



Information Research - Vol. 29 No. 2 (2024)

‘Alexa, play metal’: exploring music selection and personal information management via voice assistants

Jochen Steffens*, Jesse David Dinneen*, Sascha Donner, Tom Potthoff

*Co-first authors

DOI: <https://doi.org/10.47989/ir292646>

Abstract

Introduction. Music streaming services have changed how music is played and perceived, but also how it is managed by individuals. Voice interfaces to such services are becoming increasingly common, for example through voice assistants on mobile and smart devices, and have the potential to further change personal music management by introducing new beneficial features and new challenges.

Method. To explore the implications of voice assistants for personal music listening and management we surveyed 248 participants online and in a lab setting to investigate (a) in which situations people use voice assistants to play music, (b) how the situations compare to established activities common during non-voice assistant music listening, and (c) what kinds of commands they use.

Analysis. We categorised 653 situations of voice assistant use, which reflect differences to non-voice assistant music listening, and established 11 command types, which mostly reflect finding or refinding activities but also indicate keeping and organisation activities.

Results. Voice assistants have some benefits for music listening and personal music management, but also a notable lack of support for traditional personal information management activities, like browsing, that are common when managing music.

Conclusion. Having characterised the use of voice assistants to play music, we consider their role in personal music management and make suggestions for improved design and future research.

Introduction

For most of human history, listening to music has been almost entirely a collective experience, restricted to certain important times of day or ceremonies. Through advances in recording and transmission technology in the last century such as the radio in the 1920s, the Walkman in the late 1970s, and the emergence of smartphones and streaming services in the 2000s, there have been several changes to how music is experienced, referred to by media scholars as musicalization (Pontara & Volgsten, 2017). For example, music moved from being mostly a rare and collective live experience to being mostly an individualised, everyday, and ubiquitous activity. Further, with the increasing possibilities to store music in digital format, music collections grew considerably (Cunningham et al., 2004), thus necessitating individual listeners perform format-specific personal information management tasks like organising local MP3s or making playlists of streamable songs.

More recently still, the use of digital voice assistants such as Siri, Alexa, and Cortana, has increased, and with it increased the voice assistant-powered selection of music (Ammari et al., 2019; Dubiel et al., 2018) and ability (and sometimes necessity) to interact with digital items such as songs and files whether in local collections or remote databases. The new features, possibilities, and challenges of using voice assistants to play music and manage collections remain relatively unexplored, however (Wirfs-Brock et al., 2020). For example, it is still not well understood when and how people interact with voice assistants to play music, what kinds of commands they use, how the formulation of a voice command depends on the listening situation (e.g., present activity), and what the voice-assistant paradigm will mean for the management of music collections (i.e., personal music management).

To address the above gaps in knowledge and begin to explore voice assistants in personal music management we examined what

commands voice-assistant users might use when listening to music and in which situations. Specifically, commands and situations were collected from 248 participants through online and in-lab surveys, then the 653 usable responses (commands and situations) were categorised into 11 types each and the situations were compared to those common to non-voice assistant music listening. Together the data indicate the use of voice assistants to play new or known songs in social and hands-free situations, but also a need to perform management actions like keeping and organising songs and playlists, which are not yet well supported.

In what follows we describe the current scholarly knowledge about music listening activities and interfaces and identify the exact questions to be answered, then provide detail about the methods of our study and report its results. We then present and discuss the answers to our specific research questions and consider what they mean for voice assistant-powered personal music management and possible future design and research activities thereof.

Literature review

Studies of personal information management – how individuals store, manage, and later retrieve and reuse information (Jones et al., 2017) – have begun to examine the management of music in locally stored collections and from streaming services. Here we review digital personal music management before and after the introduction of streaming services and voice assistants.

Local personal music management

Music consumption and integration into daily life has increased across the last half century, first through format-specific players (e.g., radio, cassette, or MP3 players) and now commonly in various digital formats on personal computing devices (e.g., laptops and smartphones). Although music streaming services provide an alternative to storing music locally (Bergman et al., 2022b), as of a few years

ago it was still common to store tens of thousands of digital audio files on desktops and laptops (Dinneen & Julien, 2019), and relatively little is known about the personal information management that offline collections necessitate and online collections enable. Early studies of user interactions with digital music collections examined home media management (Sease & McDonald, 2009) and suggested which metadata are relevant to browsing personal collections (e.g., title, artist, genre, rhythm, and mood) and identified the various organisational schemes used (Cunningham et al., 2004). Fewer studies still looked at personal music management since the era when people manually synced songs to their mobile device (cf. Brinegar & Capra, 2010).

Music streaming and management via playlists

Since their introduction, music streaming services introduced a second paradigm to digital music playback, offering users a wide selection of music that was not stored locally, and thus need not necessarily be managed by them (Lee & Waterman, 2012). Rather than primarily collecting and managing music recordings (as files), on streaming services users primarily collect and manage music links in the form of bookmarks and playlists (Sesigür, 2020; playlists of local music have also long been available). Indeed, studies indicate that, regardless of a listener's age, using streaming applications like Spotify or Apple Music leads (on average) to more exploration of new music, but often at the cost of neglecting personally managed collections (Bergman et al., 2022a), and such neglect can lead to less excitement in the overall musical experience and to a less certain musical self-identity (Bergman et al., 2022b).

A playlist is 'a set of songs meant to be listened to as a group, usually with an explicit order' (Fields et al., 2010) or 'under a particular principle' (Barrington et al., 2009) such as an emotion (e.g., sad music), a location (e.g., songs for the gym), an event (e.g., wedding) or an activity (e.g., travelling; Cunningham et al., 2006). Playlists can be considered a powerful tool to organise and retrieve from a music collection (Kamalzadeh et al., 2012), and sharing

them is used for communication like expressing love or farewell wishes (Cunningham et al., 2006). Further, playlists can be generated automatically (e.g., to provide or promote new music), and so are popular in both personal and commercial contexts (Fields, 2011).

Voice user interfaces and voice assistants

By using voice interactions, digital assistants are changing how various tasks can be performed, including personal information management tasks (e.g., creating and reviewing calendar entries with verbal commands). A voice interface for music retrieval typically allows the user to query a database with artist names, song titles, genres, or keywords (Bainbridge et al., 2003), and might further add data like the user's playback history to choose the correct action or optimise the results (Lee et al., 2012; Tzanetakis & Cook, 2002). But it is not yet clear what the outcome of these changes will be nor if they genuinely address user needs (Khaokaew et al., 2022). Because of the auditory nature and lack of visual feedback in voice user interfaces, the relevant features, affordances, and allowed voice commands are learned by users through ongoing verbal interaction – trial and error and asking for help – that can cause frustration (Furqan et al., 2017; Myers et al., 2019; Myers et al., 2021; Yankelovich, 1996). Undesirable moments include the voice user interface capturing the user's commands erroneously, misunderstanding their intent, or simply failing to produce the desired response (Myers et al., 2018). In some cases, users respond by further articulating, or less frequently, further simplifying or completely changing their query, as well as resorting to using the graphical user interface (i.e., multimodal interaction; Myers et al., 2018). In positive cases, people eventually become proficient with voice user interfaces, or even intentionally explore the system's behaviour to learn it faster, but some also simply continue to struggle to complete tasks indefinitely (Myers et al., 2019b). Adaptive suggestions are one promising solution to

address these issues (cf. Myers, 2019), but have so far been explored only with Calendar software, which has a different set of data and possible interactions than music playback software. Thus, it remains unknown what kind of music-related voice commands are used in light of these affordances and constraints.

Voice assistants are personas, often powered by artificial intelligence and natural-language processing technologies that use a voice-user interface to accept and respond to a wide variety of user commands (Azzopardi et al., 2022) and generally control features across an operating system. Users have high expectations of voice assistants (Luger & Sellen, 2016; e.g., Amazon's Alexa or Google's Assistant) and thus integrate them into their daily domestic routines by (for example) requesting various information and playing music (Ammari et al., 2019).

Voice assistants and personal music management

Whether employed on local collections or for streaming music, the use of voice assistants is currently unlike retrieval via text search, which has a visual interface and presentation of results (but similarly relies on recall) and is even more unlike the more popular retrieval method of browsing, which relies on recognition and spatial cognition (Benn et al., 2015). In other words, traditional visual search results and browsing displays present visually persistent content for a user to review, whereas audio results do not (i.e., they must be read again). What this means for users doing personal music management is not yet clear, as relatively little is known about users' relevant cognition, mental models, queries and tasks when using voice assistants (Stone, 2022). It is understood that voice assistants have different properties and that this leads to different user experiences and different skills required to succeed in using voice assistants (Beirl et al., 2019; Brüggemeier et al., 2020). Because voice assistants are a relatively new and very rapidly developing technology, findings about the various features or differences may quickly become obsolete, and assistants could do more to support

participants' varied tasks (Wirfs-Brock et al., 2020).

Voice assistants thus remain relatively understudied territory for music playback and personal music management and so considerable gaps in knowledge persist. Between the novel affordances and constraints of the voice user interface, users' expectations and challenges, and the trade-off of recall and recognition, there are many uncertainties about what voice assistants can or should be like, especially from the perspective of human-computer interaction and personal information management research. Voice assistants further advance the emphasis streaming places on retrieval over keeping and organising (Bergman et al., 2022a, 2022b); but it is unclear if this results in an absence of commands to keep and maintain a collection (i.e., less personal information management), or if voice user interfaces encourage more retrievals for discovery than for known items, and if this is desirable. Further, older people search much more for computer files – people over fifty search over four times as much as those in their twenties (Bergman et al., 2019) – and navigate to them relatively little. Yet it is unclear if they prefer voice input, which relies on verbal recall rather than visual recognition. Finally, it is unknown if users attempt to find songs again, given that they do not store and thus do not re-find the item, and overall one could wonder if the voice assistant paradigm enables any new kinds of music listening (e.g., conversational music exploration) or new kinds of personal music management.

Summary

Answering the gaps in knowledge identified above requires first understanding the basics of how users interact with voice assistants to play music, but there has been a lack of studies about this (Lemström & Tzanetakis, 2017). As a result, it remains unclear what the experience of voice assistant-powered personal music management is like, including basic phenomena like which commands people might use and in which situations. Since the primary intended use of voice assistants in the context of music today seems to be just playing (i.e., finding, retrieving, or re-finding) music, and since

finding and re-finding are the start and end of the personal information management process (Bergman & Whittaker, 2016; Jones et al., 2017), we begin exploring this problem space by examining the situations and commands used when playing music rather than keeping or organising it (e.g., voice commands for building playlists or organising existing collections).

Methods

To explore personal music listening and management with voice assistants, especially its setting and users' commands, we undertook a survey study collecting listening commands and situations and employed both qualitative and quantitative analyses. In light of the knowledge gaps identified above, our specific research questions to be answered are:

RQ1. In which situations might people use voice assistants to play music?

RQ2. How do the reported situations compare to the music listening activities common when music is not selected by use of voice assistants?

RQ3. What kinds of commands do people give to voice assistants when playing music?

RQ4. To what extent can the kinds of commands be predicted by the music listening activities?

Data collection

Participants were welcomed and informed about the motivation and procedure of the study. It was explained that the study would take around 10 minutes, that all information would be treated confidentially, that no identifying information would be obtained, and that participants were free to cease participation without consequence at any time. To investigate the use of voice assistants in digital music listening and the relationship between the formulation of voice commands and the activity at-hand, participants were asked about their use of voice assistants in various situations. In detail, we aimed at capturing up to three self-selected situations in which participants would use a voice assistant to play music in day-to-day life (i.e., we elicited

real, potential, intended, and/or hypothetical use). In this way the commands would not be dependent upon the affordances and constraints of a particular voice user interface, nor constrained to specific personal information management actions (i.e., refinding may be elicited but searching, sharing, and other actions may also), and the activities and voice commands could be freely considered and formulated by participants.

The survey consisted of two parts. In the first part, participants reported their socio-demographic information (e.g., age and gender), whether the voice assistants *Siri*, *Alexa*, *Google Assistant*, *Cortana*, and *Bixby* were familiar to the participants or not, and how often they would use voice assistants for music selection (daily, weekly, monthly or never). In the second (and main) part of the survey, we asked each participant for three self-reported, day-to-day situations in which they would use voice assistants to select music and what command they would use in each situation. Both details were provided without constraints into a free text field (i.e., not selecting from pre-made options or template answers).

Recruitment and data set

The questionnaire was administered online to both remote and in-lab participants in German, using the Web-based survey tool *Limesurvey*. The invitation to participate was distributed via university mailing lists, social media channels, and via the Web platform *clickworker.de*. In total, 248 persons (155 male, 90 female, three diverse; mean age = 36.6 years, standard deviation = 12.5) participated voluntarily in the survey and reported at least one valid music listening situation. Among those, 23 graduate students filled out the questionnaire in a computing lab of Technische Universität Berlin under the otherwise same conditions to receive course credits for their participation. Users of *clickworker.de* (n = 190) received a monetary compensation of 1€.

Most respondents (184) gave complete information on all three situations in which they would use voice assistants to select music (the rest reported on only one (27) or two (37) situations, still sufficient for analysis). The final

result was $m = 653$ different voice commands that are or would be used to select music using a voice assistant. Of the 248 participants surveyed, 79.0% reported using streaming services to listen to music, 56.9% use their own digital music collection, 58.9% use the radio, 41.5% use compact discs and 14.1% use vinyl records. Information on knowledge of particular voice assistants revealed that the best-known voice assistant was *Alexa* (91.1%) followed by *Siri* (83.5%), *Cortana* (56.9%), *Google Assistant* (57.3%), and *Bixby* (15.7%). Only 2.8% of the respondents did not know any of the voice assistants. Finally, 15.0% of the respondents reported using a voice assistant daily to select and play music, while 13.8% use one every month for that purpose and 14.6% use one less than once a month; roughly half of the participants (56.5%) had never used a voice assistant to select music.

Data processing and analysis

Categories for classifying voice commands were created by identifying and examining relevant keywords among the data (e.g., song titles, genres, or particular artists). All commands were then coded by an author and a second coder not previously involved in the analysis. The reported music listening situations were then coded independently by two authors according to the classification of music listening activities derived by Greb et al. (2018), which distinguishes between *Being on the move*, *Housework*, *Working & Studying*, *Pure music listening*, *Party*, *Relaxing and falling asleep*, *Exercise*, *Coping with emotions*, *Making music*, *Social activity*, and *Others*.

To further test to what extent the (nominal-scaled) kinds of commands can be predicted by the (also nominal-scaled) music listening activities (RQ4), we computed a generalised multinomial mixed-effects model with the MCMCglmm-Package in R / R Studio (Hadfield, 2010; R Core Team, 2015). This model included the reported activities as fixed effects and a random intercept for each participant to take into account the hierarchical data structure due to the repeated reports of up to three situations by each participant violating statistical independence assumptions. To avoid sparse matrices and a lack of statistical power,

we only included activity and command categories with more than 40 observations, respectively, while omitting the vague category 'Other / Unspecific'. The command category 'Song Title' and activity category 'Being on the move' served as reference categories which were left out of the calculation. Determining a reference category has to be performed in an exclusive categorical system with n categories since the last piece of information is redundant. This also means that all effects have to be regarded as relative compared to the reference category. Finally, to determine the relative influence of stable (i.e., individual/person-related) and varying (i.e., situational) factors on the usage of different voice command categories, we computed adjusted intraclass correlation coefficients for each command category utilising the lme4 package (Bates et al., 2015). For all analyses, the significance level was set to $\alpha = .05$.

Results

Situations of voice assistants for music listening

The results of categorising the reported voice assistant-powered music listening situations into the existing 11 categories (Greb et al., 2018) are presented in Table 1. It reveals that activity reported most often was Housework (25.1%). This category includes activities such as washing up, cleaning, getting ready, taking a shower, and eating alone. The second most frequent category includes other / nonspecific activities (20.4%) which could not be coded to one of the other categories (e.g., 'at home'). Moreover, using voice assistants to listen to music when being on the move was reported in 15.3% of the situations. Relaxing and falling asleep was reported in 14.5% of the situations, including in which the main activity was relaxing, recuperating, or trying to fall asleep. This category includes situations in which participants reported to be in the car, train, subway or bike. Social activity (7.5%) includes situations in which the main activity was interacting with others, for example cooking, eating or playing with friends. Situations in which the main activity was working and studying were reported by 6.0% of the participants in the context of music listening

with voice assistants, followed by exercise (i.e., exercising or doing sports, 5.8%). Finally, the least frequent situations in which participants would use voice assistants to listen to music

included party (2.9%), pure music listening as the main activity (2.1%), and making music (0.5%). Finally, no 'coping with emotions' was reported by the participants.

Activity while using voice assistant	Percentage frequency [%]	Percentage frequency [%] Greb et al., 2018
Housework	25.1	15.0
Others / Nonspecific	20.4	12.1
Being on the move	15.3	28.4
Relaxing and falling asleep	14.5	6.5
Social activity	7.5	1.2
Working and studying	6.0	13.3
Exercise	5.8	5.8
Partying	2.9	6.8
Pure music listening	2.1	7.3
Coping with emotions	–	2.2
Making music	0.5	1.3

Table 1. The 11 activity categories and their frequencies

Note: Each situation described (N =653) was classified into one activity category according to Greb et al. (2018, p.12)

Comparison of voice assistant music listening situations to non-voice assistant music listening situations

In the next step, we compared the percentage frequency of music listening situations in which humans (would) use voice assistants with that of non-voice assistant music listening situations, as obtained by Greb and colleagues (2018) in an online study with a similar design. Here, a chi-square test of independence reveals a significant difference across the two frequency distributions, $\chi^2(10) = 26.0, p < .01$. In detail, Table 1 shows that in this study, the

frequency of participants reporting doing housework while listening to music was 10.1% higher in the current study (voice assistant) than in the one by Greb et al. (without voice assistant). The important role of manual work is further supported by 40 situation descriptions (6.1%) in which participants reported using the voice assistant to listen to music to have their hands free (e.g., when cooking) or in which they did not want to touch their listening device (e.g., smartphone) with dirty hands. Also, relaxing and falling asleep ($\Delta=8.0\%$) and social activity ($\Delta = 6.3\%$) were reported more

frequently in the context of voice assistant situations.

Another major difference in the prevalence of music listening activities in voice assistant and non-voice assistant situations was found when looking at the activity 'Being on the move'. Here, the percentage frequency is more than 13.1% higher in non-voice assistant compared to voice assistant situations. Also, we observed a higher prevalence in non-voice assistant situations for working/studying ($\Delta = 7.3\%$).

Voice commands used and prediction by music listening activities

Categories for classifying the 653 freely formulated voice commands were created by identifying and examining relevant keywords among the data (e.g., song titles, genres or particular artists), which resulted in eleven different categories which are presented in Table 2. The interrater reliability between the two coders as measured by Cohen's κ were .759, which constitutes a substantial agreement between the coders (Landis & Koch, 1977).

The most common category *Song title* (19.8%) is also the most specific one and directly refers to the music title a person wants to listen to, for example 'Siri, play *The Moldau*'. The category *Unspecific* (15.5%) comprises all commands that are only intended to play music that is not further specified (e.g., 'Alexa, play some music!'). Further voice commands refer to the *Emotional-semantic expression and the function of music listening* (14.5%) including commands such as 'Siri, play music to relax!' or 'Play music for motivation'. The *Playlist* category (12.6%) includes voice commands asking to play a particular playlist (e.g., 'Play my work playlist!'). Furthermore, the *Genre* category (8.7%) contains commands demanding a specific music genre or style (e.g., 'Alexa, play Metal'). The *Artist* category (8.3%) refers to commands referring to particular artists or composers, for

instance, 'Hey Google, play *Nightwish!*' or 'Alexa, play *Mozart!*'. A small portion of voice commands (*Others*, 5.7%) were not easily categorisable due to a lack of information nor similar enough to other commands to warrant a new category (e.g., 'Siri, skip the next 30 seconds', 'Play x on Spotify', 'Alexa, play..', 'Nothing'). The 'Others' category can be distinguished from the 'Unspecific' category in a way that the commands categorised as 'Unspecific' still refer to musical choices which could be interpreted by a voice assistant whereas the 'Others' category lacks such information or refers to the music playback in a broader sense.

Finally, a minority of situations included commands in which participants stated they would simply ask the voice assistant to turn on the *Radio* (4.9%, e.g., 'Siri, start-up Spotify and play Pop Radio'), to listen to their *Personal collection / Favourite songs* (4.3%, e.g., 'Play songs that I listen to frequently'), *Hits* (3.8%, e.g., 'Alexa, play the latest hits'), or *a particular album* (2.0%, e.g., 'Ok Google, Play *album*'). Only eight commands expressed an explicit desire to explore or listen to new music (1.2%, e.g., 'Hey Siri, play new pop music') and half of those were to play new music from already familiar artists or playlists.

Of the commands captured, 39.4 % have been used in the past, according to the self-reports. By contrast, 60.6% of the specified commands have never been used by the respondents and thus are of hypothetical nature. Results, however, indicate that the frequency of identified command categories does not differ depending on the real vs. hypothetical nature of commands, as indicated by a non-significant Chi²-Test, $\chi^2(10) = 12.7$, $p = .24$. This finding suggests that also the not (yet) used commands may have explanatory power regarding how music is or will be chosen using voice assistants.

Voice command category	Percentage frequency [%]	Interclass correlation coefficient
Song title	19.8	.956
Unspecific	15.5	.974
Emotional-semantic expression and function of music listening	14.5	.529
Playlist	12.6	.804
Genre	8.7	.914
Artist	8.3	.938
Other	5.7	.966
Radio	4.9	.931
Personal collection / Favourite songs	4.3	.920
Hits	3.8	.942
Albums	2.0	.947

Table 2. The 11 *voice command categories* identified in the qualitative analysis (n=653)

In the next step, we aimed at connecting the voice command to the aforementioned music listening activities. The interrater reliability of the coding of music listening activity categories between the two coders as measured by Cohen's κ was .755 which again constitutes a substantial agreement between the coders. We further computed intraclass correlation coefficients to estimate the relative contribution of (stable) individual and (varying) situational influences by means of the music listening activity on the type of voice command used by the participants. Therefore, we computed a null model only including a random intercept for each participant. Results displayed in Table 2 (Column 3) show that the intraclass correlation coefficient varied from .529 (Emotional-semantic expression and

function of music listening) to .974 (Unspecific), meaning that a large amount of variance (52.9-97.4%) in the use of voice commands can be attributed to stable individual factors, suggesting that participants in our study had a strong tendency to use the same voice command category repeatedly. By contrast, there is not much variance left to be explained by variables varying across situations. Nevertheless, in the next step, we computed a generalised multinomial mixed-effects model to test whether music listening activities (independent variable) can predict the obtained voice commands (dependent variable).

Table 3 displays the observed activity-dependent post means with their significances of the generalised multinomial mixed-effects

model. The table suggests several interesting effects; for instance, when performing housework (as opposed to being on the move), participants had a higher tendency to ask the voice assistant to play music from a particular playlist, genre, or artist, than calling for a specific song title (reference category). Also, when relaxing or falling asleep (in contrast to being on the move), they had a higher tendency

to call for a playlist or artist than asking for a particular song title. Finally, when being involved in a social activity (again compared to being on the move), people were more likely to demand a particular artist than a song title. For the other voice command categories 'Unspecific' and 'Emotional-semantic expression', no effects were observed.

Voice command category	Post mean (Housework)	Post mean (Relaxing & falling asleep)	Post mean (Social activity)
Unspecific	2.02	2.12	1.95
Emotional-semantic expression	1.01	0.32	1.16
Playlist	2.88**	3.06**	0.29
Genre	1.66*	1.49	-0.86
Artist	4.76***	3.97***	3.39*

Table 3. Fixed-effects outcome of the generalised multinomial mixed-effects model. The table presents parameter estimates (i.e., post means) and their significances (*: $p < .05$, **: $p < .01$, ***: $p < .001$) of the model predicting the obtained voice command categories (dependent variables) by means of music listening activity categories (independent variables).

Note: All effects presented must be regarded relative to the voice command category 'song title' (voice command) and activity category 'being on the move' (i.e. reference categories).

Discussion

In the course of the present study, we investigated which verbal commands listeners do or would use to play music via voice assistants and in which situations. Further, we explored whether and to which degree activities during music listening can predict the formulation of the obtained verbal command categories. After briefly answering the research questions and considering their implications for personal music listening with voice assistants, we synthesise points across the answers to identify implications for personal music management specifically and personal information management more broadly.

RQ1 In which situations might people use voice assistants to play music?

To answer this question we categorised 653 situations into 11 situation categories (established by Greb et al., 2018), presented above (see Table 1). These categories reflect the variety of music listening situations that occur in daily life (e.g. moments of work, leisure, socialising and transit), which demonstrates a variety of uses for voice assistants for music listening and partly mirrors non-music uses of voice assistants (Krause et al., 2014; O'Hara & Brown, 2006). However, not all situations appear equally amenable to such use. The relative frequency of situations like housework, relaxing and falling asleep, and social activity suggest that voice assistants might be

particularly attractive to participants when they are either occupied with some kind of manual work (similar to the context for hands-free search; Ammari et al., 2019) or do not want to tangibly interact with the music device. Indeed, 6.1% of the commands refer to 'hand' or 'hands', most often in the context of housework, in particular cooking, cleaning or gardening. Such activities are presumably not common contexts for personal information management, music-related or otherwise. By contrast, in contexts where participants are in transit, on the move or work in an office, although personal information management is common in such cases (e.g., retrieving email; Lanctot & Duxbury, 2021) interacting with voice assistants might be less desirable. In such cases the audio space is often shared with other people and verbal communication with a digital device is impractical or socially undesirable (i.e., direct contact with the device is preferable). For example, it is plausible that being overheard giving music commands to a voice assistant (e.g., 'create a playlist to calm me down at work') is not as acceptable as being overheard asking it to handle shared work tasks ('create an appointment') or to adjust the temperature, because the former may reveal something personal (e.g., about personal information management and/or music), violate etiquette, or both.

RQ2 How do the reported situations compare to the music listening activities common when music is not selected by use of voice assistants?

We identified some differences in the distribution of music listening activities to what was reported in prior works examining when music was played without the use of voice assistants. We identified that the situations not amenable to tangible device interaction (housework, relaxing and falling asleep, or social activities) were more than ten percent higher than in the comparable non-voice assistant study (Greb et al., 2018), which further supports our interpretation of the results of RQ1 as discussed above. A further difference is that the frequency of 'Being on the move' and 'Working/studying' situations were higher in non-voice assistant than voice assistant music

listening situations. This might be because situations like using a mobile device on a train or working at one's home computer are amenable to hands-on interaction and do not require the use of voice assistants. But it is also plausible that such situations are where more elaborate personal music management happens, like making edits to a playlist before sharing it (Bergman et al., 2022a). In other words, perhaps such tasks are difficult with a voice user interface (e.g., 'tell me what is on this playlist; stop, remove that song; resume telling me...'), or perhaps users are simply not sure if it is possible, and so these tasks might be reserved for tangible interaction (e.g., touch screen or mouse and keyboard). It is not possible to confirm this with the data collected, as discussed in the limitations below, and so such tasks should be the subject of further study.

RQ3 What kinds of commands do people give to voice assistants when playing music?

The eleven command categories in Table 2 (not to be confused with the eleven situation categories in Table 1) and the distribution of commands together indicate that the ways voice assistant users currently verbalise musical wishes can be reduced to relatively few dimensions. The musical identifiers of song title and artist are arguably unsurprising, and the presence of genre corroborates previous research on the crucial role of such terms for expressing listeners' cognitive schemes (Shevy, 2008), as a means to communicate musical and extra-musical meaning, and as a central element of music preferences (e.g., Rentfrow & Gosling, 2003). The presence of music listening function commands without such musical details (i.e., function of music listening like relaxation) suggests that listening functions may be especially important in the absence of other knowledge about the music that could be played (e.g., genre or artist), such as when a user generates an initial voice command without first reviewing options as would be done with a visual interface (i.e., if recognition is not supported and recall is difficult, discussed further below).

The presence of commands about playlists and personal collections aligns with their increasing

role as a means to organise music (e.g., Cunningham et al., 2006; Krause & North, 2014) particularly around an overall topic (e.g., ‘*Alexa, play playlist ‘Couch’*”). Notably, although participants were asked to provide commands for selecting music with voice assistants, rather than particular personal information management actions like keeping or refinding, the playlist and personal collection commands nonetheless suggest particular personal information management activities: playlist commands like ‘*Play my work playlist*’ indicate the prior creation of a playlist, and thus some keeping activities to create it, and commands about personal collections and favourite songs indicate some degree of meta-level (or organising) activity required to maintain and demarcate a collection and its parts (e.g., to designate a favourite). Indeed, playing music from one’s own collection might be an attempt to avoid neglecting or forgetting songs they like (Bergman et al., 2022b), or may be in recognition that collecting streamed songs can improve music listening enjoyment (Bergman et al., 2022a). Further, the use of playlists may be a response to (a) playlists being the only feature for doing personal music management in current streaming interfaces, or (b) there not being explicit opportunities when using voice assistants to collect or keep songs in a collection, as not collecting is the default there and in non-voice assistant streaming (Bergman et al., 2022a). In other words, some users may be attempting to overcome this default by collecting despite the relative lack of support for it. In this sense, streaming playlists are arguably overloaded, i.e., used for curating music around a topic (Cunningham et al., 2006) but also for simply collecting rather than losing (Sesigür, 2020).

The categories of commands vary regarding their specificity. Whereas song title represents a very specific musical desire with one possible retrieval result (notwithstanding different versions and interpretations of a musical piece), less specific categories like genre and function of music can lead to potentially millions of suitable songs. The most extreme case is the unspecific category, including commands in which musical properties do not seem to play a role and/or the control is completely

transferred to the voice assistant (e.g., ‘*Play a song!*’). Although such a command could be the positive outcome of a system becoming effectively personalised, the variety of listening situations and musical genres make it hard to predict its success. For example, chains of user commands (and corresponding prompts back to the user) have been found, outside of music retrieval, to lead to better task outcomes (Wu et al., 2022). Nonetheless, it may be easier for the user in particular moments than deriving a more specific request – such *satisficing* has been observed in non-voice assistant music retrieval (Bentley et al., 2006) – and so should be supported by voice assistants, and category specificity might be useful to consider when prompting a user, for example in asking what kind of music or what particular artist, as appropriate for the present situation.

The many hypothetical (i.e., not previously given) commands may indicate not only what users believe they can do with current systems, but generally how they think about and distinguish music and thus what they might do with an ideal system. For example, that music listening function was a relatively common command category may be an indication that non-voice assistant music software would benefit from supporting such functions beyond the ability to create playlists (e.g., translating a function like relaxation to a suggestion like ambient soundscapes). Further, three commands referred directly to properties of the music, such as its tempo (beats per minute), with one describing the timbre of desired song/album: ‘*Please play drones with metallic sound columns and warm pads in the background*’. That such commands were infrequent might be attributable to the difficulty of verbalising such data (i.e., putting sounds into words), which benefits from musical knowledge to articulate at least broad music demands (e.g., about the instrumentation or the harmonic complexity of a musical piece). Conversely, it could be because users have, over time, found that voice assistants are effective in translating their non-musical information to musical properties. Whether such commands are effective, and how they might be used to also organise music in voice user interfaces or visual interfaces, should be further explored.

Although streaming services encourage music discovery, very few commands reflected an interest to explore or listen to new music or build a music collection (i.e., finding and keeping in personal information management terms). The absence of such explicit commands is not a definitive indication that new music is not on voice assistant users' minds (Guest et al., 2011). Yet if the desire for new music exists, the commands nonetheless do not reflect it, and given the relatively light task of generating a few commands, participant fatigue seems unlikely to have precluded the provided commands from reflecting that desire. It is possible that 'play new music' is too broad to occur to users (only one usage in our study), but there were numerous equally broad commands in the unspecified category. More likely then is that users assume some command categories will provoke new music (e.g., commands for listening functions, wherein the artist is not specified), and/or some users lack the terminology to explore music (e.g. unknown genres). However, the lack of commands to play new music even from known artists (e.g., 'play the new single from artist X'; only one usage) suggests users are unaware of, unaccustomed to, or uninterested in such uses. If users ask voice assistants to play only the music they have already heard and remember or that suits their current activities, personal collections (even as playlists) will not be kept or grown (Steffens & Anglada-Tort, 2023); possible consequences – beyond a lack of exploration – are currently unclear but might include fatigue, stagnation or even polarisation of musical taste.

RQ4 To what extent can the kinds of commands be predicted by the music listening activities?

Here, our analyses have shown that a large amount of variance in the use of voice commands can already be attributed to stable, but so far unknown differences among individuals (i.e., as represented by the random intercept and indicated by the high intraclass correlation coefficient values > 90%), suggesting that participants had a strong tendency to use the same command category repeatedly and not leaving much room for situational variables (< 10%) such as the activity

at-hand. This finding might also indicate that humans have a mental category system of how to think about music in general which governs the way they organise their music collection and search for new music. For example, whereas person A might have the tendency to generally organise and search for music in terms of genre concepts, person B might do so with respect to emotions and semantics or associated listening contexts. This potential intersection of personal music management activity to mental structures should thus be the subject of future work, particularly as they may interact with similar individual differences like cognitive styles (e.g., 'analytic' organising by genre vs a 'wholist' use of listening context; Kozhevnikov, 2007) or personal information management styles like filing and piling (Henderson & Srinivasan, 2011).

Regardless of these highly individual command patterns, there were also trends in certain commands attributable to the music listening activity, such as requesting a playlist, genre, or artist when performing housework, requesting a particular playlist or artist when relaxing or falling asleep, to requesting a particular artist for social activities. For example, when performing housework, participants reported asking the voice assistant to play music from a particular playlist, genre, or artist rather than calling for a specific song title (as presented in Table 3). This finding might be interpreted in a way that, during certain activities such as housework, individuals call for multiple suitable songs at the same time to have their hands free for their activity at-hand (e.g., cooking, cleaning, sports exercise) over a longer period of time. Furthermore, when relaxing or falling asleep, participants showed a tendency to call for a playlist or an artist rather than asking for a particular song title. This again can be explained by the required duration of the planned activity (which is longer than one song) and also by the fact that there are already many ready-made playlists provided by digital music services which are designed to fulfil this particular listening goal.

Synthesis: voice assistants, music listening, and personal music management

Interaction with voice assistants currently depends entirely on (linguistic) recall rather than recognition. That users must know what commands are possible, applicable, and valid is a long-standing challenge of voice interaction (Yankelovich, 1996), but there are possible advantages to a recall-driven approach. For example, allowing users to play a desired song by naming it, rather than retrieving it among many potentially distracting items, emulates the benefits of search. This may be of particular benefit to groups like 'super searchers' and older users, who rely more on linguistic ability (which increases throughout life; Hartshorne & Germane, 2015) than on spatial ability (Benn et al., 2015) and thus use file search much more than navigation to retrieve desktop files (Bergman et al., 2019; Pak, 2001). However, recognition has been shown to be an important driver of music selection behaviour in non-voice assistant contexts (Steffens & Anglada-Tort, 2023), and interface features and design could further support it. Doing so may make the interfaces more useful to all users and more accessible to particular users. For example, the shift from line-based to full-screen text editors drastically reduced differences in participant performance across age groups when storing and modifying documents, perhaps because displaying more of a document reduced the demands on spatial capacity and working memory (Gomez et al., 1983). Voice assistants too could do more to compensate for such demands, whether caused by cognitive decline or development yet to come, for example by reading out options in an organised way (e.g., with numbered options like a phone menu so that complex options need not be verbally retained and repeated).

However, the reliance on recall arguably loses some of the benefits of recognition that are found in traditional personal music management, like browsing genres for inspiration or to be reminded about beloved artists. Similarly, remembering even an extremely positive musical experience apropos of nothing may be difficult, especially as music

is consumed at such a high volume today. Serendipity is likely achieved by music discovery embedded in streaming services, but further support for and personalisation of reminding may help to replicate the useful effects of browsing. This might be achieved for example by suggesting artists or genres that were listened to many times but not recently. To further improve voice assistants support for recognition, the categories observed here could be codified into the system, for example to support recognition by making suggestions to a user in order of category frequency (i.e., asking if the user wants a particular song, music for a particular function, a particular playlist...). To our knowledge, current systems (e.g., Spotify) still effectively select pre-made playlists by reviewing their titles (e.g., 'driving tunes'), and so other musical data could be used to generate suitable playlists as they are requested. To match music to a listening function, a voice assistant could solicit a description ('what are you doing?') and then translate the activity to a function or query to retrieve a suitable playlist or offer options at the suitable level of specificity (e.g., name songs, artists, genres...).

Above we have observed the user commands and situations that exist in voice-assistant powered music listening and commented on outstanding opportunities and limitations particularly with regards to music listening and management. While voice assistants have some clear benefits for music listening and personal music management – benefits that users are already taking advantage of with their commands – they also seem to currently lack support for activities that are common when managing music and personal information, which may lead to a lack of new music exploration, collection building and maintenance, sharing, and so on. It seems that until the benefits of recall and recognition are well integrated into the interactions with voice-based systems, important aspects of personal music listening and management are rather constrained and their full benefit is unrealised.

Limitations

There are numerous important factors in broad voice assistant use that could not be explored in this study (i.e., limitations of scope). First,

technical and contextual factors may influence the use of voice assistants for music and other activities, such as the device type, available interaction modes, device location, and user's culture. For example, while smartphones are omnipresent, home assistants may encourage different interactions, especially in social contexts of different kinds that not examined here (O'Hara & Brown, 2006), and whereas it is our impression that voice assistant use (including for music playback) has become common in the US, it remains relatively uncommon in e.g., Germany. To understand voice assistants more broadly would thus require exploring such factors, which were beyond the scope of this study.

Second, we solicited commands to play music generally, rather than asking about particular personal information management actions like copying a song or organising a playlist. Although the term play is appropriate to music and our exploratory study of voice assistants, its use limits the breadth of conclusions that can be drawn in the present study. It is likely that future studies asking about particular tasks would produce fine-grained data about those tasks, and we therefore encourage future investigations of more specific voice assistant-powered personal information management tasks and in formats beyond music (e.g., sharing photos; Bentley et al., 2006). Third, it is likely that voice assistant users give different voice commands when *prompted* (as in our study) than during independent use of voice assistants, and that speculative commands differ from those seen during *in situ* use. It is not clear to what extent the commands might differ, but it is reasonable to assume they will. Thus, to facilitate ecological validity of data in future studies, it is recommended to collect given commands with methods like logging (Chernov et al., 2008) or experience sampling (Greb et al., 2019), which was not possible in the course of the present study.

Further, it is possible that participant familiarity with different voice assistants could change the commands our participants suggested, for example because different voice assistants have different features or affordances (Brüggemeier et al., 2020). Although our intuition is that this

has little to no effect – we have the impression current voice assistants are fairly similar – a comparison of the collected commands across participant groups using different voice assistants would reveal any possible differences.

Finally, it is important to note that only 38% of the reported voice commands were reported to come from a real situation, whereas the remainder were hypothetical commands users *would* give a voice assistant to select music. Since only an eighth of the respondents reported using a voice assistant on a daily basis, the findings above should be interpreted cautiously: because voice assistants are becoming increasingly popular (Tiwari et al., 2020), the data of this study may not reflect current or later use of voice assistants. Nonetheless, the relative consistency in commands – between participants who had used voice assistants and those who had not – suggests the hypothetical commands may also be realistic.

Conclusion

The command categories and listening situations identified in this study suggest voice assistants are promising for music listening and for personal music management, but with some limitations and risks. Although they enable interactions that can benefit particular situations and user groups, voice assistants offer limited support for personal information management actions like keeping and browsing, particularly because of the affordances and features of current voice user interfaces, the situations in which it is desirable to use them, and perhaps users' awareness of the possible commands. From this initial perspective on voice assistants for music listening and personal music management we have made suggestions for how the systems might be improved, especially to better provide the benefits of browsing.

Additionally, this study contributes to the literature on personal music management by considering an emerging mode of interaction and integrating perspectives from music psychology. It is a starting point for further research on the use of voice assistants in

personal information management and among few studies of voice-based music listening and management. Despite the variety and popularity of mobile devices and ubiquity of personal information management, there has been little research into mobile personal information management in comparison to stationary computing environments (Dinneen & Julien, 2020). Due to the limitations of mobile display sizes and situations in which the use of a visual interface is less appropriate, the use of voice assistants could offer new possibilities to interact with personal collections. We hope that the personal information management

research community will take up an exploration of these possibilities.

Finally, from a music-psychological perspective, the study contributes to the understanding of how listeners mentally categorise music when visual recognition is not available and how they use music and music technology in different situational contexts. Here, future research might look into person-related (e.g., musical sophistication, openness) and further situational variables (e.g., mood, presence of other people) associated with the usage of commands (e.g., their specificity) and the expected musical outcomes.

About the authors

Jochen Steffens is a professor of musical acoustics and Director of the Institute of Sound and Vibration Engineering at Hochschule Düsseldorf, Germany. He is interested in how we listen to sound and music, how we can predict and understand music listening behaviour, and how listening affects our well-being as well as our emotions, perceptions, and behaviour. He can be contacted at jochen.steffens@hs-duesseldorf.de.

Jesse David Dinneen is a Junior Professor in the Berlin School of Library and Information Science at Humboldt-Universität zu Berlin, where he researches personal information management and information ethics. He can be contacted at jesse.dinneen@hu-berlin.de.

Sascha Donner is a PhD student at the Berlin School of Library and Information Science at Humboldt-Universität zu Berlin. He is interested in how human-AI collaboration will affect human information processing, communication, and knowledge.

Tom Potthoff is a graduate of the Audio Communication Group at Technische Universität Berlin, Germany.

References

Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, search, and IoT: how people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction*, 26(3), 1–28.

<https://doi.org/10.1145/3311956>

Azzopardi, L., Aliannejadi, M., & Kanoulas, E. (2022). Towards building economic models of conversational search. arXiv preprint arXiv:2201.08742.

<https://doi.org/10.48550/arXiv.2201.08742>

Bainbridge, D., Cunningham, S. J., & Downie, J. S. (2003). *How people describe their music information needs: a grounded theory analysis of music queries*. Paper presented at Fourth International Conference on Music Information Retrieval (ISMIR 2003). (Archived by ResearchGate at https://www.researchgate.net/profile/David-Bainbridge/publication/220723261_Analysis_of_queries_to_a_Wizard-of-Oz_MIR_system_Challenging_assumptions_about_what_people_really_want/links/00b7d52)

[9717a75a552000000/Analysis-of-queries-to-a-Wizard-of-Oz-MIR-system-Challenging-assumptions-about-what-people-really-want.pdf](https://doi.org/10.1002/97811187552000000/Analysis-of-queries-to-a-Wizard-of-Oz-MIR-system-Challenging-assumptions-about-what-people-really-want.pdf)

Barrington, L., Oda, R., & Lanckriet, G. R. (2009). Smarter than genius? Human evaluation of music recommender systems. Paper presented at 10th International Society for Music Information Retrieval Conference (ISMIR 2009), Kobe, Japan. (Archived by ISMIR Archives at <https://archives.ismir.net/ismir2009/paper/000014.pdf>)

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>

Beirl, D., Rogers, Y., Yuill, N. (2019) *Using voice assistant skills in family life*. Paper presented at 13th International Conference on Computer Supported Collaborative Learning - A Wide Lens: combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, CSCL 2019. Lyon, France. (Archived by ResearchGate at https://www.researchgate.net/profile/Oleksandra-Baga/publication/363539397_Home_Digital_Voice_Assistants_use_cases_and_vulnerabilities/links/6321e85a70cc936cd309bebf/Home-Digital-Voice-Assistants-use-cases-and-vulnerabilities.pdf)

Benn, Y., Bergman, O., Glazer, L., Arent, P., Wilkinson, I. D., Varley, R., & Whittaker, S. (2015). Navigating through digital folders uses the same brain structures as real world navigation. *Scientific Reports*, 5(1), 14719.

Bentley, F., Metcalf, C., & Harboe, G. (2006). Personal vs. commercial content: the similarities between consumer use of photos and music. In R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries & G. Olson (Eds.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 667-676). <https://doi.org/10.1145/1124772.1124871>

Bergman, O., Israeli, T., & Whittaker, S. (2019). Search is the future? The young search less for files. *Proceedings of the Association for Information Science and Technology*, 56(1), 360-363. <https://doi.org/10.1002/pra2.29>

Bergman, O., & Whittaker, S. (2016). *The science of managing our digital stuff*. Mit Press. <https://doi.org/10.17723/0360-9081-81.1.233>

Bergman, O., Whittaker, S., & Gradovitch, N. (2022a). Evidence for the merits of collecting streaming music. *Personal and Ubiquitous Computing*, 1-12. <https://doi.org/10.1007/s00779-022-01692-y>

Bergman, O., Whittaker, S., & Tish, G. (2022b). Collecting music in the streaming age. *Personal and Ubiquitous Computing*, 26(1), 121-129. <https://doi.org/10.1007/s00779-021-01593-6>

Brinegar, J., & Capra, R. (2010). Understanding personal digital music collections. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-2. <https://doi.org/10.1002/meet.14504701378>

Brinegar, J., & Capra, R. (2011). Managing music across multiple devices and computers. In *Proceedings of the 2011 iConference* (pp. 489-495). <https://doi.org/10.1002/meet.14504701378>

Brüggemeier, B., Breiter, M., Kurz, M. & Schiwy, J. (2020). User experience of Alexa, Siri and Google Assistant when controlling music – comparison of four questionnaires. In C. Stephanidis, A. Marcus, E. Rosenzweig, P. P. Rau, A. Moallem & M. Rauterberg, (Eds.), *HCI International 2020 - Late Breaking Papers: User Experience Design and Case Studies* (S. 600-618). Springer International Publishing. https://doi.org/10.1007/978-3-030-60114-0_40

- Chernov, S., Demartini, G., Herder, E., Kopycki, M., & Nejd, W. (2008, April). *Evaluating personal information management using an activity logs enriched desktop dataset*. Paper presented at 3rd Personal Information Management Workshop, Florence, Italy. (Archived by ResearchGate at https://www.researchgate.net/profile/Gianluca-Demartini/publication/251546606_Evaluating_Personal_Information_Management_Using_an_Activity_Logs_Enriched_Desktop_Dataset/links/00b7d52a1a36656f25000000/Evaluating-Personal-Information-Management-Using-an-Activity-Logs-Enriched-Desktop-Dataset.pdf)
- Cunningham, S. J., Bainbridge, D., & Falconer, A. (2006). "More of an art than a science": supporting the creation of playlists and mixes. Paper presented at 7th Conference of the International Society of Music Information Retrieval (ISMIR), Victoria, Canada. (Archived by [academia.eu](https://www.academia.edu) at <https://www.academia.edu/download/44050057/content.pdf>)
- Cunningham, S. J., Jones, S., and Jones, M. (2004). *Organizing digital music for use: an Examination of personal music collections*. Paper presented at 5th International Symposium on Music Information Retrieval. Archived by the Internet Archive at <https://web.archive.org/web/20240524124423/https://www.ee.columbia.edu/~dpwe/ismir2004/CRFILES/paper221.pdf>
- Dinneen, J. D., & Julien, C.-A. (2019). What's in people's digital file collections? *Proceedings of the Association for Information Science and Technology*, 56(1), 68-77. <https://doi.org/10.1002/pr2.64>
- Dinneen, J. D., & Julien, C.-A. (2020). The ubiquitous digital file: A review of file management research. *Journal of the Association for Information Science and Technology*, 71(1), E1-E32. <https://doi.org/10.1002/asi.24222>
- Dubiel, M., Halvey, M., & Azzopardi, L. (2018). A survey investigating usage of virtual personal assistants. *arXiv preprint arXiv:1807.04606*. <https://doi.org/10.48550/arXiv.1807.04606>
- Fields, B. (2011). *Contextualize your listening: the playlist as recommendation engine* (PhD thesis). Goldsmiths College (University of London). (Archived by ResearchGate at https://www.researchgate.net/profile/Benjamin-Fields-4/publication/266271406_Contextualize_Your_Listening_The_Playlist_as_Recommendation_Engine/links/561e2d5108aef097132b32bc/Contextualize-Your-Listening-The-Playlist-as-Recommendation-Engine.pdf)
- Fields, B., Lamere, P., & Hornby, N. (2010). Finding a path through the juke box: the playlist tutorial. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, Netherlands. (Archived by Internet Archive at https://archive.org/details/field_juke_box)
- Furqan, A., Myers, C., & Zhu, J. (2017). Learnability through adaptive discovery tools in voice user interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1617-1623). <https://doi.org/10.1145/3027063.3053166>
- Gomez, L. M., Egan, D. E., Wheeler, E. A., Sharma, D. K., & Gruchacz, A. M. (1983, December). How interface design determines who has difficulty learning to use a text editor. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 176-181). <https://doi.org/10.1145/800045.801605>
- Greb, F., Schlotz, W., & Steffens, J. (2018). Personal and situational influences on the functions of music listening. *Psychology of Music*, 46(6), 763-794. <https://doi.org/10.1177/0305735617724883>

- Greb, F., Steffens, J., & Schlotz, W. (2019). Modeling music-selection behavior in everyday life: a multilevel statistical learning approach and mediation analysis of experience sampling data. *Frontiers in Psychology*, 10, 390. <https://doi.org/10.3389/fpsyg.2019.00390>
- Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied Thematic Analysis*. Sage.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* 33, 1-22. <https://doi.org/10.18637/jss.v033.i02>
- Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, 26(4), 433-443. <https://doi.org/10.1177/0956797614567339>
- Henderson, S., & Srinivasan, A. (2011, January). Filing, piling & structuring: strategies for personal document management. In 2011 44th Hawaii International Conference on System Sciences (pp. 1-10). IEEE. <https://doi.org/10.1109/HICSS.2011.205>
- Jones, W., Dinneen, J. D., Capra, R., Diekema, A., & Pérez-Quiñones, M. (2017). Personal information management (PIM). In J. D. McDonald & M. Levine-Clark (Eds.) *Encyclopedia of Library and Information Science* (4th ed., pp. 3584-3605). CRC Press. (Archived by asXiv at <https://arxiv.org/pdf/2107.03291>)
- Kamalzadeh, M., Baur, D., & Möller, T. (2012). A survey on music listening and management behaviours. In *Proceedings of the 13th International Society for Music Information Retrieval (ISMIR) Conference*, Porto, Portugal. (Archived by academic.edu at <https://www.academia.edu/download/67744370/ismir12.pdf>)
- Khaokaew, Y., Holcombe-James, I., Rahaman, M. S., Liono, J., Trippas, J. R., Spina, D., ... & Salim, F. D. (2022). Imagining future digital assistants at work: a study of task management needs. *arXiv preprint arXiv:2208.03443*. <https://doi.org/10.1016/j.ijhcs.2022.102905>
- Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style. *Psychological Bulletin*, 133(3), 464-481. <https://doi.org/10.1037/0033-2909.133.3.464>
- Krause, A. E., & North, A. C. (2014). Contextualized music listening: playlists and the Mehrabian and Russell model. *Psychology of Well-Being: Theory, Research and Practice*, 4(22), 1-16. <https://doi.org/10.1186/s13612-014-0022-7>
- Krause, A., North, A., & Hewitt, L. (2014). Music selection behaviors in everyday listening. *Journal of Broadcasting & Electronic Media*, 58(2), 306-323. <https://doi.org/10.1080/08838151.2014.906437>
- Lanctot, A., & Duxbury, L. (2021). When everything is urgent! Mail use and employee well-being. *Computers in Human Behavior Reports*, 4, 100152. <https://doi.org/10.1016/j.chbr.2021.100152>
- Lee, C.-H., Soong, F. K., & Paliwal, K. K. (2012). *Automatic speech and speaker recognition: advanced topics*. Springer Science & Business Media. Archived by the Internet Archive at <https://web.archive.org/web/20170914122733/https://link.springer.com/content/pdf/bfm%3A978-1-4613-1367-0%2F1.pdf>
- Lee, J. H., & Waterman, N. M. (2012, October). Understanding user requirements for music information services. In *International Society for Music Information Retrieval Conference (ISMIR 2012)* (pp. 253-258). Archived by the Internet Archive at

<https://web.archive.org/web/20220131234014/https://zenodo.org/record/1417625/files/LeeW12.pdf>

Lemström, K. & Tzanetakis, G. (2017). Music information retrieval. In J.D. McDonald & M. Levine-Clark (Eds.) *Encyclopedia of library and information science*. CRC Press.
<https://doi.org/10.1081/E-ELIS4-120043656>

Luger, E., & Sellen, A. (2016, May). "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* (pp. 5286-5297).
<https://doi.org/10.1145/2858036.2858288>

Myers, C. M. (2019). Adaptive suggestions to increase learnability for voice user interfaces. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion* (pp. 159-160). <https://doi.org/10.1145/3308557.3308727>

Myers, C. M., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J. (2018). Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-7). <https://doi.org/10.1145/3173574.3173580>

Myers, C. M., Furqan, A., & Zhu, J. (2019a). The impact of user characteristics and preferences on performance with an unfamiliar voice user interface. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-9). <https://doi.org/10.1145/3290605.3300277>

Myers, C. M., Grethlein, D., Furqan, A., Ontañón, S., & Zhu, J. (2019b). Modeling behavior patterns with an unfamiliar voice user interface. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 196-200).
<https://doi.org/10.1145/3320435.3320475>

Myers, C. M., Laris Pardo, L. F., Acosta-Ruiz, A., Canossa, A., & Zhu, J. (2021). "Try, try, try again:" sequence analysis of user interaction data with a voice user interface. In *CUI 2021-3rd Conference on Conversational User Interfaces* (pp. 1-8). <https://doi.org/10.1145/3469595.3469613>

O'Hara, K., & Brown, B. (Eds.). (2006). *Consuming music together: social and collaborative aspects of music consumption technologies* (Vol. 35). Springer Science & Business Media. (Archived by the academia.eu at https://www.academia.edu/download/46877669/Distributing_the_Process_of_Music_Choice20160629-18002-vmoh5g.pdf#page=91)

Pak, R. (2001, October). A further examination of the influence of spatial abilities on computer task performance in younger and older adults. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45(22), pp. 1551-1555.
<https://doi.org/10.1177/154193120104502203>

Pontara, T., & Volgsten, U. (2017). Musicalization and mediatization. In O. Driessens, G. Bolin, A. Hepp, & S. Hjarvard (Eds.), *Transforming communications - studies in cross-media research. dynamics of mediatization: institutional change and everyday transformations in a digital age* (Vol. 20, pp. 247-269). Springer International Publishing. https://doi.org/10.1007/978-3-319-62983-4_12

R Core Team (2015). *R: A language and environment for statistical computing*. Archived by the Internet Archive at <https://web.archive.org/web/20240503151610/https://www.r-project.org/>

- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, 84(6), 1236–1256. <https://doi.org/10.1037/0022-3514.84.6.1236>
- RStudio Team (2015). RStudio: *Integrated development for R* [Computer software]. Archived by the Internet Archive at <https://web.archive.org/web/20240503151352/https://posit.co/products/open-source/rstudio/>
- Sease, R. and McDonald, D. W. (2009). Musical fingerprints: collaboration around home media collections. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. ACM, New York, NY, 331-340. <https://doi.org/10.1145/1531674.1531724>
- Sesigür O (2020) How to approach collecting music on streaming services. *Interactions: Studies in Communication & Culture* 11(1), 65–74. https://doi.org/10.1386/iscc_00006_1
- Shevy, M. (2008). Music genre as cognitive schema: extramusical associations with country and hip-hop music. *Psychology of Music*, 36(4), 477–498. <https://doi.org/10.1177/0305735608089384>
- Steffens, J., & Anglada-Tort, M. (2023). The effect of visual recognition on listener choices when searching for music in playlists. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication: <https://doi.org/10.1037/aca0000562>.
- Stone, M. (2022). Understanding and evaluating search experience. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 14(1), 1-105. <https://doi.org/10.1007/978-3-031-79216-8>
- Tiwari, V., Hashmi, M. F., Keskar, A., & Shivaprakash, N. C. (2020). Virtual home assistant for voice based controlling and scheduling with short speech speaker identification. *Multimedia Tools and Applications*, 79(7-8), 5243–5268. <https://doi.org/10.1007/s11042-018-6358-x>
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293–302. <https://doi.org/10.1109/TSA.2002.800560>
- Wirfs-Brock, J., Mennicken, S., & Thom, J. (2020, April). Giving voice to silent data: designing with personal music listening history. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-11). <https://doi.org/10.1145/3313831.3376493>
- Wu, T., Terry, M., & Cai, C. J. (2022, April). Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *CHI Conference on Human Factors in Computing Systems* (pp. 1-22). <https://doi.org/10.1145/3491102.3517582>
- Yankelovich, N. (1996). How do users know what to say? *Interactions*, 3(6), 32-43. Archived by the Internet Archive at <https://web.archive.org/web/20240524124104/https://www.carl.angiolillo.net/portfolio/projects/MessengerMail/files/readings/Yankelovich%20-%20How%20do%20Users%20Know%20What%20to%20Say.pdf>