

“PLAY MUSIC”: USER MOTIVATIONS AND EXPECTATIONS FOR NON-SPECIFIC VOICE QUERIES

Jennifer Thom¹ Angela Nazarian² Ruth Brillman¹
Henriette Cramer¹ Sarah Mennicken¹

¹ Spotify, USA

² University of California, Davis

jennthom@spotify.com

ABSTRACT

The growing market of voice-enabled devices introduces new types of music search requests. As voice assistants can potentially support conversational requests, music requests can be more ambiguous than requests in typed search interfaces. However, these systems may not be able to fulfill ambiguous requests in a manner that matches the user need. In this work, we study an example of ambiguous requests which we term as non-specific queries (NSQs), such as “play music,” where users ask to stream content using a single utterance that does not specify what content they want to hear. To better understand user motivations for making NSQs, we conducted semi-structured qualitative interviews with voice users. We observed four themes that structure user perceptions of the benefits and shortcomings of making NSQs: the tradeoff between control and convenience, varying expectations for personalization, the effects of context on expectations, and learned user behaviors. We conclude with implications for how these themes can inform the interaction design of voice search systems in handling non-specific music requests in voice search systems.

1. INTRODUCTION

Voice assistants and smart speakers are rapidly becoming ubiquitous. Globally, an estimated 600 million people use voice assistants at least once a week [7]. In the U.S., roughly a quarter of adults own a smart speaker, such as an Amazon Echo or Google Home [16, 28]. One of the most popular use cases for smart speakers is music listening [1, 5]. If a user approaches a music search with a specific piece of content in mind and requests it by artist and track name [14], current voice assistants are typically able to fulfill such requests.

However, voice assistants also provide users with the opportunity to search for music in a conversational and

open-ended manner without mentioning specific entities. We focus on non-specific queries (NSQs) in this research, which are requests to play music but which lack any specifications in the user’s language about which music to return. An example of a non-specific voice query would be “Play music” (since no specifics are provided) but not “Play me some hip hop,” since “hip hop” specifies a genre.

Due to the current limitations of voice search systems, interactions that do not exactly specify what music to request are less likely to be successful. When faced with a potentially unbounded space of conversational interactions with a voice assistant, users may simply not know what to say [9, 25]. As a result, users can end up falling back on habits they developed when they first obtained their voice assistants [5] and resort to more simple interactions. In addition, users’ mental models of the voice assistant’s capabilities do not always match its actual capabilities [22]. This makes it difficult for users to know what types of voice searches will end up providing the desired results.

Our research questions are as follows: (1) what are user motivations for making non-specific queries (NSQ) and (2) what do users expect from the system when they make NSQs? To study these questions, we conducted a qualitative interview study with users who make such requests. Finally, we conclude with implications for how voice assistants can better meet these user needs and propose avenues for future work.

2. RELATED WORK

2.1 Voice Assistants and Music Search

Voice has become a common interaction modality for users to search for information. In particular, music has emerged as a popular domain for voice search. This has been observed across device types, ranging from mobile phones to smart speakers. Guy’s [11] analysis revealed that music videos were the top triggered results on voice search on mobile phones. In the case of smart speakers, music listening is not only the most commonly requested functionality but requests for information about music (e.g., “Who sings this song”) also emerged as a prevalent search [1]. However, while voice interactions can potentially enable quick

¹ The third author is now at Google and has not contributed to the paper after joining Google.



music playback on a smart speaker, issues such as difficult to pronounce artist names can pose a challenge for users, making these interactions more effortful [32].

2.2 Music Search and Discovery

To observe how users would use natural language to search for music and to better understand users’ information needs, researchers have turned to general domain repositories such as Google Answers. Bainbridge et al. [3] observed that users typically employ a variety of metadata to form music searches, such as bibliographic information including performer, title of work, or date of recording. Lee [19] also observed that music searches in Google Answers are typically known-item queries.

Music discovery on music streaming services also occurs through personalized recommendations, which can be potentially supported through natural language voice queries. Such discovery often requires a user to be receptive to novelty, suggesting that listeners are selective about the situations where this would be a positive experience [18]. Lee and Price [20] observe that music listeners with more ‘adventurous’ music habits are more positive about novelty in recommendations while more ‘discerning’ listeners expect recommendations to not be novel enough to meet their tastes. Because users of voice assistants can use natural language for ambiguous queries for a variety of reasons, these systems will have to account for a listener’s appetite for discovery.

We extend the research on music search and discovery by studying a common query employing non-specific language that may not necessarily indicate a tolerance for exploration. In addition, we note that the prior research covered in this section has predominantly focused on typed search and modalities that offer visual feedback.

3. METHOD

3.1 INTERVIEWS

To elicit a rich descriptive user-centric dataset about non-specific queries that may carry meaning to users despite their simple format, we chose to employ semi-structured interviews. The data collection took place over a one month period in 2018 and consisted of 17 in-person 60 minute interviews. We sent a recruitment email to a random sample of Spotify users in a city in the Northeastern United States. The user selection criteria were: (1) owning a voice-enabled device and (2) making at least one NSQ on that device within the previous 30 days. The recruitment email also asked users to indicate what voice-enabled devices they owned and which music streaming services they used.

We sampled participants who used smart speakers, such as Google Home, Amazon smart devices as well as mobile voice assistants to cover a broader range. To better reflect the diversity of the listener population, we selected participants from a range of ages, occupation, and gender. Participant age ranged from 21 to 52 (mean = 31) with a wide range of occupational backgrounds. Table 3.1 summarizes

P#	Age	Gender	Job or Industry	Used Device(s)
1	29	f	Health care	AA, MVA, Other
2	23	m	Student	GH, MVA
3	39	m	Technology	GH, MVA
4	26	f	Asst Director	MVA
5	21	f	Student	MVA
6	25	f	Operations	GH, MVA
7	26	f	Clerk	MVA
8	35	f	Director	AA, AA
9	52	m	Education	AA
10	39	m	Dispatcher	GH, MVA
11	26	m	Advisor	MVA
12	29	m	Analyst	AA, MVA
13	33	f	Legal	AA, MVA, Other
14	34	m	Education	MVA
15	25	m	Student	GH, MVA
16	32	m	Manager	MVA, Other
17	38	m	Design	MVA, Other

Table 1. Demographic information of our study participants and their voice assistant usage. Google Home (GH), Amazon Alexa (AA), Mobile Voice Assistant (MVA)

the demographic characteristics of the participants. Participants were paid \$100 for their time, consistent with the compensation level for industry user research.

Our interviews consisted of two parts. The first part focused on presenting scenarios to elicit user motivations for making NSQs. The second part employed specific examples of common NSQs as probes to investigate user expectations for the results of NSQs.

3.1.1 Part 1: Scenarios

The interviewer began by asking questions related to user’s daily listening habits to develop an understanding of each user’s unique music listening style. The participant would then answer questions about their voice device habits regarding general domain voice interactions as well as those specific to music streaming and recommendation. We probed users’ motivations for requesting music through voice, and any factors that might discourage them. Following this discussion, we prompted participants with eight different scenarios in randomized order to describe whether and how they would choose to issue NSQs to their voice devices. We included the following scenarios: listening to music right now, during a workout, with friends at party, in the morning, after work or school, during chores, during the commute, while starting the day.

3.1.2 Part 2: Utterances as probes

Next, the interviewer presented five different wordings of NSQs selected from a pool of the 50 most common NSQs appearing in logged data in the 30 days preceding the month of the study. The interviewer then asked the participants to provide the utterances verbatim to the voice assistant in the room, which matched the assistant that participants reported owning and interacting with most frequently in our recruitment email. The five utterances used dur-

ing this part of the interview were: “Play Spotify playlist”, “Turn on the music”, “Play music from Spotify”, “Play music”, “Play my Spotify.”

3.2 Analysis

To identify broader themes encompassing the original annotation codes, researchers analyzed the interview transcripts using Braun & Clarke’s 6-step framework [6]. Thematic analysis [24] was used to develop annotation codes to identify common themes and pervasive concepts. Three coders verified the codes by annotating the transcripts independently. The annotation tags were subsequently grouped into larger conceptual categories to provide insight into participants’ behaviors, motivations, expectations and desired experiences.

4. FINDINGS

Our findings can be categorized along four themes that characterize user perceptions of the benefits and shortcomings of NSQs.

4.1 Trade-offs Between Effort and Control

We observed that our participants perceived NSQs as a convenient way to start listening to music with little effort by ceding control to the voice assistant. On the other hand, participants refrained from making NSQs when they believed that the returned content would be unpredictable. In this section, we describe how participants actively weigh the tradeoffs between user control and effort when deciding whether to make an NSQ.

4.1.1 Starting music effortlessly is more important than a specific outcome

Participants reported that they made non-specific queries as a lower effort way to request music. For instance, they described wanting to easily start a ‘lean back’ (or ‘hands off’) music listening session where the music comprised a soundtrack or background effect to the user’s activity. In these situations, participants prioritized the convenience of making a lower effort NSQs rather than requesting something specific.

I just need something to play in the background and accompany my morning. - P6, 25

Users also reported a desire to not add an additional level of effort to a current activity, especially when that activity required the use of the user’s hands or eyes. In these cases, any keyboard or typed search would require an extra level of effort, as it would involve interrupting their current activity. Below, a participant enumerates the situations where they would choose voice search and NSQs over text-based search.

Yeah, I would definitely say cooking, cleaning, housework type stuff, where I probably have my hands full. - P8, 36

Some users identified NSQs as an effortless way to discover novel music. Here, a user describes using NSQs as a fallback to specific queries when they needed an accessible and low-effort way to seek music that is new to them. For this user, a willingness to make an NSQ for low-effort music discovery is closely tied to their understanding of Spotify’s reputation. When this user does not see one service as allowing for new music discovery, they switch to an alternative service when experiencing this NSQ intent.

I’ve been using [music service X] longer than [music service Y]. And I consider [X] my go-to, let’s say...But [Y] I like when I don’t want to be so hands on...so, I would say if I can find new music, it’s normally through [Y], but I create lists – playlists on [X]. And so, a little music that I’ve basically cultivated for at least a decade...I typically go to [Y] when I’m trying to be mindless about what I’m listening to - P16, 32

Participants reported providing NSQs when they did not have something specific in mind that they wanted to hear but still wanted to begin a listening session without having to expend decision-making effort.

I don’t want to make a decision necessarily, or I can’t really think of what I want to listen to at that point. - P1, 29

This effect was particularly strong when the act of choosing a more specific piece of content felt like an obstacle or burden.

[An NSQ] also sort of takes the edge off of having to make your own playlist, which is something I like doing, but if I didn’t feel like putting the effort then. - P11, 26

This is consistent with findings observed by Hosey et al. [14] in which users reported being open to various results when they did not have anything specific in mind when searching for music. In the current study, we observed this as well, in particular when participants also wanted to initiate listening to music with as little effort and mental load as possible.

4.1.2 Desire for predictable outcome and associated feedback

Users consistently expressed the belief that they could not exercise as much control through voice search, compared to keyboard or touch searches. This was partly because users felt more context was available through typed queries.

I’m a visual person. I guess I kind of like to read all the songs that might come up in a playlist or something like that. Whereas if I’m [providing a voice request] I can’t really do that, it’s just whatever [music streaming service] picks out. - P2, 23

This perceived lack of control is at least partially tied to real-world limitations of voice assistants, which can break down at multiple points. For instance, ASR systems can struggle to correctly transcribe the non-standard spellings (e.g., for artist “6lack”), difficult or ambiguous pronunciations (e.g., for the hip hop duo “Rae Sremmurd”), foreign languages and non-Latin alphabets (e.g., Hebrew, Arabic, Greek) [32]. Named entity-recognition systems can struggle with ambiguous requests, e.g., “Play Changes.” Here, the voice system might still return an unintended result, such as playing “Changes” by David Bowie rather than “Changes” by 2Pac.

In an attempt to avoid these errors, users often make voice requests that they perceive as *safe bets*, such as artists they expect the system to understand. Below, a participant discusses not requesting a particular album because ASR systems often incorrectly transcribe it.

I definitely default to playing an artist rather than a specific album...I try to have it play like ‘Citsuoka’ [an album by My Morning Jacket]—there’ve been times when I’ve struggled and it hasn’t been able to recognize that. And so some of it is definitely a habit that I’ve developed where it’s more reliable to just go with the artist. - P3, 39

These experiences with voice systems can influence how a user will engage with NSQs. Users who trust the recommender system may start to consider NSQs as *safe bets*. Users who experience a sense of distrust due in part to NLU errors may refrain from making NSQs, believing that they would be too difficult for the system to fulfill.

4.2 Expectations for Personalization

Participants’ perceptions of NSQs are tied to their expectations of whether or not the requests would yield a personalized result. In particular, we observed that participants had varying levels of trust in the quality of recommendations that they would receive in the first turn of their voice interactions.

4.2.1 Trust that generic queries would lead to personalized results

Participants who were open to musical exploration or felt comfortable relinquishing control over their listening session often used NSQs. By making such open-ended requests, users were aware of the possibility of hearing something new or unexpected by requesting music in such an open-ended manner. This was particularly marked with participants who expressed high degrees of confidence in the service’s recommendation algorithms.

I think the algorithms on here are super sharp and I think that it does a really great job of condensing what I’m into, what I have saved, and pulling something up. - P4, 26

We observed that participants created their own theories about how they would receive personalized results from

NSQs, consistent with Eslami et al. [8]. For instance, participants hypothesized that they received personalized recommendations based on prior music listening behavior.

I’m assuming it bases it off of whatever kind of music I like. I mean, it started playing a song that I was okay with...Or, I assume, based on whatever the app thinks that I would like based on the kind of music that I currently like and have favorited and stuff. - P10, 39

Participants also expressed a varying level of comfort with proactive recommendations that could be fulfilled with an NSQ. For instance, a participant describes a desire for an NSQ result that could be personalized to meet a user need for music discovery and exploration to return something novel.

I would know at this point ‘play my [music streaming service]’ is my default where I’m at at that time, where I usually and if I want something different in the usual time and space, I have to look for something specific. Because it doesn’t know that I don’t want the usual. - P17, 38

In addition, some participants expected that personalization could be sufficiently proactive to be predictive of user need without explicit input from the user.

I think the ideal situation is I want [music streaming service] to know what I want to hear before I know what I want to hear. - P4, 26

When participants had positive expectations about the personalization capabilities of the voice assistant, they were more optimistic about making NSQs and receiving a successful result.

4.2.2 Fear that generic requests might lead to unfamiliar content

In contrast, participants who were less open to musical exploration refrained from making NSQs out of a reported fear that the returned content will be outside the range of music they normally listen to. As a result, familiar content would be more likely to result in an enjoyable listening experience to these participants.

I’m usually really specific because I like certain things. Well, as you get older, you realize this is what you like, and there’s enough music for me to go back to and things I like. I don’t discover too much anymore. Occasionally, but not as much as I used to. - P9, 52

I’m not too adventurous with my music, so I opt and like to listen to songs that are similar to songs that I have heard before. - P6, 25

These findings are consistent with Li et al.’s [21] observations that users who conduct non-focused music

searches consume more novel songs. When users want to hear familiar content, they will not use non-specific language to request it.

Finally, our qualitative interview responses indicate that users do not perceive all NSQs identically and had different expectations for outcomes depending on unique meaningful linguistic markers in the NSQ. For instance, there was a consensus among users that the presence of a personal pronoun in a non-specific request was tied to an elevated desire to hear personalized content that is suited specifically to their musical taste.

I just assumed that ‘play my [music streaming service]’ would yield a different thing than just turning on the music because the ‘my’, to me, indicates the music that I’ve already liked or downloaded. - P11, 26

4.3 Context Affects User Motivation to Make NSQs

Our participants reported on different contextual factors that affected their motivation to make a non-specific request.

4.3.1 More openness to discovery at certain times of day

Our participants reported that certain times of day were more conducive to openness to new music and searching for music in an exploratory mindset in the manner described by Hosey et al. [14].

I can see myself doing this in the mornings when I just need something to play in the background and accompany my morning. When I’m working out or when I’m sleeping or when I’m reading, I kind of want a certain mood, but in the mornings, I’m more open to exploration. - P6, 25

Participants further noted that there were certain activities that would occur at certain times of the day, which may make time of day a potentially good proxy for sensing context. This is consistent with other research suggesting that listening preferences change throughout the day [27].

4.3.2 Less individual control needed in a group setting

Social context can influence how participants perceive the utility of making NSQs. Below, a participant expressed her willingness to make NSQs with others in a group.

If I’m with a bunch of other people, my friends, they’re like, ‘Oh, you’re still listening to this playlist? Put something else on.’ Just having – it’s like a nice neutral third party, where it’s like, ‘Oh, I’m not dominating the radio, and neither are you.’ - P8, 36

Here we see social context play a role in how participants would make NSQs. In social settings, participants

reported using non-specific queries to purposely relinquish control over their music listening experience. In this specific context, being able to abdicate control allowed the music streaming service to step in as a DJ and provide music for their social listening session. This was motivated not only by the aforementioned desire for an effortless experience, but also for users to avoid being judged for their personal taste in music.

4.3.3 Desire for specific results to fit moods and activities

Because participants perceived the results of NSQs to be unpredictable, they were less willing to make NSQs when they wanted to hear music that would fit their current mood or emotional state. This user motivation aligns with previous work on how music serves to regulate affect [15,27,29,33]. We distinguish this motivation from the observation that participants make NSQs as a low-effort way to start a background music session. Here, the stakes felt higher to participants when the resulting music did not match one’s mood, potentially even altering into an unwanted state.

Some users do not expect that music recommendation systems can accomplish this level of emotional congruence in the context of NSQs. Below, a participant describes how unexpectedly hearing an album called “Planetarium” as the content returned for an NSQ—music which she describes as having a deeply emotional quality and unpleasant associations—would have a negative impact on her if she wasn’t prepared to hear it.

I would probably be more specific than that because the last thing I want is to be like, ‘Play music’ and it puts on ‘Planetarium’ and now I’m on my back on my bed with a jug of ice cream like, ‘How did this happen?’ - P4, 26

To avoid such situations, she exercises control by asking for specific artists or albums that fit her current mood rather than handing control to the music streaming service.

4.4 Learned User Behaviors

As users become more familiar with the capabilities of voice assistants, they fall fairly quickly into settled routines and habits [5]. We observed similar behaviors with our participants in how they adopted NSQs over time.

4.4.1 Tool to learn about boundaries of the voice assistant

Participants reported making NSQs out of initial curiosity. Consistent with Mennicken et al. [23], we observed that participants perceived NSQs as an opportunity to test the limits of the system and playfully learn the rules for what they can accomplish through their voice-enabled device.

I just said ‘Hey Google play some good music’ or something like that just to see what it would do, or I’ve also just [said] ‘play music’ before just because I was curious what it would play. - P2, 23

Participants felt that if they received a result that they deemed enjoyable, this would increase the likelihood that they would use these types of queries in the future again. This aligns with prior research in general web search that suggests that users will change their request strategies if their results are unsatisfactory [2, 13].

4.4.2 Simple requests facilitate habit formation

Participants who reported making NSQs frequently developed their mental models of how the system would respond and then learned to integrate NSQs into their daily routine.

I think I've kind of just made it a habit. Go up the stairs, set your keys down, set the backpack down, talk to [music streaming service], take the shoes off, so I think it's kind of ingrained into that coming home routine. - P8, 36

Similar to developing hypotheses about how the voice assistant would return personalized content, participants also created folk theories of how the system would interact with them depending on prior behavior.

I expect it to be some default behavior that happens all the time, whether it's resuming whether it's starting from like alphabetical like the stuff I've tagged in my library [...] different stuff like over time that I've gotten used to. - P3, 39

For those participants who were more familiar with possible results of NSQs, a final motivation for issuing NSQs was to resume a previous listening session.

When I come home from work [and provide an NSQ] it picks up [...] whatever thing I was listening to on my phone, so that's kind of nice. It picks up where you left off. - P8, 36

For days like that, again, where I would prefer to use [an NSQ], it's when I'm in and out of the Jeep constantly. That way I could just jump in the Jeep, I throw the phone up on the dashboard, I tell the Jeep to start playing the music and I'm already rolling out of the parking lot. If I'm gonna be in the Jeep for an extended period of time like when I'm on my way to work, then I just pick it manually. But, [...] Running errands and stuff where you're constantly in and out, it'd be nice to actually just say, [an NSQ] and it just picks up right where I left off. - P10, 39

5. DISCUSSION

Based on our findings, we discuss (1) how the user experience of NSQs could be improved through a better integration of feedback; and (2) what additional linguistic cues could be considered in the NSQs to better address user needs.

5.1 Design for Feedback

When participants perceived a lack of control over their results, they were less inclined to make NSQs. One possible way to address that concern through design is to provide contextualized results when appropriate. For example, voice output, such as text-to-speech, can be used to give additional information about how the content is personalized. Metadata, such as an artist's collaborators or genre, can be used to feed into text-to-speech contextualization of the results of an NSQ [4].

Our results revealed that participants were willing to make NSQs when they wanted to start a low-effort listening session. However, this was balanced with a reluctance to receive overly surprising results. For users who make NSQs on a device with text-to-speech output, conversational search can help guide users by offering personalized options in a dialogue. This guidance can help balance a listener's need to retain control while minimizing cognitive load [12]. To generate candidate responses in a conversational agent for music discovery and exploration, systems can employ various NLP techniques to take advantage of the semantic relationships between entities found in music corpora and catalogs [26]. This can be helpful for the 'discerning' users, observed by [20], who are not satisfied with their recommendations or users who are willing to put in effort for exploration, as observed by [14, 17].

5.2 Leverage Linguistic Cues

Our findings can also be applied to queries slightly more specific to the ones currently studied. Our observations about NSQs can provide insight into descriptive music queries, such as "Play hip hop" or "Play something calming." While there are indeed categorically 'wrong' answers for descriptive queries such as returning a classic rock song for "Play hip hop," we suggest that expectations of personalization and the trade-off between effort and control remain important motivators for users who make descriptive queries.

Another future research direction related to non-specific language would be requests such as "recommend me something". While this type of request does not specify the type of content, the intent expressed by "recommend" suggests that a user is open to new content or to personalized results. Prior research [10] has uncovered different user goals for discovery-related content, suggesting that fulfillment of these voice requests should be aware of the different user needs behind these recommendation-related requests. Additionally, if the assistant has knowledge of user context such as location or time of day, the fulfillment of parsimonious "recommend me something" utterances can potentially be highly personalized [30, 31].

In summary, our research suggests that user intent when making ambiguous and non-specific requests can vary depending on user tolerance for effort and novelty, and that this tolerance can impact the level of user trust with the system. Future directions include iterating on system design to better support NSQs and further investigating utterance language to better understand user intent.

6. REFERENCES

- [1] T. Ammari, J. Kaye, J. Tsai, and F. Bentley. Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(3, Article 17):1–28, 2019.
- [2] A. Aula, R.M. Khan, and Z. Guan. How does search behavior change as search becomes more difficult? In *Proc. of the 2010 CHI Conference on Human Factors in Computing Systems*, pages 35–44, Atlanta, Georgia, USA, 2010.
- [3] D. Bainbridge, S.J. Cunningham, and J.S. Downie. Analysis of queries to a Wizard-of-Oz MIR system: Challenging assumptions about what people really want. In *Proc. of the 4th Int. Society for Music Information Retrieval Conf.*, Baltimore, Maryland, USA, 2003.
- [4] M. Behrooz, S. Mennicken, J. Thom, R. Kumar, and H. Cramer. Augmenting music listening experiences on voice assistants. In *Proc. of the 20th Int. Society for Music Information Retrieval Conf.*, pages 303–310, Delft, The Netherlands, 2019.
- [5] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3, Article 91), 2018.
- [6] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.
- [7] E. Cherian and J. Pounder. Speak easy mindshare report. Technical report, J. Walter Thompson, London, UK, 2017.
- [8] M. Eslami, K. Karahalios, C. Sandvig, K. Vaccaro, A. Rickman, K. Hamilton, and A. Kirlik. First I "like" it, then I hide it: Folk theories of social feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2371–2382, San Jose, California, USA, 2016.
- [9] A. Furqan, C. Myers, and J. Zhu. Learnability through adaptive discovery tools in voice user interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1617–1623, Denver, CO, USA, 2017.
- [10] J. Garcia-Gathright, B. St. Thomas, C. Hosey, Z. Nazari, and F. Diaz. Understanding and evaluating user satisfaction with music discovery. In *Proc. of the 41st Int. ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 55–64, Ann Arbor, MI, USA, 2018.
- [11] I. Guy. Searching by talking: Analysis of voice queries on mobile web search. In *In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 35–44, Pisa, Italy, 2016.
- [12] A. N. Hagen. The playlist experience: Personal playlists in music streaming services. *Popular Music and Society*, 38(5):625–645, 2015.
- [13] A. Hassan, R.W. White, S.T. Dumais, and Y. Wang. Struggling or exploring?: Disambiguating long search sessions. In *Proceedings of the 7th ACM Int. Conference on Web Search and Data Mining*, pages 53–62, New York, NY, USA, 2014.
- [14] C. Hosey, L. Vujović, B. St. Thomas, J. Garcia-Gathright, and J. Thom. Just give me what I want: How people use and evaluate music search. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 299:1–299:12, Glasgow, Scotland UK, 2019.
- [15] P.N. Juslin and J. Sloboda, editors. *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press, Oxford, UK, 2011.
- [16] B. Kinsella and A. Mutchler. Smart speaker consumer adoption report. Technical report, 2019.
- [17] J. Kiseleva, K. Williams, A. Hassan Awadallah, A.C. Crook, I. Zitouni, and T. Anastasakos. Predicting user satisfaction with intelligent assistants. In *Proc. of the 39th Int. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 45–54, Pisa, Italy, 2016.
- [18] A. Laplante. Users' relevance criteria in music retrieval in everyday life: An exploratory study. In *Proc. of the 11th Int. Society for Music Information Retrieval Conference*, pages 601–606, Utrecht, Netherlands, 2010.
- [19] J.H. Lee. Analysis of user needs and information features in natural language queries seeking music information. *Journal of the American Society for Information Science and Technology*, 61(5):1025–1045, 2010.
- [20] J.H. Lee and R. Price. User experience with commercial music services: An empirical exploration. *Journal of the Association for Information Science and Technology*, 67(4):800–811, 2016.
- [21] A. Li, J. Thom, P. Chandar, C. Hosey, B. St. Thomas, and J. Garcia-Gathright. Search mindsets: Understanding focused and non-focused information seeking in music search. In *Proc. of the World Wide Web Conference*, pages 2971–2977, San Francisco, CA, USA, 2019.
- [22] E. Luger and A. Sellen. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5286–5297, San Jose, CA, USA, 2016.
- [23] S. Mennicken, R. Brillman, J. Thom, and H. Cramer. Challenges and methods in design of domain-specific voice assistants. In *2018 AAAI Spring Symposia*, Palo Alto, CA USA, 2018.

- [24] S.B. Merriam. *Qualitative Research in Practice: Examples for Discussion and Analysis*. Jossey-Bass Inc, Hoboken, NJ, 2002.
- [25] C. Myers, A. Furqan, and J. Zhu. The impact of user characteristics and preferences on performance with an unfamiliar voice user interface. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 47:1–47:9, Glasgow, Scotland UK, 2019.
- [26] S. Oramas, L.E. Anke, F. Gómez, and X. Serra. Natural language processing for music knowledge discovery. *CoRR*, abs/1807.02200, 2018.
- [27] M. Park, J. Thom, S. Mennicken, H. Cramer, and M. Macy. Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, 3:230–236, January 2019.
- [28] Edison Research and NPR. The smart audio report. Technical report, Somerville, NJ, 2019.
- [29] S. Saarikallio and J. Erkkilä. The role of music in adolescents' mood regulation. *Psychology of Music*, 35(1):88–109, January 2007.
- [30] M. Schedl and A. Flexer. Putting the user in the center of music information retrieval. In *Proc. of the 13th Int. Society for Music Information Retrieval Conference*, pages 385–390, Porto, Portugal.
- [31] M. Schedl, A. Flexer, and J. Urbano. The neglected user in music information retrieval research. *Journal of Intelligent Information Systems*, 41(3):523–539, December 2013.
- [32] A. Springer and H. Cramer. "Play PRBLMS": Identifying and correcting less accessible content in voice interfaces. In *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, pages 296:1–296:13, Montreal, Quebec, 2018. ACM.
- [33] M. Zentner, D. Grandjean, and K.R. Scherer. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4):494–521, August 2008.