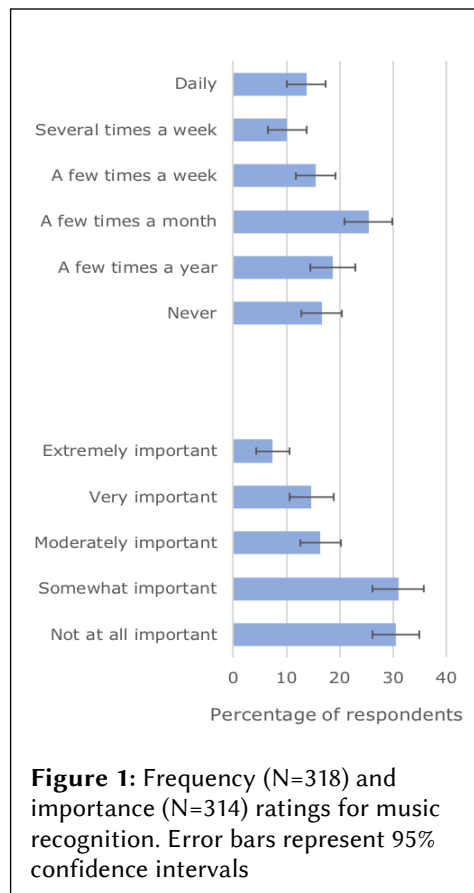
Measuring user experience (UX) is a proven way to test assumptions about what users need and expect from novel products. While there is demonstrated value in integrating user feedback in all stages of the product development cycle [2] there is a need for more evaluation methods that can be used in early phases of product development [10]. Another challenge to UX evaluation comes from the rise in technology-driven innovation, whereby new products are inspired by technological advancement rather than by validated user needs. User-centered design evangelist Don Norman has argued that because conceptual product breakthroughs are driven by the development of new technologies, consumer research should only take place after a product direction has been established [6]. Although this point was controversial, it highlighted the difficulty of incorporating user feedback when working in tech-driven environments. To this effect, we introduce Triptech, a UX research method compatible with the conceptual phase of the design process as well as with tech-first innovation practices. We review how it was used to evaluate user needs and to collect feedback on early design concepts for machine learning (ML) applications. We present the findings used to draft UX requirements and to outline future design and research activities for *Now Playing*, a music recognition system for the Pixel 2 [1].

2 METHOD

Part 1: Surveys

Materials

The first step of a Triptech study is to assess the relative frequency and perceived importance of user needs independently from how we plan to address them. This part serves two purposes. First, it forces the team to produce a clear articulation of the need or problem in a concise, tech-agnostic way. Second, the surveys yield generalizable results used as a first-pass filter to prioritize user needs and related concepts. This can be particularly useful when there are more problems or needs



than what can be tested with in-person study participants (Part 2). We began by formulating a need or problem statement for each of the design concepts. We outlined nine user needs representing nine ML concepts. The need for the music recognition concept that became *Now Playing* and the focus of this case study was: “*When I hear music I like, I want to know the artist and title (e.g. a song playing in a coffee shop)*”. See side bar for tips for formulating need statements.

Participants

3,003 US respondents (307 to 479 unique participants per survey) completed online surveys in exchange for access to premium content on a network of sites or for Google Play credit (www.google.com/surveys).

Procedure

Respondents completed one of nine three-question surveys: one for the music recognition need and eight for other ML smartphone concepts (between-subjects design). The first question was used to screen out respondents who did not have a smartphone. The other two questions asked respondents to rate the frequency and importance of the need, shown here using music recognition as an example:

1. How often do you encounter the following problem: When I hear music I like, I want to know the artist and title (e.g. a song playing in a coffee shop)? 6-point scale [Daily; Very often (several times a week); Often (a few times a week); Sometimes (a few times a month); Rarely (a few times a year); Never].
2. How important is it to address the following problem: When I hear music I like, I want to know the artist and title (e.g. a song playing in a coffee shop)? 5-point scale (Not at all important; Somewhat important; Moderately important; Very important; Extremely important).

Findings

Figure 1 shows the distribution of ratings for music recognition. When shown the statement “*When I hear music I like, I want to know the artist and title (e.g. a song playing in a coffee shop)*”, 39.4% of respondents answered that they experience this problem a few times a week or more. 38.5% of respondents answered that it’s at least *moderately important* to address the problem. To compare these results to that of the other eight needs in this study, we plotted frequency as a function of importance. We bisected each axis at the median to form four quadrants (median frequency = 20.0%, median importance = 37.7%; Figure 2, top panel). Music recognition was the second most frequent and fourth most important user need of the nine tested. In the present study, we were able to include all needs and design concepts in the focus group part of Triptech (Part 2), however, practitioners who need to reduce the number of ideas before moving on to Part 2 can use the importance x performance chart to deprioritize the concepts that address needs in the bottom left quadrant (relatively lower frequency and importance).

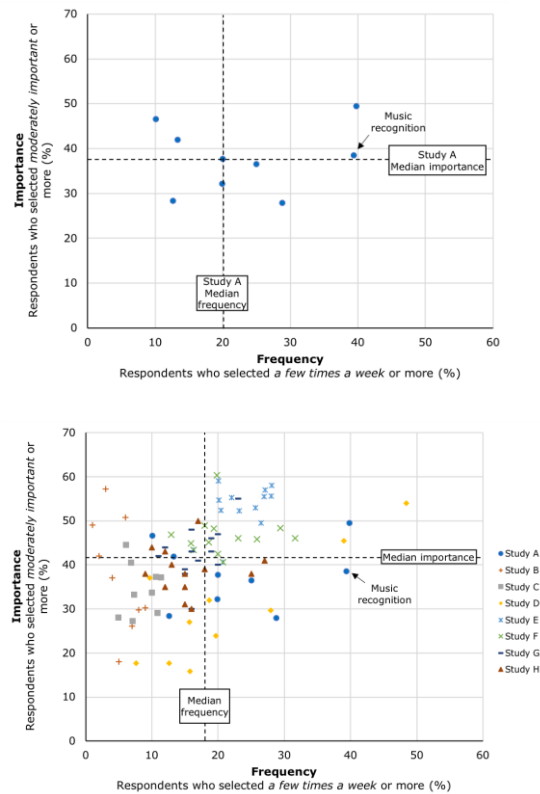


Figure 2: Importance (percentage of respondents who answered it's at least moderately important to address the need or problem) as a function of frequency (percentage of respondents who report experience the need a few times a week or more) for the study described in this paper (Study A; top panel) and for 89 needs across eight studies (bottom panel).

Part 2: Focus group

In Part 2, small groups of participants share individual and group feedback about the design concepts that address the needs tested in Part 1. For each concept, participants rate the desirability and perceived utility of the idea before moving on to the group interview. The group interview methodology is based on a mix of three existing methods: storyboarding [8,9], focus groups [5], and Speed Dating. Speed Dating (SD) is a design method used to explore applications strategies for ubiquitous computing [3]. Triptech draws inspiration from its *needs validation stage*, in which groups of target users react to concepts presented using storyboards. Unlike SD, in which the discussion is focused on the user need, how it is triggered, and how important it is to address, Triptech discussions are focused on the design concepts themselves. Participants share their reactions, likes, dislikes, potential use cases, questions, and concerns about the proposed solution. This feedback is then used to outline UX requirements and to craft future design and research efforts. Comparing the ratings for the user needs to feedback about the design concepts can reveal instances in which there is a mismatch between the problem and how the team plans to solve it.

Materials

We presented each design concept using short, three-frame storyboards, which are optimal for conveying a single feature or concept [7]. Figure 3 shows the storyboard for music recognition and the *Now Playing* concept. The first frame of the storyboard included the user need—worded exactly as it was in Part 1—and a simple illustration of the problem. The second frame and its caption showed how the proposed concept addresses the need. The third frame and its caption highlighted the value proposition of the concept by showing the happy resolution of the user need.

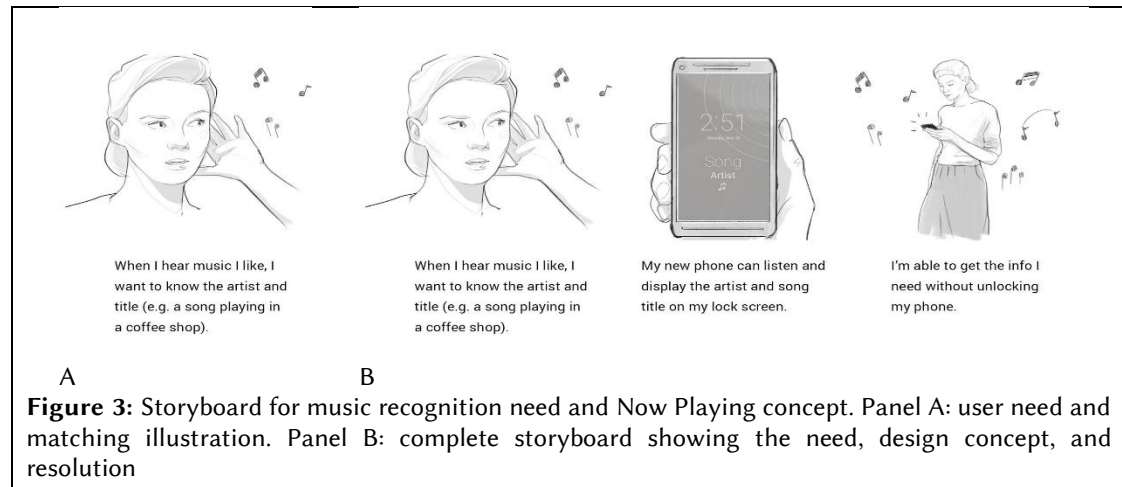


Figure 3: Storyboard for music recognition need and Now Playing concept. Panel A: user need and matching illustration. Panel B: complete storyboard showing the need, design concept, and resolution

Participants

10 participants took part in group interviews held in Google's Seattle Washington office on October 13th and 18th 2016 (6 and 4 per session, respectively). Participants were recruited using Google's user experience participant sign-up (google.com/usability). Participation took two hours.

Procedure

After a short ice breaker, we presented the first storyboard and collected user feedback following the steps below. We then repeated the procedure for each of the remaining nine concepts.

User need ratings. We showed participants the first frame of the first storyboard (Figure 3, Panel A) and asked them to rate the frequency and importance of the need using the same scales used in Part 1. We did this in order verify that the ratings given by focus group participants were representative of the ratings given by survey respondents.

Individual feedback. After all participants had entered their frequency and importance ratings, we revealed the remaining two frames of the storyboard: the concept and the resolution (Figure 3, Panel B). Participants could ask clarifying questions about the concept but were asked to refrain from sharing their initial reactions with the group. They answered the following questions individually:

- Describe your first impression of this feature
- How much do you want this feature? (7-point scale from I don't want it at all to I absolutely want it)
- How useful is this feature? (7-point scale from Of no use to Of great use)

Group feedback. We asked the following questions to provide structure to the discussion. For the first question, we asked participants to share their answers with the group, one at a time, in clockwise order. We encouraged participants to react to previous answers as well as to add their own. Once every participant had the chance to share, we opened the discussion to the group.

- Name something you like or something you dislike about this feature.
- What's an example of how you would use it?
- What questions do you have about this idea?
- What concerns do you have about this idea?
- Is there anything else you would like to share?

Findings

We analyzed participants' reactions to each one of the nine need/concept pairs. Below, we present the analysis method and the findings specific to music recognition and the *Now Playing* concept. Practitioners can repeat this process for every storyboard evaluated using the focus group portion of Triptech.

- Would it use cellular data?
- Would it drain the battery?
- How would it work?
- Is it recording anything?
- What does this mean for privacy?
- Would it work offline?
- Does it send notifications or make sounds?
- Can it be turned off?
- Can it be used to track the music I've heard?

Figure 4: Questions and concerns about the music recognition concept

User need evaluations. We conducted independent sample t-test to compare the frequency and importance ratings from the focus groups to the ratings collected using surveys in Part 1. Results confirmed that the focus group participants' ratings for music recognition were representative of the ratings given by survey respondents.² We repeated this for the remaining eight user needs and confirmed that in-person participants' ratings were representative of survey ratings.

Individual feedback. There was a strong correlation between desirability and perceived utility ratings ($r = 0.924$). We analyzed participants' first impressions of the music recognition concept according to their ratings for utility and desirability: rating of 3 or 2 = high; 1, 0, or -1 = neutral; -2, -3 = low). Those who rated the music recognition concept highly (2 out of 10 participants) liked the passive nature of the feature. "*Fantastic. I was thinking this when I saw the problem and thought 'How could this be more than just another app or command?'*"—P6. Participants who rated the concept as neutral (4 out of 10) had mostly positive reactions but expressed concerns such as battery life and data usage: "*Sounds like a battery drain -- would this work only on wifi? or pull mobile data?*"—P5, "*This feature is cool, but doesn't seem very necessary in my opinion.*"—P3. The four participants who rated the feature as low desirability/utility found it wasn't a necessity in their lives. They also mentioned concerns with battery: "*I know a lot of people who would love this feature, but I'm not a huge music buff so I would rarely use it.*"—P1, "*Sounds interesting, but sounds like it will drain the battery.*"—P8.

Group feedback. Participants reacted positively to the concept. They found the idea "cool" and were fond of its passive nature. Participants said there are already ways to find out what song is playing (e.g. launching an app, asking a digital assistant) but some felt that existing methods take too long or too many steps to activate. Participants mentioned locations in which they would like to recognize ambient music, including at a concert, in a coffee shop, at the gym, while shopping, and at work. Figure 4 shows questions and concerns raised by the group.

UX recommendations. We considered first impressions and key discussion themes to make must have and nice to have UX recommendations for music recognition. Must have: a) address user concerns about privacy, b) address concerns about battery drain, c) not consume cellular data, d) not make noise or interrupt the user, e) be easy to dismiss or ignore. Nice to have: a) ability to log or keep track of music heard, b) ability to buy the music, c) ability to rate songs, d) a way to note music of interest when driving or otherwise occupied.

3 DISCUSSION

The Triptech study influenced the development of ML smartphone concepts in general, and *Now Playing* specifically, in several ways. The surveys helped us identify which of the nine

² There was no significant difference between frequency ratings from the surveys ($M = 2.23$, $SD = 1.63$) and from the focus groups ($M = 2.90$, $SD = 1.73$); $t(414) = -1.282$, $p = .201$, $d = 0.400$, indicating that the in-person participants' responses were in line with that of survey respondents. Similarly, there was no significant difference between importance ratings from the surveys ($M = 1.31$, $SD = 1.24$) and from the focus groups ($M = 1.70$, $SD = 1.42$); $t(409) = -0.977$, $p = .390$, $d = 0.292$.

concepts were anchored on the most frequent and important user needs. As a result, we deprioritized the work on certain features (those that fell in the bottom left quadrant of Figure 2, top panel) and reallocated resources to focus on needs that received higher ratings, (e.g. music recognition, which ranked second in terms of frequency and fourth in terms of importance).

Data from the focus group sessions helped identify factors that likely influenced users' expectations, the desirability for the concept, perceived utility, and overall impression, which in turn informed the product roadmap for *Now Playing*. For example, to minimize the impact on battery life, the system uses a gatekeeper music detector that prevents the relatively computationally expensive music recognizer from triggering unless music is present. *Now Playing* addresses user privacy and data consumption concerns by running entirely on-device, never sending audio data to a server. See [1] for more details.

4 LIMITATIONS

The Triptech method has limitations that must be considered when conducting the study and analyzing its results. First, Triptech relies on participants' subjective ratings of user needs. It does not measure the degree of overlap between perceived and observed user needs [3]. This is why we encourage practitioners to draw from existing user research (e.g. foundational studies, ethnography) to hypothesize user needs or problems whenever possible.

Another limitation of the survey portion of Triptech is that it only provides a relative evaluation of frequency and importance. It can identify which needs are most frequent and most important, but only in relation to other needs in the study. This means that the thresholds for high importance and frequency will vary from study to study. We suggest that practitioners keep track of frequency and importance ratings across studies to build a picture of the range of responses for their target users. For example, the bottom panel of Figure 2 shows the survey results for 89 needs tested across nine studies with similar recruiting criteria (median frequency threshold = 18.0%; median importance threshold = 41.9%). While it is still not possible to establish absolute thresholds for high and low frequency and importance, this sort of tracking can help set targets for future product explorations.

Lastly, Triptech is subject to the limitations of the focus group methodology [5]. Results are not predictive of how well a product will fare in the market once it is released. Instead, the purpose of the group discussion is to identify areas that warrant more exploration, such as user expectations and concerns. Another is that focus groups cannot be used to quantify the relative magnitude of each topic or concern. Therefore, we recommend that practitioners report all key discussion themes and refrain from making claims about which ones mattered most to study participants.

5 Conclusion & What's next

Triptech draws from existing research methodology—surveys, storyboarding, SD, and focus groups—to form a UX evaluation method used to collect feedback during the conceptual phase of the product innovation process. In its first stage, it provides a relatively quick way to gain generalizable knowledge about the frequency and magnitude of user needs. In its second stage,

Triptech is used to assess potential users' reactions to design concepts. Its merits include the ability to identify concepts that resonate with users but also to deprioritize solutions that don't solve the needs of the target audience, thereby saving teams from making costly investments. Triptech can be adapted to the needs of various settings and audiences, e.g. [4]. Beyond its impact on *Now Playing*, the Triptech method has helped several teams conduct UX evaluation in the early stages of product development. To date, it has been used to evaluate over 100 needs across a dozen Google projects. Its surveys have reached more than 35,000 respondents and Triptech focus groups have been conducted in the United States, England, and India. We hope it will continue to be a valuable tool for collecting user feedback in the fast-moving, tech-driven environments of today's innovative companies.

ACKNOWLEDGMENTS

We would like to thank Don Barnett, Chelsea Brady, Alison Lentz, Kristen Olson, Oliver Siy, and Tim Wantland for their help preparing and conducting the first study with authors JAS and KP; and Megan Torkildson for conducting a pre-cursor study with author AS. We would also like to thank Brandon Barbello and Tom Hume for their role in incorporating Triptech data into the planning for *Now Playing* and other ML concepts.

REFERENCES

- [1] Blaise Agüera y Arcas, Beat Gfeller, Ruiqi Guo, Kevin Kilgour, Sanjiv Kumar, James Lyon, Julian Odell, Marvin Ritter, Dominik Roblek, Matthew Sharifi, Mihajlo Velimirovic. 2017. Now Playing: Continuous low-power music recognition. *Computing Research Repository (CoRR)*.
- [2] Petra Bosch-Sijtsema, and Jan Bosch. 2015. User involvement throughout the innovation process in high-tech industries. *Journal of Product Innovation Management* 32(5), 793-807.
- [3] Scott Davidoff, Min Kyung Lee, Anind K. Dey, and John Zimmerman. 2007. Rapidly exploring application design through speed dating. *International Conference on Ubiquitous Computing*. Springer, Berlin, Heidelberg.
- [4] Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, and 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.
- [5] Richard A. Krueger and Mary Anne Casey. 2014. Focus groups: A practical guide for applied research. Sage publications.
- [6] Donald A. Norman. 2010. Technology first, needs last: the research-product gulf. *Interactions* 17(2), 38-42.
- [7] Truong, Khai N., Gillian R. Hayes, and Gregory D. Abowd. 2006. Storyboarding: an empirical determination of best practices and effective guidelines. *Proceedings of the 6th conference on Designing Interactive systems*. ACM, 2006.
- [8] Corrie Van der Lelie. 2006. The value of storyboards in the product design process. *Personal and ubiquitous computing* 10(2-3), 159-162.
- [9] Rachel Krause. Storyboards Help Visualize UX Ideas. Nielsen Norman Group, www.nngroup.com/articles/storyboards-visualize-ideas/.
- [10] Arnold POS Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. 2010. User experience evaluation methods: current state and development needs. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. ACM.