Expectation Conformance in Online Sound Therapy: Designing Tools for Users of Mental Wellbeing Applications

Arjun Arunasalam
Purdue University
West Lafayette, IN, USA
aarunasa@purdue.edu

Jason Tong
Purdue University
West Lafayette, IN, USA
tong72@purdue.edu

Habiba Farrukh
University of California, Irvine
Irvine, CA, USA
habibaf@uci.edu

Muslum Ozgur Ozmen
Arizona State University
Tempe, AZ, USA
ozmen@asu.edu

Koustuv Saha
University of Illinois
Urbana-Champaign
Urbana, IL, USA
ksaha2@illinois.edu

Z. Berkay Celik
Purdue University
West Lafayette, IN, USA
zcelik@purdue.edu

Abstract

Mental wellbeing technologies are prevalent in digital spaces, such as content creation websites (e.g., YouTube) and mobile apps. Many users leverage such technologies and thus develop expectations for what they should provide. However, tools to verify whether these technologies conform to user expectations remain largely unexplored. We investigate this problem in the domain of binaural beats - a popular mental wellbeing technology. Using results from preliminary research, where we establish commonly held user expectations and introduce a method to measure expectation conformance, we explore the design of a tool to be used by binaural beats listeners. Through a pre-design survey with 43 participants and a user study with 28 participants, we demonstrate how tools leveraging commonly held user expectations can help users of mental wellbeing applications make informed decisions.

CCS Concepts

• **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods*; *Empirical studies in collaborative and social computing*; *Social media*;

Keywords

Mental Wellbeing; User Expectations; Online Content

ACM Reference Format:
Arjun Arunasalam, Jason Tong, Habiba Farrukh, Muslum Ozgur Ozmen, Koustuv Saha, and Z. Berkay Celik. 2024. Expectation Conformance in Online Sound Therapy: Designing Tools for Users of Mental Wellbeing Applications . In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24), November 9–13, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3678884.3681903

CSCW Companion '24, November 9–13, 2024, San Jose, Costa Rica

ACM ISBN 979-8-4007-1114-5/24/11
https://doi.org/10.1145/3678884.3681903

1 Introduction

Mental wellbeing applications have emerged in online spaces in various forms. These applications are diverse, ranging from guided meditative applications to calming tones uploaded to websites such as YouTube. Given that online mental wellbeing content has grown in digital spaces, end-users are often exposed to such content. Consequently, end-users consume this mental wellbeing content and develop expectations surrounding what this content should deliver.

Prior works in online mental wellbeing are diverse. One line of work has investigated how users leverage digital spaces to improve their online mental wellbeing [1, 10, 22]. Another research line evaluates the impact of wellbeing content (e.g., meditation apps) on users via methods such as expert evaluation and autoethnography [6, 11, 20]. Despite this, there is a lack of research on user expectations surrounding mental wellbeing technology. Additionally, tools to inform users when tracks do not conform to commonly held user expectations are largely absent.

The aforementioned research gap motivates our exploration into designing tools for end users. We focus on binaural beats, an emerging form of mental wellbeing technology that relies on two tones of differing frequencies, played in the left and right ear to produce a third, perceived frequency. Through preliminary research (detailed in Section 2.2), we (1) discover what experienced binaural beats listeners' expectations are for binaural beats and (2) that users have concerns that online binaural content does not adhere to expectations. We also design a method to verify whether a track conforms to commonly held expectations. Given users are concerned whether tracks conform to expectations, it is fair to assume users would benefit from being able to verify a track. Thus, we sought to explore how users can perform such verification.

One method to allow end-users to verify wellbeing applications would be through a mental wellbeing tool, designed for users who consume binaural beats content. However, such tools need to have widespread use and should not just cater to experienced binaural listeners. Thus, we focus on designing a tool to cater to inexperienced listeners - individuals who have yet to start/just started listening to binaural beats. Consequently, we focus on answering the following research question:

RQ: How effective is a tool that communicates commonly held expectations in aiding inexperienced listeners?

To answer this research question, we first conduct an exploratory pre-design survey. In this survey, we assess the potential benefit of designing a tool that caters to inexperienced binaural beats listeners. After surveying 43 participants, we determined that inexperienced listeners display information-seeking behaviour online for mental wellbeing technology and that they found commonly held expectations surrounding binaural beats believable.

Our exploratory results motivate us to encapsulate our methodology for verifying expectation conformance in a UI tool we call Apollo. We conduct a study with 28 participants, showing them binaural tracks with and without Apollo's output. We find that Apollo's output causes a significant difference in how likely users are to listen to a track, demonstrating it's ability to guide users to make informed decisions.

Our work presents broader implications for the HCI and CSCW community when dealing with mental wellbeing applications. First, we highlight the need for mental wellbeing stakeholders to carefully design tool-based interventions. Second, we showcase the potential of using expectations of experienced users to ground design of tools for inexperienced users. Finally, we recommend future work should explore the applicability of our methods and tool design to other mental wellbeing applications outside the domain of binaural beats.

2 Background

Our research question builds on (1) existing work on the binaural beats phenomenon and (2) preliminary research we conducted on users' expectations for binaural beats.

2.1 Binaural Beats

Binaural beats are an emerging mental wellbeing technology. They refer to an auditory illusion where two tones of different frequencies are played in a listener's left and right ear [17]. The brain perceives a net frequency between the two tones, known as the binaural beat. For example, if the left ear receives a 180Hz tone and the right ear receives a 170Hz tone, the brain absorbs a 10Hz tone. Exposure to this third frequency causes the brain's neurons to emit the same frequency. This synchronization of brain frequencies to this sound is known as entrainment [3, 21]. The entrainment effect has resulted in binaural beats' emergence as a mental wellbeing technology, with users believing entrainment towards specific frequencies can bring about different mood enhancements and induce mental states[1].

2.2 Preliminary Research and Motivation

Our efforts to answer our research question are motivated by preliminary research on binaural beats and users' expectations. First, we crawl online data to scope what commonly held user expectations experienced binaural beats listeners have for binaural beats. Second, we devise a method to ascertain whether a binaural beats track conforms to these expectations [2].

Understanding User Expectations. In an inductive analysis of over 2K URLs pointing towards online discussion forums or blogs, we discovered that experienced binaural beats listeners have established expectations on (1) what goals they expect binaural beats can achieve for them and (2) what specific binaural beats are tied to these goals, as noted in Table 1 below.

Table 1: Listeners' goals and binaural beats expected.

Goal	User Expectations
`Sleep`	A single period of delta in track's duration OR Sleep cycle frequencies
`Relaxation`	A single period of alpha OR theta in track's duration
`Concentration`	A single period of beta in track's duration
`Cognitive Enhancement`	A single period of gamma in track's duration

Users have four mental state goals: sleep, relaxation, concentration and cognitive enhancement. Each goal has expectation(s) on the types of brainwaves. We note that named brainwaves in Table 1 have associated frequencies, which users also specify. We also discovered that users have concerns about the quality of the binaural beats tracks they listen to - specifically, whether the tracks adhere to their expectations.

Evaluating Track Conformance. After understanding what users expect, we synthesize a method to verify expectation conformance: whether a given track conforms to commonly held user expectations. First, we extract the specific mental state a track is advertising, from track metadata. The track is classified into one of the six classes (sleep, relaxation, concentration, cognitive enhancement, multi-label, and outlier) through lexicons constructed for each mental state. Relaxation, sleep, concentration, and cognitive enhancement are mental states accepted among binaural beats listeners. Multi-labeled denotes the tracks classified into more than one mental state, and outlier indicates the tracks with unclear/vague intents.

Second, we performed a spectral analysis to extract the time-frequency model of the binaural beats track. We develop a process that combines audio signal processing techniques while addressing spectral leakage and frequency precision challenges to accurately capture dominant frequencies in the left and right audio channels. This allows us to obtain a time-frequency model, where frequency represents the binaural beats produced.

Third, we use five detection rules designed to verify whether a track does not conform to expectations. If any of these rules are flagged, we deem the track a *deceptive* track. These rules operate on a track's intent and time-frequency model and are as follows:

- **F1 Goal Deviation:** When a track does not advertise a user expected intent (outlier).
- **F2 Goal Conflict:** When a track claims to achieve multiple goals (multi-labeled)[2].
- **F3 Deceptive Absence:** When a track does not contain binaural beats.
- **F4 Goal-Expectation Negligence:** When a track's binaural beats do not match the advertised intent.
- **F5 Goal-Expectation Contradiction:** When a track's binaural beats match a contradicting intent.

2.3 Research Goals

Given that users have concerns about the binaural tracks they listen to, we now focus on our expectation conformance methodology's applicability. Thus, our research goal becomes to ascertain if a tool

[1]We do not claim to evaluate the efficacy of binaural beats as mood enhancers. Our research is grounded on user expectations and focuses on discrepancies between user expectations and what binaural beats are present in a track.

[2]Tracks inducing more than one state is not supported by our user expectation findings, and thus multi-label is indicative of a deceptive track.

encapsulating such methodology is useful for end users who listen to binaural beats. However, such tools should be designed to cater to the broad public - a binaural beats listener does not just include experienced listeners but inexperienced listeners as well. We take an exploratory approach to investigate: **how effective is a tool that communicates commonly held expectations in aiding inexperienced listeners?**

First, in a pre-design survey, we conduct a preliminary investigation to assess the potential benefit of such a tool. After assessing this benefit, we encapsulate our expectation conformance methodology in a UI tool and conduct a user study. We expose users to binaural beats tracks, providing them with and without our prototype UI design to determine the influence of commonly held user expectations on their decisions.

3 Pre-Design Survey

Our pre-design survey aims to develop a preliminary understanding of whether a tool leveraging experienced listeners commonly held expectations would help inexperienced listeners. We note that we take a proxy approach to evaluate if users would find such a tool helpful, grounded on the following two assumptions:

- Users should demonstrate information-seeking behaviour via online forums for mental wellbeing technology.
  - In our preliminary research, we extract experienced listeners' expectations from online resources. If inexperienced listeners are likely to consume such online resources, it is reasonable to assume that they are exposed to commonly held expectations.
- Users should find commonly held expectations believable.
  - Once exposed to common expectations, if users find them believable (e.g., a track with specific brainwaves can achieve a specific goal) there exists potential for such a tool to help them make informed choices.

3.1 Survey Overview and Participants

Our survey comprised five sections. First, we asked preliminary questions about the use of mental wellbeing and binaural beats applications. Second, we provided users with an introduction to binaural beats. Third, we probe how users are likely to consume information on mental wellbeing technology. Fourth, we displayed images of binaural beats tracks found on YouTube. Specifically, we selected tracks that promised one of the four discovered mental state goals. For each track, we asked users if they believed the stated goal could be achieved via binaural beats. Fifth, we probe users for what information surrounding a binaural beats track would be useful to them. Our survey was deemed exempt by our IRB.

We advertised our study through internal campus email listings and Slack channels to recruit participants who are not experienced binaural beats listeners. Participants were required to be at least 18 years old and have access to a computer to complete the survey online. In total, 43 participants qualified for our study. These participants included 30 males, 10 females, and three non-binary individuals ranging from 18 to 55 years old.

3.2 Survey Results

All participants were inexperienced listeners. They had self-reported not having listened to binaural beats before or less than five times. 14 participants recorded being aware of binaural beats. These participants mentioned having heard of binaural beats *"[from] colleagues"* or *"from reading [on] the web."* 64.2% of them had listened to tracks on platforms such as YouTube or Spotify. The remaining 29 participants had never heard of binaural beats before. When asked how likely they were to learn about them from forums/blogs, 82% of users stated that information on forums/blogs helps them understand binaural beats. Further, 82% of these users stated they would be inclined to listen to binaural beats if exposed to accounts of positive experiences on forums and blogs.

83.9% of participants believed that binaural beats with delta waves could help with sleep. 80.6% believed binaural beats would help them relax, while 77.4% stated that the presence of theta and alpha waves would accomplish this. For concentration and cognitive enhancement, 77.4% and 67.5% of participants believed binaural beats were able to achieve these goals, respectively.

We found that 61.0% of participants find that the title of a binaural beats track would influence their decision to listen to it, with users noting *"the purpose of the binaural beat track"* and *"the resulting effect of the track, the type of binaural beats used"* as criteria for an informative title. 71.0% and 67.7% of participants acknowledged that knowing the beats present in the track and their association with a specific mental state goal would influence their decision.

Takeaway:
- Individuals unfamiliar with binaural beats are likely to learn about them from online forums and blogs.
- The four mental state goals are accepted as believable among inexperienced listeners.
- Participants find track titles informative, but access to information about binaural beats produced in a track and whether they are associated with a mental state goal would influence their decision to listen to the track.

4 Impact of a Mental Wellbeing Tool on Users

Our pre-design survey demonstrates that a tool to aid individuals who are not experienced listeners has the potential to be effective. Thus, we encapsulate our expectation conformance methodology into a UI tool we refer to as Apollo, as overviewed in Figure 1. Apollo accepts a track as input, conducts track metadata analysis to extract a track's intent (❶) and spectral analysis to produce the time-frequency model (❷). Then, Apollo verifies conformance via our detection rules (❸). Apollo outputs results of expectation conformance via the UI, an example of which is seen in Figure 2.

We then investigate how helpful Apollo is in informing users about the deception of a binaural beats track compared to track metadata on various online/mobile streaming platforms.

4.1 Participants Recruitment

We advertised our study through internal campus email listings and Slack channels. We diversified our survey exposure via snowball sampling (participants forwarded our contact information to interested individuals) and non-university-related public channels (e.g., social groups). Participants were required to be at least 18 years of age and have access to a computer to complete the online

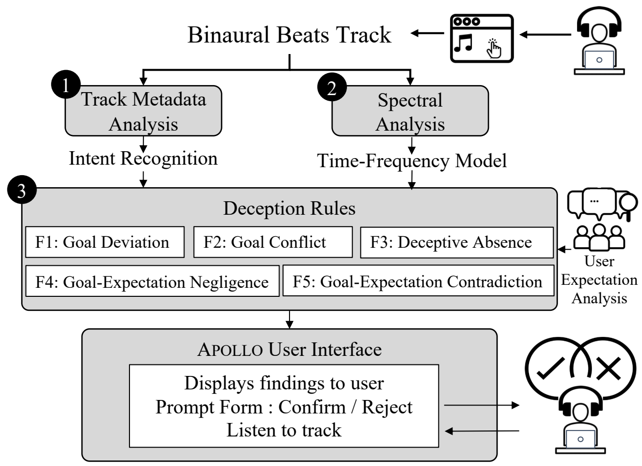

Figure 1: Architecture of Apollo: a binaural beats verification tool to function as an informative conduit between users and audio streaming providers.

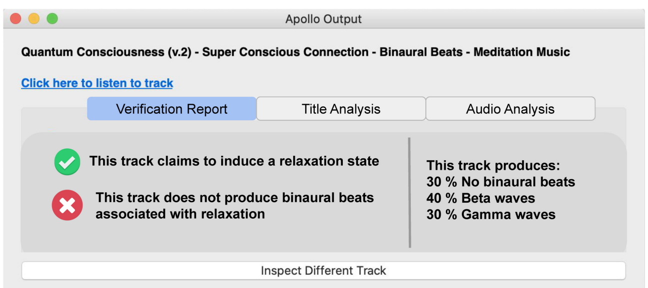

Figure 2: Apollo's UI which is communicated to the user to request approval or denial of listening to the track.

survey. 28 participants qualified for our study, a statistically significant sample size to study Apollo's effectiveness [14]. Participants comprised 20 males and 8 females, aged between 18 to 55 years old. 6 of 28 participants recorded listening to binaural beats tracks before. All participants self-reported being inexperienced listeners.

4.2 Survey Overview

Our survey was distributed via the Qualtrics platform [18] and comprised two sections. First, we introduced participants to binaural beats, their use in mental wellbeing apps, and an introduction to Apollo. Next, we presented each participant with 10 different tracks randomly sampled from our dataset from three categories: deceptive titles (G1), deceptive audio (G2), and tracks that conform to user expectations (G3). Deceptive titles (G1) refer to tracks flagged by Apollo for F1 or F2; tracks that contain deceptive metadata. Deceptive audio (G2) refers to tracks flagged by Apollo for F3, F4 or F5; tracks with audio that does not align with the track metadata. G3 tracks pass Apollo's verification process and are not flagged.

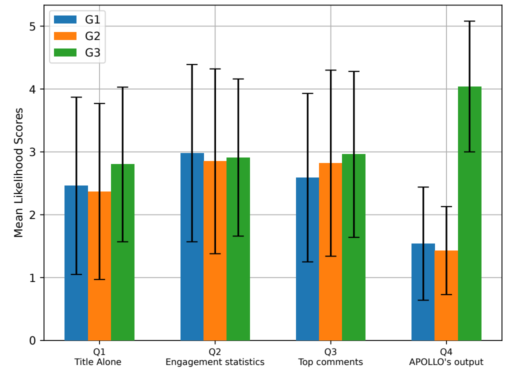

Figure 3: Mean scores and SD across different levels of information (Q1-Q4), for track categories G1-G3.

We designed our survey to study Apollo's influence on users' decisions when compared to track metadata. Thus, our null hypothesis remained: *Presenting Apollo's output does not change the user's likelihood of listening to the track.* For each track, participants scored the likelihood (from 1-5, with 5 representing most likely) of them listening to the track when presented with: (Q1) the track title alone, (Q2) the track title and its number of views, likes, and dislikes, (Q3) the track title and its top comments, and (Q4) the track title and a visual of Apollo's user interface for the track.

We allowed participants to add comments (e.g., reasons for choosing their score) for each question. We compare the participants' likelihood to listen to a track when presented with track metadata (Q1, Q2 and Q3) and Apollo's output (Q4) to determine if Apollo impacts their decisions. We allow participants to develop an opinion of the track before being exposed to Apollo's output (Q4 is last) to minimize participant response bias. Lastly, we asked the participants to answer: (1) *"How likely are you to use Apollo?"*, and (2) *"How useful is Apollo in helping you decide to listen to a track?"* on a scale of 1 to 5, with higher scores representing higher confidence in Apollo's output. Our survey was deemed exempt by our IRB.

4.3 Survey Results

Figure 3 shows the participants' mean likelihood and standard deviation (SD) for each question with respect to the three track categories (G1-G3). We found that with only the track title (Q1), participants reported a mean likelihood score of 2.46 out of 5 across all track categories. This shows that participants were indecisive about listening to the given track regardless of the type of claims made in the track title. However, when presented with a track's number of views, likes, and dislikes with the title (Q2), participants reported a higher mean score of 3.02 (22.7% increase from the track title alone). Similarly, participants expressed a higher likelihood mean score of 2.80 (13.8% increase from track title) when presented with the given track's most upvoted comments (Q3). We observed that higher view counts, a high likes to dislikes ratio, and positive comments motivated participants to listen to a track, irrespective of whether the track's intent was deceitful or not. For instance, one participant stated that *"results experienced by commenters are highly suspicious, but drives up curiosity"* when justifying a high score for a track (titled "Repair Nerve Damage Binaural Beats") with a deceitful intent but positive comments.

In contrast to track metadata, we found that the participants reported more decisive scores when presented with Apollo's output for a given track. Participants expressed a lower likelihood of listening to tracks that were flagged as deceptive by Apollo (G1, G2), and assigned a mean score of 1.54 and 1.43 for track categories G1 and G2 respectively. The low scores illustrate Apollo's success in informing users on deception (e.g., *"Apollo's analysis confirms that I would not benefit from this track."*, *"Well the app tells me that this video is bogus so I probably wouldn't watch it"*). When presented with a track that had been validated by Apollo (G3), participants reported a higher confidence and interest in listening to the given track, with a mean likelihood score of 4.04 out of 5, suggesting Apollo's ability to instill confidence in conforming tracks (e.g., *'If what I'm looking for is as described in the title. I'll actually [listen to] it"*, *"...This increases my confidence."*). We also observe that SD for all categories is lowest in Q4, indicating that variability across participants' responses is lowest when presented with Apollo's output among the four questions.

Participants reported strong confidence in Apollo's report and gave an average score of 4.13 out of 5 for its usefulness. Likewise, participants gave a score of 3.39 (mean) when asked if they were likely to use Apollo. Participants who provided lower scores were motivated by their apprehension of mental wellbeing audio. For example, a participant who assigned a low score noted that *"if [he/she] did start getting into these types of videos, then [he/she] would most likely use the software."*

We use the Wilcoxon Signed-Rank test [27] to test our null hypothesis. We compared participants' scores when presented with track metadata (Q1, Q2, Q3) and Apollo's output (Q4) and observed that p-value < 0.05 across three categories (G1, G2, G3). Thus, we reject the null hypothesis and determine a significant difference between how users use track metadata and Apollo when deciding if they should listen to a track.

Takeaway:
- Our tool communicates binaural beats present in a track and leverages user expectations in a UI.
- Providing users expectations conformance information significantly changes their likelihood of listening to a track.

5 Discussion

In this section, we synthesize the key takeaways of our study as well as discuss our study's limitations.

5.1 Key Takeaways

Expectation Conformance as a Standard. Our pre-design survey and Apollo user study demonstrate the reliability of expectation conformance as a standard for mental wellbeing technology. There is a plethora of mental wellbeing tech, and each tech would elicit its own set of expectations. Thus, it is reasonable to assume that users would seek online information sources to learn about such technology, as evidenced by our pre-design survey. As a result, leveraging commonly held user expectations is able to positively influence users in making decisions, evidenced in our study showing participants Apollo's output. We advocate that commonly held expectation conformance be considered as a core UX criterion in domains outside mental wellbeing (e.g., stakeholders can leverage online discussion forums to understand how to validate the design of user-facing systems such as social media privacy settings).

Our work adds to existing research on applying a human- and community- centered lens to mental wellbeing applications. We crowdsource expectations from experienced listeners' conversations online and use them to inform inexperienced listeners who consume mental wellbeing content. Our user study findings demonstrate how leveraging the insights of experienced users can significantly impact how users choose/avoid mental wellbeing content.

Designing for Mental Wellbeing Users. The key takeaway from our study is that tools that leverage commonly held user expectations can influence users' decision-making process in the context of mental wellbeing applications (e.g., how likely they are to listen to a track). Our user study demonstrates that compared to track metadata, Apollo has a signifcant impact on users' decision-making process. We envision that our tool can be fine-tuned to provide users with more actionable information. For instance, once deception is detected, perhaps Apollo could recommend a deception-free binaural beats track. However, whether users would trust an alternative suggestion would still need to be explored.

Additionally, our study bears implications for mental wellbeing applications outside the binaural beats domain. We show that information beyond application-provided metadata is helpful to users. Thus, future work should explore the applicability of mental wellbeing tools to other wellbeing applications. For example, a tool for visual-based therapy may integrate computer vision components to measure conformance to what users are expecting to see. Similarly, tools to verify applications providing guided meditation (where recorded audio provides instructions [26]) can leverage components such as large language models (LLMs). Here, such tools can apply an LLM on audio transcripts to verify if specific criteria are met. For instance, a user may wish to filter out tracks that include words of affirmation and only listen to tracks that instruct on breathing exercises, both common patterns in guided meditation applications.

5.2 Limitations and Future Work

We envision our user study to determine the influence of Apollo on users' decisions as a first step towards designing a mental wellbeing tool for binaural beats listeners. One limitation is that users are unable to customize the pre-defined expectations. Future work will explore a tab where users can define custom expectations that are automatically converted into deception detection rules.

Another limitation is that our recruited population for our user study with Apollo was inexperienced listeners of binaural beats. This population suited the purpose of our study as we aimed to design a tool with a wide target audience. However, it is likely that experienced listeners would desire tools such as Apollo. Here, they may desire more granular information and additional features. Future work will focus on conducting a pilot study with experienced listeners and redesigning Apollo to cater to this target population.

Moreover, it is important to acknowledge that the framing of output for mental wellbeing tools can impact users. To illustrate, an authoritative framing (e.g., *"Don't listen to this track. It does not follow user expectations"*) vs. a more measured framing (e.g., *'This track does not follow commonly held expectations. You may want to consider other options."*) can elicit different responses. This impact can be exacerbated if an authoritative framing is used to provide false information, e.g., user expectations that are out of date. Future work can experiment with different output framing and framing that can account for the quality of user expectations (e.g., expectations from a small sample size, expectations that are outdated) and their corresponding impact on users.

6 Related Work

User-Centered Design. Our work builds on user-centered design (UCD) [25], an established theoretical framework that advocates for leveraging users' insights in designing technology. Prior works have used this framework to advocate for better design of technology, such as mobile apps [8], APIs [19] and web interfaces [16]. We extend to the current literature by grounding on UCD to synthesize a tool that uses experienced users' insights. Our tool informs non-experienced listeners with commonly held expectations to aid them in making an informed choice.

Designing and Evaluating Mental Wellbeing Technology. Mental wellbeing technologies has been a long research focus within the HCI and CSCW community. To illustrate, one line of research studies mental wellbeing support among social media and online forum participants, focusing on data such as rhetoric and support-seeker characteristics of users [1, 4, 7, 9, 10, 22–24]. Methods such as field studies and expert evaluation to design mental wellness interventions also exist [5, 12, 13, 15, 28].

There have been efforts focusing on probing online wellbeing intervention claims. Methods to achieve this range from qualitative analysis grounded on scientific credibility [11], electroencephalography (EEG) [20], and auto-ethnography [6]. These studies rely on extra equipment and manual expert analysis.

In contrast, Apollo is an automated end-to-end system, that leverages rules synthesized from user expectations to inform users of any discrepancies found in a binaural beats track.

7 Conclusion

By leveraging commonly held user expectations, we design a tool to aid users in making informed decisions. We first conduct a pre-design survey to assess the value of synthesizing a tool. After ascertaining its potential benefit, we encapsulate our expectation conformance methodology in a tool, Apollo and evaluate it with 28 participants. Our study sheds light on the need for mental wellbeing stakeholders to carefully design interventions for users.

Acknowledgments

We thank our anonymous reviewers for their valuable feedback. This work is supported by startup funding from Purdue University.

References

[1] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *CHI conference on human factors in computing systems*.
[2] Arjun Arunasalam, Jason Tong, Habiba Farrukh, Muslum Ozgur Ozmen, Koustuv Saha, and Z. Berkay Celik. 2024. Understanding User Expectations for Binaural Beats. (2024).
[3] Ann-Katrin Becher, Marlene Höhne, Nikolai Axmacher, Leila Chaieb, Christian E Elger, and Juergen Fell. 2015. Intracranial electroencephalography power and phase synchronization changes during monaural and binaural beat stimulation. *European Journal of Neuroscience* (2015).
[4] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* (2020).
[5] Prerna Chikersal, Danielle Belgrave, Gavin Doherty, Angel Enrique, Jorge E Palacios, Derek Richards, and Anja Thieme. 2020. Understanding client support strategies to improve clinical outcomes in an online mental health intervention. In *CHI conference on human factors in computing systems*.
[6] Claudia Daudén Roquet and Corina Sas. 2018. Evaluating Mindfulness Meditation Apps. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*.
[7] Jessica L Feuston and Anne Marie Piper. 2018. Beyond the coded gaze: Analyzing expression of mental health and illness on instagram. *ACM on Human-Computer Interaction* (2018).
[8] Andrea K Graham, Sarah W Neubert, Angela Chang, Jianyi Liu, Emily Fu, et al. 2021. Applying user-centered design methods to understand users' day-to-day experiences can inform a mobile intervention for binge eating and weight management. *Frontiers in Digital Health* (2021).
[9] Meeyun Kim, Koustuv Saha, Munmun De Choudhury, and Daejin Choi. 2023. Supporters First: Understanding Online Social Support on Mental Health from a Supporter Perspective. *ACM on Human-Computer Interaction* (2023).
[10] Taisa Kushner and Amit Sharma. 2020. Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums. In *ACM web conference*.
[11] Mark Erik Larsen, Kit Huckvale, Jennifer Nicholas, John Torous, Louise Birrell, Emily Li, and Bill Reda. 2019. Using science to sell apps: evaluation of mental health app store quality claims. *NPJ digital medicine* (2019).
[12] Emily G Lattie, Rachel Kornfield, Kathryn E Ringland, Renwen Zhang, Nathan Winquist, et al. 2020. Designing mental health technologies that support the social ecosystem of college students. In *CHI Conference on Human Factors in Computing Systems*.
[13] Reeva Lederman, Greg Wadley, John Gleeson, Sarah Bendall, and Mario Álvarez-Jiménez. 2014. Moderated online social therapy: Designing and evaluating technology for mental health. *ACM Transactions on Computer-Human Interaction (TOCHI)* (2014).
[14] Nielsen 2021. How Many Test Users in a Usability Study? https://www.nngroup.com/articles/how-many-test-users/. [Online; accessed 1-May-2024].
[15] Kathleen O'Leary, Stephen M Schueller, Jacob O Wobbrock, and Wanda Pratt. 2018. "Suddenly, we got to become therapists for each other" Designing Peer Support Chats for Mental Health. In *CHI conference on human factors in computing systems*.
[16] Jonas Oppenlaender, Thanassis Tiropanis, and Simo Hosio. 2020. CrowdUI: Supporting web design with the crowd. *ACM on Human-Computer Interaction* (2020).
[17] Gerald Oster. 1973. Auditory beats in the brain. *Scientific American* (1973).
[18] Qualtrics 2024. Qualtrics XM - Experience Management Software. https://www.qualtrics.com/. [Online; accessed 14-May-2024].
[19] Lavanya Ramakrishnan, Sarah Poon, Valerie Hendrix, Daniel Gunter, Gilberto Z Pastorello, and Deborah Agarwal. 2014. Experiences with user-centered design for the Tigres workflow API. In *IEEE International Conference on e-Science*.
[20] Daniel Reinhardt, Cordula Baur, Sara Klüber, and Jörn Hurtienne. 2020. MindPeaks: Formative Evaluation Method of Mindfulness Meditation Apps. In *Companion Publication of the ACM Designing Interactive Systems Conference*.
[21] Bernhard Ross, Takahiro Miyazaki, Jessica Thompson, Shahab Jamali, and Takako Fujioka. 2014. Human cortical responses to slow and fast binaural beats reveal multiple mechanisms of binaural hearing. *Journal of neurophysiology* (2014).
[22] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *ACM on Human-Computer Interaction* (2017).
[23] Koustuv Saha and Amit Sharma. 2020. Causal factors of effective psychosocial outcomes in online mental health communities. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 590–601.
[24] Koustuv Saha, John Torous, Eric D Caine, and Munmun De Choudhury. 2020. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *Journal of medical internet research* (2020).
[25] Edward Tenner. 2015. The design of everyday things by donald norman. *Technology and Culture* (2015).
[26] Wikipedia 2024. Guided Meditation. Retrieved 10-July-2024 from https://www.headspace.com/meditation/guided-meditation
[27] RF Woolson. 2007. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* (2007).
[28] Dong Whi Yoo, Michael L Birnbaum, Anna R Van Meter, Asra F Ali, Elizabeth Arenare, et al. 2020. Designing a clinician-facing tool for using insights from patients' social media activity: Iterative co-design approach. *JMIR Mental Health* (2020).