Proceedings of the Nineteenth International AAAI Conference on Web and Social Media (ICWSM 2025)

Deceptive Sound Therapy on Online Platforms: Do Mental Wellbeing Tracks Conform to User Expectations?

Arjun Arunasalam[1], Jason Tong[1], Habiba Farrukh[2], Muslum Ozgur Ozmen[3], Koustuv Saha[4], Z. Berkay Celik[1]

[1] Purdue University, West Lafayette, IN, USA
[2] University of California, Irvine, Irvine, CA, USA
[3] Arizona State University, Tempe, AZ, USA
[4] University of Illinois Urbana-Champaign, Urbana, IL, USA
aarunasa@purdue.edu, tong72@purdue.edu, habibaf@uci.edu, moozmen@asu.edu, ksaha2@illinois.edu, zcelik@purdue.edu

Abstract

The rising popularity of mental wellbeing technologies has led many individuals to explore binaural beats—an emerging form of sound therapy proliferating on web and mobile platforms. However, it currently remains unknown whether users can trust binaural tracks on online platforms, or if they deceive unsuspecting users. Our research aims to address this problem by understanding (1) what binaural beats listeners expect from tracks and (2) whether online tracks conform to these expectations. To understand user expectations, we perform thematic analysis on online forum threads and blog posts to extract binaural beats goals and expectations tied to these goals. Next, we design a methodology to measure binaural beats tracks' conformance to commonly held user expectations. This methodology comprises, (1) obtaining a track's intent to induce a mental state through track metadata analysis, (2) extracting a track's binaural beats time-frequency model using Fast Fourier Transform, (3) mapping user expectations to rules that identify deceptive tracks, and validating them on the track's extracted intent and time-frequency model. We evaluate ~7K binaural beats tracks and find that only 7.5% conform to commonly held user expectations, while the remaining 92.5% deceive users with deviant claims (e.g., eroticism, weight loss) or deliver contradicting binaural beats. Our study underscores the significance of understanding users' expectations and verifying conformance of online wellness technologies to expose discrepancies in expectations.

1 Introduction

People have increasingly turned to mental wellbeing technology to improve their quality of life. The upward trend in mental health awareness is largely influenced by the COVID-19 pandemic. Symptoms of loneliness, depression, and anxiety have become highly prevalent due to limited social contact, disruption in routines, and income reduction (Palgi et al. 2020). To cope with the unfortunate rise in such symptoms, the demand for mental wellbeing apps has skyrocketed. Global user spending on meditation apps has increased by more than 50%, surpassing $195 USD million in 2020 (empeek). Additionally, listening time for mental health tracks has also increased by 50% on audio platforms (e.g., Spotify), causing a 30% increase in traffic (Digital Music News).

Among mental wellbeing apps, apps leveraging a claimed auditory effect known as binaural beats have become increasingly popular. Binaural beats are presented as an auditory illusion where the brain, upon hearing two tones with differing frequencies played in each ear, perceives a separate tone, known as the binaural beat. Here, we note that the science behind the binaural effect and its overall neuroscientific impact on users is mercurial. Although some have advocated its usefulness, others have deemed it a placebo. Regardless, binaural beats have emerged as an extremely popular form of sound therapy, resulting in a large community that claim they use it to improve their wellbeing.

To tap into the expanding binaural beats community, many content creators have started producing binaural beats tracks that intend to satisfy users' expectations. These tracks are available across various web and mobile platforms (e.g., YouTube, Spotify, BrainTap) for users to consume, either for free or for a fee. Most of these platforms allow any user to upload content for public viewing. While these platforms offer community guidelines and procedures for regulating content, tools and algorithms for evaluating whether content creators' uploaded tracks conform to their listeners' expectations are largely absent. This absence raises concerns about whether users can be deceived by consuming tracks that fail to deliver the expected binaural beats.

The research community has long explored the impact of emerging technology on users, especially when these technologies do not work as anticipated. For instance, work has been done to detect deceptive videos that target and harm children (Papadamou et al. 2020; Singh et al. 2019) and usability issues with two-factor authentication (Turner et al. 2023; Reese et al. 2019). However, research efforts to examine the user impact of claimed mental wellbeing technologies and corresponding deception are largely absent. For example, binaural beats tracks may make misleading claims to lure listeners for monetization (via advertisement income). Similarly, listeners could be redirected (via track descriptions) to harmful websites where bad actors steal credentials. Also, tracks may be absent of binaural beats or beats that deviate from accepted norms - the specific frequencies users associate with an intent. Here, listeners are deceived as they are victimized by false advertising and exposed to outlandish claims, resulting in time/money loss.

Given that no prior work has focused on answering whether binaural beats tracks abide by user held beliefs, we focus on answering the following research questions:

RQ1 What are users' expectations surrounding binaural beats?

RQ2 Do binaural beats tracks on online platforms conform to commonly held user expectations?

To answer the challenging question of what mental wellbeing goals users expect (**RQ1**), we crawl the web for forum threads and blogs related to binaural beats and develop an understanding of user expectations. We then execute thematic analysis on this collected data. We identified four commonly accepted mental state goals (relaxation, sleep, concentration, and cognitive enhancement) associated with binaural beats and the respective expectations (types of binaural beat frequencies) listeners associate with each mental state. From **RQ1**, we also discover users are concerned about deception, e.g., concerned about spreading false binaural content, erosion of trust with platforms, and that users desire verification.

To answer **RQ2**, we first had to synthesize a methodology to verify if a given binaural beats tracks conforms to user expectations. To achieve this, we introduce a three-step method. First, we perform track metadata analysis using mental state lexicons to recognize a track's intent. To address the challenge of capturing the perceived binaural beat, we then apply Fast Fourier Transform to translate track waveforms into a time-frequency model. This model contains the peak frequencies on both left and right channels for the track's duration and serves as an intermediate representation for further analysis. From this, we map user expectations to systematically develop five rules before leveraging a track's intent and time-frequency model to validate a binaural beats track.

We applied this methodology on over ~7K binaural beats tracks available on popular audio/video-sharing platforms (YouTube and Spotify). We also evaluate paid binaural beats tracks purchased via e-commerce platforms (some of which make their tracks available through mobile apps). We find that only 7.5% of the tracks conform to user expectations. The remaining 92.5% tracks deceive users for monetary benefit, providing them binaural beats that contradict users' expectations and advertising deceptive claims. Our analysis also uncovered five common deceptive advertising categories, such as weight loss, sexual performance, and financial growth, used by content creators to lure users into listening to their tracks. In this paper, we make the following contributions[1]:

- We identify and characterize four mental wellbeing goals and respective expectations binaural beat listeners have via thematic analysis of forums and blog entries.
- We introduce an expectation conformance methodology for binaural beats tracks, which extracts a track's intent and time-frequency model and discovers where tracks deviate from expectations.
- We evaluate over 7K tracks from popular audio/video streaming platforms and expose deceptive tracks.

[1]We make the supplementary materials of this research project available at https://osf.io/ks4yv/.

2 Background

2.1 Related Work

Technology that Users Leverage for Mental Wellbeing. Everyday users often leverage various forms of technology with the overall goal being to improve their mental wellbeing, thus, becoming a focal point of the CSS and HCI community. Prior research has focused on online mental wellbeing communities to investigate factors such as peer support (Saha and Sharma 2020; De Choudhury and De 2014; Saha et al. 2020). Another line of work has focused on investigating the claims of digital mental health/wellbeing interventions, via auditing methods involving qualitative analysis and expert evaluation (Larsen et al. 2019; Daudén Roquet and Sas 2018)

Despite widespread adoption of mental wellbeing technology, it is important to note that the neuroscientific evidence on this technology's impact on users is largely disputed. Mental wellbeing tech is often referred to as pseudoscience, with the perceived impact attributed to the placebo effect (NY Times a,b). Although methods evaluating wellbeing tech may account for placebos (e.g., via control groups), research has argued that the impact of placebos on mental wellbeing interventions is hard to understand (Boot et al. 2013). Similarly, prior work has shown that placebo treatments can significantly improve ones medical condition (Bschor et al. 2024).

User Expectations of Technology Services. Our efforts to understand if binaural beats on online platforms conform to user expectations is grounded in two theoretical frameworks. First, the User-Centered Design (UCD) framework (Tenner 2015) argues that users should be significantly involved in product design, including technology services. Various factors should be accounted for, such as age, gender, and, closely related to our work, users' expectations. UCD has been a leveraged framework for understanding digital mental health interventions (Cornet et al. 2020; Vial, Boudhraâ, and Dumont 2022). These works outline the lack of attention paid to user perspectives during intervention design.

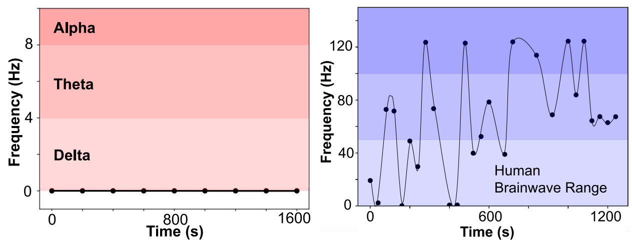
Second, the Technology Trust Framework (Mcknight et al. 2011) builds on UCD to synthesize a model of fundamental principles of user trust in technology. This framework argues that for users to be able to trust technology, such technologies should "operate reliably and without failing". The Technology Trust Framework has been leveraged in understanding user adoption of emerging technologies such as Covid-19 tracing apps (Riemer et al. 2020).

We draw on these two theoretical frameworks to adopt the following **motivating principle**, *"For users to be able to trust binaural beats tracks and find them reliable, these tracks should conform to commonly held user expectations."*

When binaural beats technology fails to adhere to this principle, we consider that users have been deceived. Our research leverages these two theoretical frameworks to (1) understand what users expect from binaural beats technology and (2) correspondingly introduce a methodology to verify if tracks deviate from user expectations.

2.2 Binaural Beats

Among the plethora of mental wellbeing technologies, binaural beats tracks have gained credibility from users who claim their effectiveness in overcoming anxiety, improving



(a) No binaural beats (b) Random waveforms

Figure 1: Time-frequency representation of deceptive binaural beats tracks that target listeners.

relaxation, increasing focus, and inducing a meditative state. Binaural beats proponents claim this technology to be an auditory illusion that occurs when two tones of different frequencies are separately played into each ear, causing the human brain to interpret the resonance between two frequencies as an illusion of a third sound, known as binaural beats (Oster 1973). These proponents deem this phenomenon as entrainment – the synchronization of brain frequencies (brainwaves) to a sound (Ross et al. 2014; Becher et al. 2015).

Following this logic, listening to binaural beats requires headphones/earphones to ensure that the left and right ears receive the different tones. For instance, if the left ear receives a 200Hz tone and the right ear receives a 210Hz tone, the claim is that brain absorbs a 10Hz tone. Here, it is important to note that the effect of binaural content on someone's mood is often contested, with many arguing that it is pseudoscientific and attributed to a placebo effect. This mirrors the sentiment surrounding most mental wellbeing technology, where efficacy is often disputed, as stated in Section 2.1.

2.3 Research Motivation

Despite claims of pseudoscience, binaural beats have gained widespread popularity. The phrase "binaural beats" has doubled in popularity since 2020 (Google Trends). Binaural tracks are listened to on various platforms, with some amassing view counts in the tens of millions. For instance, searching "binaural beats" on Spotify returns the top 1,483 related playlists that are frequented by $\sim$1.4 million listeners. The same search query on the Google Play Store returns the top 88 binaural beats apps that have $\sim$213 million combined downloads. These tracks have recently made numerous appearances in various media outlets (BrainTap), signifying their growing popularity. This popularity has also led to an influx of commercial products that combine audio with flashing light via strobe light goggles, e.g., the BrainTap headset (BrainTap) uses light pulses and binaural beats for brainwave entrainment. Additionaly, there exists various online guides to help content creators produce binaural beats tracks (YouTube; 344 Audio). Growing popularity and ease of production makes binaural beats prime for investigation.

Given there exists a growing body of consumers who leverage binaural beats as mental wellbeing technology, it becomes imperative to understand "Do binaural tracks online conform to what most users expect?" We note that abiding by our motivating principle means that for users to be able to trust binaural tracks, these tracks should conform to commonly held user expectations, regardless of the neuroscientific impact of binaural beats. For instance, Figure 1 previews the binaural beat frequency of two online binaural tracks with respect to time (synthesis of which we explain in Section 4). In (a) no binaural beats are produced and in (b) the binaural beats produced do not follow any identifiable pattern. Creators may also craft track titles that contain outlandish claims in an attempt to lure users into listening (Appendix A.1).

Given that such tracks may plague online services and guided by our motivating principle, we aim to bridge an existing research gap by first (1) understanding what users expect from binaural beats (**RQ1**) and (2) evaluating how well binaural beats tracks conform to user expectations (**RQ2**).

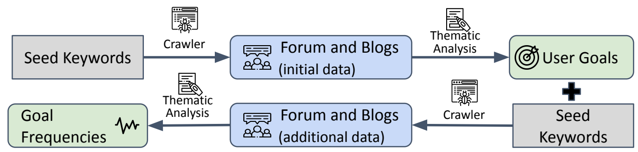


Figure 2: Overview of methodology to understand user expectations surrounding binaural beats,

3 RQ1: Examining User Expectations

To answer RQ1, we draw on digital/online ethnography methodology as a method to understand user expectations. Additionally, prior work has also shown that participants of online communities display information seeking behaviour (Hwang and Foote 2021) while another line of research has also leveraged the use of online discussion platforms as a tool with a high likelihood of ecological validity – capturing authentic interactions between users (Hinchcliffe and Gavin 2009; Colom 2022).

Building on the above, Figure 2 overviews our methodology to understand user expectations. We start with seed keywords to crawl the web to identify online communities: forum and blog data, to thematically analyze (1) user expectations for binaural beats goals. We repeat our crawling process with goals as additional keywords to collect additional blog and forum data which we thematically analyze to understand (2) the frequency expectations for each goal.

3.1 Analysis of Online Communities

Prior work shows that information-seeking behavior for health-related topics is common among users (Social Life; Fox and Fallows 2003). Thus, we pivot to online resources to understand expectations. We implemented a web crawler to identify posts/conversations about binaural beats on public forums. We used the Google Trends API (Google Trends) to collect the top 20 phrases that included the keywords "binaural beats" (Appendix A.2). We collected the top 100 URLs obtained in response from Google Search API (Google Search) for each phrase. Through this, we collected 823 unique URLs representing binaural beats-related content.

We sampled 20% of URLs for further analysis. We noticed that collected URLs point to diverse websites. Websites primarily pointed to text-based content, e.g., scientific papers, online forums, blogs, and audiovisual content, e.g., audio content on various platforms and binaural beats applications.

We qualitatively analyzed both text-based and audio-visual content. During this phase, we discovered that audio-visual content lacked metadata appropriate to ground user expectations. For instance, comments on binaural beat tracks on YouTube did not indicate user expectations - comments were predominantly unrelated (e.g., *"Thank you, [content creator channel name]! Happiest holidays ahead to all"*). However, discussion of binaural content amongst *users* originated in forum threads and blog entries. Thus, the two authors manually filtered the URLs to select only those associated with forum threads and blog entries. Through this filtration, we obtained 615 URLs for further analysis of listeners' expectations. These URLs comprise 388 unique domain names. The most frequent domain in our dataset ($n = 85$) was reddit.com spanning 35 subreddits (subforums).

Extracting Binaural Beat Goals. We deductively coded collected pages to identify the goals that binaural beats listeners hope to achieve. We first randomly sampled 10%, with two authors jointly developing a codebook by manually analyzing each webpage, generating initial codes, and reiterating until both authors achieved codebook stability. The authors met over multiple sessions to refine codes and reconcile disagreements. At each iteration, we note high agreement ($\kappa > 0.8$) (McDonald, Schoenebeck, and Forte 2019). Finally, the authors clustered codes that shared a common theme as a binaural beats goal. We continue sampling, repeating our thematic analysis process. Between analyzing the fourth and fifth iterations (305 URLs), we observed that no new binaural beats goals had emerged as a theme. This indicated that we had reached thematic saturation, suggesting that our initial crawl had sufficiently captured what users expect to gain from listening to binaural beats. We continued qualitatively analyzing the remaining 310 URLs and through this, confirmed that no new themes emerged.

We follow standard practice in thematic analysis by setting a threshold of 30% (Fugard and Potts 2015) This allows us to avoid including peripheral themes and improve the reliability that themes presented are indeed commonly held user expectations. The peripheral themes in our data appear in percentages $< 10\%$. From our thematic analysis, we found that users listen to binaural beats tracks to achieve one of the following mental state goals: (1) relaxation, (2) sleep, (3) concentration, and (4) cognitive enhancement.

3.2 Goals and their Expectations

Mental State Goals. The first mental state goal binaural beats listeners express is **relaxation**; 68.2% of the URLs in our dataset presented themes pertaining to relaxation. Listeners claim that these beats help them achieve a meditative state or provide relief from anxiety. One listener stated that *"[binaural beats] really do help [him/her] get into a meditative state faster than without them"* while another *"[became] dependent on them [for] a deep meditative state."* Binaural beats listeners (59.5% of URLs) also claim that binaural beats help them with **sleep**/hypnosis, where users *"[use them] to sleep effectively"* and claim a *"very satisfying sleep experience"*. There also exists a belief among listeners that binaural beats allow one to **concentrate** on a specific task, a theme found in 82.0% of the URLs. For instance, on one Reddit thread concerning methods to improve studying, a listener states that they *"... use [beats] when working if [he/she] is easily distracted."* In our analysis, we also found that binaural beats listeners believe that binaural beats can **enhance their cognitive ability**, allowing them to solve complex problems and perform cognitive tasks. Binaural listeners (36.2% of URLs) claim that binaural beats are *"important for learning, memory and information processing"*. We note that the percentages pertain to the relative frequency of themes discovered during our thematic analysis. Thus, the most prominent theme does not reflect the most prevalent binaural content consumed by users. However, they still provide an understanding of the *common* and *widespread* binaural beats goals users expect when they consume binaural content. Here, we note that our analysis also uncovers that users believe goals are distinct – a track cannot help a user achieve more than one goal.

Goal Specific Expectations. After identifying mental state goals, we focus on what listeners expect for specific goals. To do so, we repeated our web crawling process. We leveraged thematic analysis codes as search keywords. Similar to our analysis for identifying goals, we selected URLs related to forums and blog entries, and two of the authors manually coded the collected data. We collected $2K$ URLs via this method for further analysis. While deductively coding, we grounded on expectations behind the mental state goal. We focus on the mention of brainwave patterns.

Interestingly, we discovered that when users express the brainwaves related to their goal, they also note the frequencies associated with these waves. Listeners often quote scientific studies when associating a wave type with a frequency range (detailed in Section 3.3). We note that the frequencies associated with brainwaves match these studies. After coding, we filtered out themes with a prevalence lower than 10% to extract as many binaural beats specific user expectations as possible. Similar to our thematic analysis for binaural beats goals, we iteratively sampled 10%, reached thematic saturation at the 6th iteration, and then completed the thematic analysis of all URLs to confirm no new themes were realized.

Our final codebook revealed six distinct user expectations: two for relaxation, two for sleep, one for concentration, one for cognitive enhancement, with each expectation unique to a specific mental state goal. We found that listeners expect **alpha** (8-12Hz) waves or **theta** waves (4-8Hz) in binaural beats for relaxation. Many binaural beats listeners expect binaural beats for sleep/hypnosis to be in the **delta** wave range (0.5-4Hz), citing research that claims that the delta wave is commonly associated with the sleep state. Listeners also expect binaural beats for sleep to follow frequencies of the **sleep cycle**. The sleep cycle refers to the different stages of sleep, with each stage having an association with different brainwaves (Carskadon, Dement et al. 2005). Listeners often cite scientific research to explain or justify the change in frequencies in accordance with the sleep cycle. Listeners who want to be in a focused/concentrated state expect their tracks to comprise of frequencies in the **beta** range (12-38Hz). Similarly, listeners associate the **gamma** range (38-42Hz) with increased cognitive capability and expect this brainwave when listening to binaural beats for cognitive enhancement.

3.3 User Reasoning and Concern

To better understand expectations, we inductively coded the $2K$ URLs, thus exposing two additional themes: (1) research-driven beliefs and (2) user concern and suspicion.

Research-Driven Beliefs. We discovered that listeners believe that binaural beats are scientifically proven to be able to achieve relaxation, sleep, concentration, and cognitive enhancement. To illustrate, listeners state that their confidence in the discovered goals above is due to information from trusted sources. For example, many listeners reference popular news websites (e.g., Washington Post) and other journalism websites (e.g., Popular Science) that promote binaural beats when justifying their beliefs. Such news outlets provide users pointers to other resources such as research papers and links to medical websites (e.g., WebMD, Healthline). We highlight two examples in Appendix A.3, where users are pointed to resources such as research papers and medical websites. Listeners also mention peer-reviewed papers when supporting claims such as *"scientific studies suggest [binaural beats] have an effect on the brain"*. Given that listeners' beliefs of accepted goals are based on user-trusted sources, listeners expect specific brainwaves (within range of frequencies) to meet their expectations.

User Concern and Suspicion. Another overarching theme consistent across studied forums is **user suspicion and concern** surrounding the binaural beats tracks hosted on platforms. First, users demonstrate confusion about the presence of binaural beats - questioning whether promised frequencies are present in the track. To illustrate, one user asked *" I need to be sure [a track] is truly made with binaural beats."*, with another stating that *"[they] would like to confirm the hz [of a track]"*. Users also express concerns of deception - they do not trust the binaural beats tracks they consume. For example, users ask questions such as *"How [to] recognize [legitimate] beats from junk?"* with another stating they use binaural beats but suspect *" [they] think [they] tried some fake ones...."*. Similarly, one user stated *"don't expect YouTube [binaural beats tracks] to be effective"* while another expresses skepticism (e.g., *"I'm quite the skeptical guy.."*)

Interestingly, one key component of users' concerns is that users do not know when they are deceived. As a consequence, users fear that they may falsely share false binaural content. It also results in an erosion of trust in the online platforms that users visit to consume binaural beats. More interestingly, lack of knowing when they are deceived results in users requesting pointers to "legitimate" binaural beats sources. For instance, one concerned user asked *"Any youtube channels that have binaural beats videos that are real?"*, suggesting general apprehension towards binaural beats due to the lack of a credible streaming source. We note that users who express these concerns do not doubt the ability of binaural beats in mental wellbeing, but raise doubts about whether tracks on streaming platforms contain the advertised audio, e.g., specific frequencies associated with titles. In short, users express (1) uncertainty about binaural content frequencies, (2) concerns about being deceived, and (3) consequences of deception, which result in a desire for verified resources.

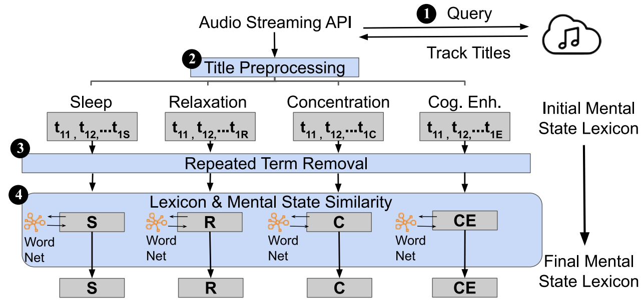


Figure 3: Overview of track metadata analysis (**S**: Sleep, **R**: Relaxation, **C**: Concentration, **CE**: Cognitive Enhancement)

4 Conformance Analysis Methodology

In answering **RQ1**, we discovered that users have specific goals and expectations. However, verifying track conformance requires understanding what goal a track is advertising and whether the track abides by expectations tied to this goal. To solve this, we introduce a three-step methodology.

First, a track is classified into one of 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 binaural beats tracks classified into more than one mental state, and outlier indicates the tracks with unclear or vague intents. Second, 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 produces the binaural beats and patterns produced throughout the track duration. Third, we map user expectations to five rules. We validate a binaural beats track's conformance to user expectation against these rules, using its extracted intent and time-frequency model.

4.1 Track Metadata Analysis

A track's metadata typically consists of a track's title, description, tags, and thumbnail, depending on the audio streaming platform. Titles often contain descriptive words that relate to the track. Our analysis on popular platforms streaming binaural beats tracks (e.g., Spotify, YouTube), shows that track metadata such as descriptions/tags contains information irrelevant to the track, such as subscription guides to apps/channels, URLs to external websites, and attention-grabbing keywords for search-engine optimization. The reasons above influence our decision to leverage the track title in understanding the mental state goal the binaural beat track is advertising to induce. Figure 3 illustrates our natural language processing (NLP) techniques to curate mental state lexicons to correctly extract a track's intent.

Intended Mental State Goal Extraction. We classify a binaural beats track's title into one of the following classes: sleep, relaxation, concentration, cognitive enhancement, multi-class, and outlier. To accomplish this, we build a lexicon that includes a set of unique words for each accepted goal discovered (sleep, relaxation, concentration, and cognitive enhancement). If a track title contains a word/words from only one lexicon, it is labeled with the corresponding mental state. Additionally, we introduce two classes to capture tracks that do not conform to user expectations. If a title contains words from multiple lexicons, it is assigned to multi-label. This is as our findings from RQ1 show us that users believe that one track cannot achieve multiple goals. Additionally, titles that do not contain any words from lexicons are labeled outliers, as they represent tracks which do not promise to achieve one of the four accepted goals.

To build a lexicon for each mental state, we initially include a mental state's expected brainwave(s) to its lexicon (e.g., beta is added to concentration) as these brainwaves express their respective mental state. We then search binaural beats tracks on YouTube and Spotify, the most common streaming platforms for binaural beats, and process their titles. We perform queries for strings such as "binaural beats" with a target state excluding other mental states to obtain the track titles (❶). We repeat the search process for each mental state. For instance, for sleep mental state, we use the query "binaural beats sleep - relaxation - concentration - cognitive enhancement" through YouTube Data and Spotify Developers APIs. Sampling tracks from these platforms allows us to collect an extensive overview of titles and ensures we build a comprehensive lexicon that includes words unique to each mental state. Overall, we obtained titles of 1255 binaural beats tracks, including 389 sleep, 232 relaxation, and 384 concentration and 250 cognitive enhancement tracks.

To obtain the word candidates from these tracks, we first tokenize the title into words, remove stopwords from the words, and use stemming to remove inflectional endings from words (❷). Resulting words represent the candidates for lexicons. However, we found that 26.4% of the words appear in lexicons for more than one mental state. For instance, the word "peace" is found in both sleep and relaxation track titles. To ensure the lexicons for each mental state are independent, we remove keywords shared across multiple lexicons (❸), resulting in a set of words unique to each lexicon.

Lexicon and Mental State Similarity. To improve the likelihood that all extracted words from titles are relevant to their respective mental state, we conduct a similarity-check. A word in the lexicon is relevant if it is semantically similar to its corresponding mental state (❹). Consider the title "(Extremely Powerful) Self Connection Meditation Binaural Beats Meditation" used in building our relaxation lexicon. After tokenizing and removing shared keywords, we find that our lexicon contains words such as "extremely" and "connection", not relevant to relaxation.

To address this challenge, we computationally identify non-relevant keyword candidates to remove them from the lexicon. We leverage the Wu-Palmer algorithm that calculates the lexical-semantic relatedness amongst two words (Wu and Palmer 1994). Wu-Palmer is a path-based algorithm based on the depths of two words and their depths to the most specific ancestor node in the taxonomies. To compute similarity scores, we use the WordNet taxonomies (Miller 1995), a lexical database of semantic relations between 155K words organized in 175K synsets (logical groupings). The returned similarity score ranges from $0 < \texttt{similarity} \leq 1$; Scores closer to 0 indicate that the word and mental state are not semantically similar and a score closer to 1 signifies high similarity. To find the thresholds for each lexicon, we compute the mean similarity score and add its standard deviation (SD). This enables us to obtain relevant words semantically related to each mental state.

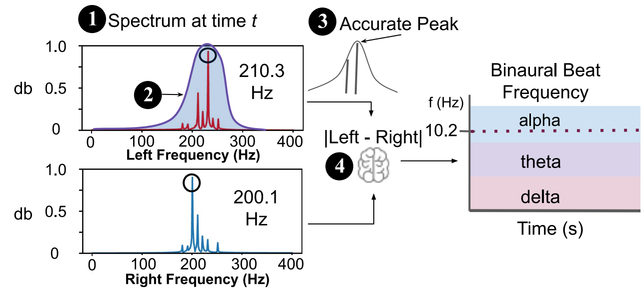


Figure 4: Time-frequency model extraction process.

Lexicon Confirmation Process. To validate the lexicons, two authors label words in each lexicon by (a) checking the relevance of words absent in the WordNet taxonomy (resulting in a score of 0), and (b) confirming our automated process builds correct lexicons. This allows us to ensure that mental state lexicons include the correct words to infer the intent of the tracks. Each annotator independently inspects the words that are absent in the WordNet to determine whether the word is relevant to one of the mental states. They then independently inspect the complete set of words in the lexicons to determine whether they are relevant to the mental states. We find a Cohen κ (McDonald, Schoenebeck, and Forte 2019) agreement of 1, denoting "perfect" agreement of lexicons.

Validating Intent Extraction To validate our methodology, we validate with 120 manually crafted strings and 120 additional titles collected from the YouTube Data API (total of 40 per class). No validation string was used in our initial lexicon building. All strings are passed into the intent extraction component: our component correctly labels all 240 strings.

4.2 Extracting Binaural Beats from Audio

We extract binaural beats frequencies produced over the track's duration from its audio, as shown in Figure 4. This process begins by dividing the track's audio waveform into its left and right channels (tones) (❶). To obtain the accurate peak frequency of left and right channels at a given period of time, we apply Fast Fourier Transform (FFT) with respect to a timestamp. We then identified two problems that impact the quality of extraction: spectral leakage and resolution of frequencies. To address spectral leakage, we use a Blackman-Harris window (❷). We also use parabolic interpolation to increase the frequency measurement resolution (❸). Applying these methods produces a time-frequency model of a track's binaural beats by calculating the absolute difference of the channels' frequencies (❹), expressed as $f(t) = f_t = |l_t - r_t|, t \in [t_0, t_n]$. Here, f_t is the binaural beat frequency, l_t and r_t are the left and right channel frequencies at time t (seconds). We set t to be 1 second.

Extracting Time-frequency Model. The track waveform is defined on the basis of amplitude versus time (seconds), with the overall audio length representing the total number of

#	User Expectations	Formal Expression (LTL)
S1	A single period of delta in track's duration	$\Box \mathtt{s} = \mathtt{delta}$
S2.1	Cycle to start with alpha before transitioning to theta	$\Box(\mathtt{s} = \mathtt{alpha}) \rightarrow \circ((\mathtt{s} = \mathtt{alpha}) \vee (\mathtt{s} = \mathtt{theta})), \Box(\mathtt{s} = \mathtt{alpha}) \rightarrow \Diamond(\mathtt{s} = \mathtt{theta})$
S2.2	Theta to continue before transitioning to delta	$\Box(\mathtt{s} = \mathtt{theta}) \rightarrow \circ((\mathtt{s} = \mathtt{theta}) \vee (\mathtt{s} = \mathtt{delta})), \Box(\mathtt{s} = \mathtt{theta}) \rightarrow \Diamond(\mathtt{s} = \mathtt{delta})$
S2.3	Cycle to end with delta or transition back to theta	$\Box(\mathtt{s} = \mathtt{delta}) \rightarrow \circ((\mathtt{s} = \mathtt{delta}) \vee (\mathtt{s} = \mathtt{theta}))$
R1	A single period of alpha in track's duration	$\Box \mathtt{s} = \mathtt{alpha}$
R2	A single period of theta in track's duration	$\Box \mathtt{s} = \mathtt{theta}$
C1	A single period of beta in track's duration	$\Box \mathtt{s} = \mathtt{beta}$
CE1	A single period of gamma in track's duration	$\Box \mathtt{s} = \mathtt{gamma}$

Table 1: Goal-Expectation Negligence (F4) rules for each mental state to mitigate user deception.

samples in the audio signal. To determine the peak frequencies between left and right channels from raw audio data, we use FFT to conduct spectral analysis on each channel of a binaural beats track (Brigham 1988). We define the number of frequency bins as half the number of audio samples collected for the input audio signal. We set the frequency resolution (Δf) to $f_{\text{max}}/N_{\text{bins}}$, where f_{max} is the maximum frequency, and N_{bins} is the number of bins.

An audio signal's maximum frequency is defined as half the audio's sampling rate with the Nyquist-Shannon sampling theorem (Landau 1967). Because the human hearing range is 20Hz to 20kHz, the sampling rate of audio tracks is typically between 44.1kHz and 48kHz. We define the number of audio samples equal to the audio sampling rate. We set the FFT window size to be equal to the audio sampling rate, for a time resolution of $1s$, a resolution suitable given that we track frequencies every 1 second. The total number of frequency bins is equivalent to half that of the FFT window size, resulting in a frequency resolution of 1Hz. This allows us to represent each frequency within the human hearing range as an individual frequency bin. Yet, 1Hz intervals limit the precision of peak frequencies to integer values, which may result in spectral leakage and inaccurate peak frequencies.

Addressing Spectral Leakage. The FFT converts a binaural beats track waveform to its periodogram, visualizing it as a continuous signal that repeats its pattern. However, pattern completion occurs only with an integer frequency. Without it, the FFT truncates the waveform, leading to spectral leakage—energy leakage from one frequency bin to its neighbors (Lyon 2009). To address this, we apply the minimum 4-term Blackman-Harris window (Harris 1978) over the track every second. This method yields a prominent peak in the periodogram and compresses FFT side lobes, reducing the impact of spectral leakage from non-integer frequencies.

Addressing Peak Frequency Inaccuracy. We further mitigate peak frequency inaccuracy by determining interpolated peak frequencies. We use parabolic interpolation (Gasior and Gonzalez 2004) to approximate non-integer frequencies. To compute the interpolated peak frequency of a periodogram, we must determine the abscissa that illustrates the highest energy. Let the frequency bin with the highest amplitude be $\mathtt{k_m}$, and $\mathtt{S}[\mathtt{k}]$ be the energy of a frequency bin within the transform. Because the shape of a signal waveform resembles parabolas, we use the amplitude of the three frequency bins, $\mathtt{k_m} - 1$, $\mathtt{k_m}$, and $\mathtt{k_m} + 1$, to interpolate actual maximum amplitude and obtain interpolated peak frequencies.

Validating Time-Frequency Extraction To validate the correctness of our frequency extraction tool, we synthesize our own binaural beats tracks, which we know the ground truth of. Specifically, we leverage the Python module AccelBrainBeat to synthesize 200 binaural beat tracks (across different frequencies and track lengths). Our frequency extraction tool is able to accurately capture the beats produced within 0.05 Hz error. We also note that we find no impact of different frequencies and track duration on accuracy.

4.3 Expressing User Expectation Goals

To analyze a track, we have to map our discovered user expectations to said track. To do this, we apply a set of user expectation rules (rules hereafter) on a track's intent and/or time-frequency model to categorize tracks that do or do not conform to user expectations.

Goal Deviation (F1). Our thematic saturation (**RQ1**) demonstrates that users expect one of four mental state goals. A track title that deviates from these goals deceives listeners into believing promises not commonly associated with binaural beats. We define the F1 rule ($\mathtt{intent} = \mathtt{outlier}$) to flag tracks that deviate from an accepted goal and detect promises not commonly associated with binaural beats.

Goal Conflict (F2). Users expect that a single track can only achieve one goal. Thus, track titles that promise multiple goals contradict user expectations, thus deceiving them. We define F2 to flag tracks that have such conflicting goals, $\mathtt{intent} = \mathtt{multi\text{-}label}$.

Deceptive Absence (F3). Audio advertised as a binaural beats track must contain differing left and right peak frequencies. However, content creators may deceive users by providing misleading track titles to lure users into listening to the tracks. We define F3 to flag the tracks that contain 0Hz in its duration, $\forall \mathtt{t}\ \mathtt{f}(\mathtt{t}) = 0,\ \ \mathtt{t} \in [\mathtt{t_0}, \mathtt{t_n}]$.

Goal-Expectation Negligence (F4). We define a set of rules to check whether a track with a goal produces binaural beats that correspond with brainwaves users expect. The rules are expressed for each user expectation discovered on each mental state goal identified in Section 3. Table 1 presents the user expectations and their linear temporal logic (LTL) representation. LTL is a language that allows model checking where our intermediate data is expressed as a discrete model of time, thus allowing reasoning about the relative order of events (Piterman and Pnueli 2018). Expressing user expectations with LTL allows us to verify whether the extracted time-frequency model conforms to a set of expectations. We first map each frequency from a track's time-frequency model ($f(t)$) to the corresponding brainwave where w_t is the brainwave produced at time t, producing a time-brainwave model.

$$w(t) = w_t = \begin{cases} delta, 0.5 < f(t) \leq 4, \;\; theta, 4 < f(t) \leq 8, \\ alpha, 8 < f(t) \leq 12, \;\; beta, 12 < f(t) \leq 35, \\ gamma, 35 < f(t) \leq 42 \end{cases}$$

We then define a single or multiple rules for each mental state based on user expectations. To illustrate, for concentration, we define `C1` that states that users expect the track to contain a single period of beta waves. Similarly, for the sleep mental state, we define `S2` where users expect a sleep track's brainwave cycle to start with alpha waves and transition to theta waves before transitioning into delta waves and ends with either delta waves or theta waves.

Goal-Expectation Contradiction (`F5`). We define `F5` to flag tracks with a clear goal that conforms to expectations of a different mental state goal. For instance, a track with an intent of sleep that satisfies brainwaves expected for a different intent $((\Box \mathtt{s} = \mathtt{beta}) \vee (\Box \mathtt{s} = \mathtt{gamma}) \vee (\Box \mathtt{s} = \mathtt{alpha}) \vee (\Box \mathtt{s} = \mathtt{theta}))$ is flagged for contradicting expected brainwaves associated with sleep.

Verifying our Rule Validation Methodology. For each rule, we synthesize positive and negative samples using AccelBrainBeat. Positive samples refer to tracks that should be flagged for the specific rule, while negative samples refer to tracks that should not be flagged. We synthesize 20 positive and 20 negative samples for 40 samples per rule (200 samples for all rules). Our rule validation correctly flags all positive samples and validates all negative samples.

4.4 Conformance for a Binaural Beats Track

To process a track, we first extract the intended mental state from the track's metadata. We then process the track's audio to produce the time-frequency model. Intent and time-frequency models are provided as input for rule validation.

Each rule is checked consecutively. The first two rules (`F1-F2`) only operate on results of a track's intent extraction - $\mathtt{intent} = \mathtt{outlier}$ or $\mathtt{intent} = \mathtt{multi} - \mathtt{label}$ are flagged. The third rule (`F3`) only operates on the time-frequency model - we flag tracks where the binaural frequency is 0Hz for all timestamps $(\forall \mathtt{t}\ \mathtt{f(t)} = 0,\ \ \mathtt{t} \in [\mathtt{t_0}, \mathtt{t_n}])$.

For the fourth and fifth rules (`F4`, `F5`), we pass each rule's formula to an LTL API alongisde the time-brainwave model ($\mathtt{w(t)}$). The API returns a true/false flag noting whether the formula is satisfied. At most, one of `S2`, `R1`, `R2`, `C1`, `CE1` (Table 3) is satisfied (or none). We note that to satisfy `S2`, flags for `S2.1`, `S2.2` and `S2.3` need to be true. We check whether LTL rules associated with the extracted intent are true/false, e.g., a relaxation track should either adhere to `R1` or `R2`. If deception is flagged, we log the rules violated. We note specific libraries used in Appendix A.4 and performance details of our methods in Appendix A.5.

5 RQ2: Conformance of Online Platforms

To answer **RQ2**: Do binaural beats tracks on online platforms conform to commonly held user expectations?, we leverage our expectation-conformance methodology on over 7K binaural beats tracks from diverse sources.

5.1 Data Collection

We collect binaural beats tracks from diverse sources in three categories: (`C1`) 4,678 tracks from YouTube, (`C2`) 1,301 tracks on Spotify, and (`C3`) 1,070 tracks available for purchase through online e-commerce platforms and mobile apps. These categories allow us to comprehensively evaluate against diverse tracks at different platforms at different price points (free, subscription-based, and paid).

Cat.	Property	Max	Average	SD
C1	# of Views	27.43M	203K	1.11M
	Duration (mins)	732	122	170
	# of Likes	414K	2.1K	11.4K
	Fraction of Likes to Dislikes	1	0.94	0.08
C2	# Subscribers	470K	5.9K	29K
C3	Price Per Track	$10	$0.99	-
	Price Per Album	$15	$5	-

Table 2: Summary statistics of binaural beats tracks

For `C1`, we used YouTube's Data API (Google Developers). We perform search queries using the strings "binaural beats" and "binaural beat" while also including binaural beats-related search queries, obtained using the Google Trends API (Google Trends). We discard queries that are interrogative phrases such as "what are binaural beats?" to avoid collecting data that informs users about binaural beats. We obtained a list of 30 distinct search queries related to binaural beats tracks. We then performed a search for each query with YouTube's Data API and obtained the first 500 tracks returned by each query. We finalize a dataset of 4,678 tracks from 1,551 different YouTube channels.

We use the same search query set with the Spotify Developers API (Spotify) to collect the top 1,307 tracks for `C2`. To identify tracks in category `C3`, we use the same query set using the Google Search API and manually annotate returned results, verifying if they link to an e-commerce platform. This results in the top 6 binaural beats online stores - 3 of them provide their tracks via mobile apps. To consider the tracks found on these online stores (total of 1,070), we ensure they are obtained for a fee (per-track basis or per-album basis).

Dataset Properties. Table 2 summarizes dataset properties. `C1` allows for more metadata such as view counts, track category and likes/dislikes due to the expressive nature of the YouTube API (Google Developers) (See Appendix A.6). Each track has been listened to on average $\sim$ 200K times, with creators often classifying binaural beats tracks in categories outside "music" (e.g., entertainment). Additionally, 50% of the videos are more than 60 minutes long, while the average duration is 120 minutes. Moreover, 90% of the tracks have a fraction of likes (Likes/(Likes+Dislikes)) greater than 0.9. The high view count and like-dislike fraction demonstrate the popularity of such binaural beats tracks. However, it is important to note that popularity does not necessarily mean that users are not deceived - deception, as we define it, refers to when tracks deliver what users do not expect, and users do not always know when they are deceived (when a track does not contain promised beats), as echoed in **RQ1**.

Engagement statistics are not available for `C2` and `C3` (from the Spotify API or online stores respectively). However, `C2` tracks are from 320 top binaural beats artists who have a combined following of ~1.89 million monthly listeners. The average artist has 5.9K subscribers, supporting Spotify's business model of allowing smaller artists to access a streaming platform. `C3` tracks are available for purchase at

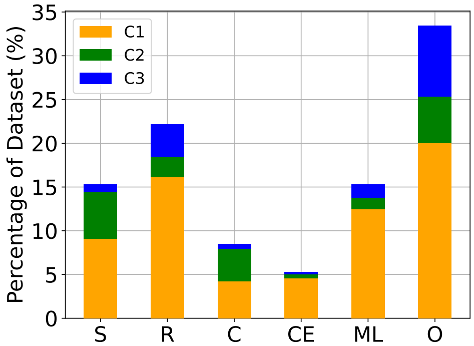


Figure 5: Percentage of track intents. (**S**: Sleep, **R**: Relaxation, **C**: Concentration, **CE**: Cognitive Enhancement, **ML**: Multi-Label, **O**: Outlier)

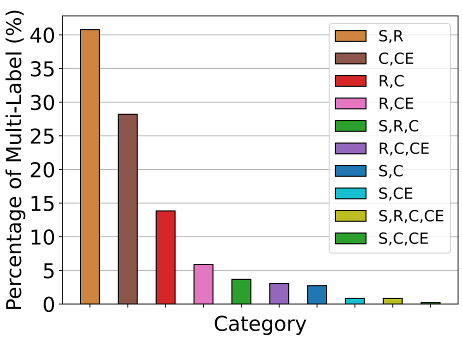


Figure 6: Fine-grained breakdown for multi-labeled tracks.

different price points. Albums (collection of tracks) average around $5 USD. Individual tracks average $0.99 per track.

5.2 Intent Extraction

Figure 5 shows the percentages of tracks classified into the six classes; (aggregate of `C1`, `C2`, `C3`). Tracks that claim to induce an intent that aligns with user expectations (sleep, relaxation, concentration, cognitive enhancement) account for 51.2% (3,602) of the dataset. Among these tracks, 43% (1,558) are classified as relaxation, and 30% (1,077), 17% (596), 10% (373) of the tracks are classified as sleep, concentration, and cognitive enhancement, respectively.

Interestingly, we identified that 49.8% (3,532) of tracks are classified into multi-label and outlier classes. Particularly, 15.34% (1,055) of the tracks are classified as multi-label that contain words from multiple lexicons (e.g., both sleep and concentration), and 33.4% (2,321) of the tracks are classified as outliers that do not contain any words from our lexicons.

Analysis of Multi-labeled Tracks. We analyze each multi-labeled track's title to understand their intent. Figure 6 illustrates the different combinations of accepted intent keywords included in multi-labeled track titles. For instance, we observe 40.78% (430) of the tracks claim to aid in sleep and relaxation, 28.2% claim to help with concentration and cognitive enhancement and 3.67% of the tracks claim to induce sleep, relaxation and concentration.

Analysis of Outlier Tracks. We perform k-means clustering on titles of outlier tracks to better understand their claims.

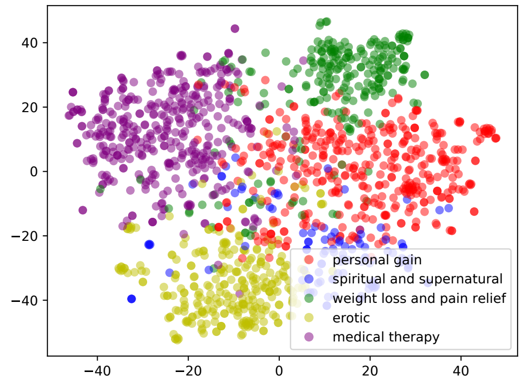


Figure 7: Different deceptive clusters in outlier tracks.

To pre-process the track titles for clustering, we tokenize titles into words and remove tokens such as non-English words, stop words, punctuation, numbers, and generic words (e.g., "binaural beats", "brain", and "entrainment") as they are shared in titles but do not contribute to their intent. We found that 8.7% of outlier tracks do not include any words in their titles after pre-processing (e.g., "Binaural Beats Brain Entrainment"). We eliminate such tracks from further processing. For the remaining tracks, we obtain the vector representation for each token in the pre-processed titles through Glove (Pennington, Socher, and Manning 2014), an unsupervised learning algorithm to create vector representations, where similar words have a similar vector representation.

Figure 7 overviews a visualization of the five clusters. Two authors manually inspected each cluster to determine the titles' claims. They assigned the following labels based on the words each cluster title contains: *personal gain*, *spiritual and supernatural phenomena*, *weight loss and pain relief*, *erotic*, and *medical therapy*. To detail, weight loss and pain relief tracks (15.8%) claim to help listeners with body fat reduction, weight loss and relieving body aches. Personal gain tracks (25.4%) claim to aid in financial growth, prosperity and happiness. Spiritual and supernatural tracks (6.9%) claim to induce psychic abilities, clairvoyance and spiritual clarity. Medical therapy tracks (24.9%) recommend users alternative medical therapy for common diseases such as kidney pain and flu. Lastly, 18.3% of outlier tracks are categorized as erotic as they claim to help listeners improve their libido and sexual performance. We additionally found overlapping clusters that shows track titles including multiple words from different clusters (e.g., the track title "Subliminal Programming: Attract Wealth And Sexy People With Binaural Beats" overlaps with erotic and personal gain clusters). Our findings show that track creators mislead listeners with outlandish claims for monetization and channel expansion purposes.

5.3 Deceptive Tracks

Table 3 presents the results of deception detection as an aggregate across `C1`, `C2` and `C3` (category specific breakdowns are available in Appendix Tables 1, 2, 3). Our analysis uncovers a large amount of tracks have deceptive track metadata. All outlier and multi-label tracks are flagged for `F1` (Goal Deviation) and `F2` (Goal Conflict), respectively, as stated by their definitions. 33.4% of all tracks make claims not accepted

Track Intent	F1	F2	F3	F4	F5
Outlier	100 %	✗	6.8%	✗	✗
Multi-Label	✗	100 %	2.13%	✗	✗
Sleep	✗	✗	1.76%	75.3%	3.06%
Relaxation	✗	✗	1.99%	92.43%	17.78%
Concentration	✗	✗	1.01 %	82.05%	13.26%
Cognitive Enhancement	✗	✗	1.88%	90.35%	49.33%
Total	33.4 %	15.3%	3.5%	43.7%	8.1%

Table 3: Distribution of deception detected (C1,C2,C3)

among binaural beats listeners (F1). Interestingly, tracks in C3 (paid services) present the largest proportion of this deception, suggesting that exploitative creators make outlandish claims to entice users for monetary benefit. 15.3% of tracks promise listeners the ability to achieve multiple states (F2), with C1 showing the highest percentage (18.5%). 3.5% of tracks falsely advertise that a track contains binaural beats, when in reality, the left and right channels produce the same frequencies (F3, Deceptive Absence).

Alarmingly, most tracks with a clear goal do not produce binaural beats associated with the said goal (F4, Goal-Expectation Negligence). At least 70% of all sleep, relaxation, concentration, and cognitive enhancement tracks are flagged for F4, indicating many creators do not meet users' expectations for these goals. Finally, 8.1% of all tracks are flagged for F5 (Goal-Expectation Contradiction). 2.20%, 20.03%, 18.60% and 51.66% of sleep, relaxation, concentration and cognitive enhancement tracks respectively produce binaural beats associated with a contradicting state (not their own state). 92% of all F5 flagged tracks satisfy sleep expectations (and have an intent other than sleep).

We find no significant percentage differences across F3, F4 and F5 between the three categories (C1–C3). However, we observe that paid tracks (C3) have the smallest proportion of F4 (35.04%), indicating that when these creators advertise valid mental wellbeing goals, they are more likely (than C1 and C2) to deliver the associated brainwaves.

Our results indicate that creators deceive users either by deceptive track metadata, audio that does not conform to user expectations, and in some cases, fail to produce binaural beats. We also discovers that deception is equally present on all platforms, regardless of purchase prices (free or paid).

Engagement in Deceptive Tracks. Although C2/C3 tracks do not provide interaction statistics between users and tracks, we present statistics for C1, given that the Youtube API provides view counts, number of comments and likes. Our analysis uncovers that engagement across deceptive tracks is high. Deceptive tracks are viewed an average of 245K times, with the highest viewed video having 340K views. The average views for tracks flagged with F1 is 206K. Other deceptive tracks have similar high-average views, with F2 at 321K, F3 at 40K, F4 at 270K, F5 at 153K.

Multi-label tracks have the highest average view count of 321K. Noticeably, tracks with outlandish claims (F1) have the highest average number of comments (289). The high average fraction of likes (0.95) suggests that listeners may be unaware that they are being deceived. Moreover, promotion of deceptive tracks is a widespread problem, with these tracks coming from 1403 unique channels.

Engagement in Tracks Free of Deception. Only 6.7% of binaural beats tracks in our dataset are not flagged for deceptions. For C1, the average number of views for such tracks is 172K (max of 830K), 73K lower than for deceptive tracks. This suggests that without knowledge of the binaural beats present in a track, users unknowingly consume deceptive tracks over tracks that conform to commonly held user expectations. The lower average number of comments (217 [max of 25K]) and average fraction of likes (0.94), when compared to deceptive tracks, also support this.

These tracks come from 147 unique channels out of 1,551. Many of their channel homepage descriptions are expressive, detailing their effort to carefully produce tracks (e.g., *"We are musicians and neurologists that help listeners."*, *"This channel is dedicated to the creation of pure binaural tones."*). Tracks found on these channels also have informative track descriptions, include references to information of brainwaves and explanations of effects (e.g., *"Beta waves will help to concentrate"*, *"This session contains frequencies which will greatly assist with Brain Power."*), and guidance for users (e.g., *"For Best Results Please Wear Headphones"*, *"Listen to this music in a quiet place for at least 20/25 minutes daily"*). Similarly, all safe tracks in C3 belong to a platform that runs a blog explaining binaural content in depth.

6 Discussion

6.1 Practical and Design Implications

Expectations through Online Communities. Our work bears implications for leveraging online resources to understand community expectations. We identify goals and expectations users have. Given the plethora of mental wellbeing technologies (e.g., light-based therapy, guided meditation), each community may form its own set of expectations.

Traditional methods for understanding user expectations such as focus groups, interviews, or surveys, face challenges, e.g., recruitment difficulties (Atkinson and Flint 2001). Therefore, our methodology serves as a reliable alternative given that we draw from *publicly* accessible resources (e.g., online forums). Future work on mental wellbeing communities can leverage such methodology to ground users' expectations and concerns about existing tools/technology.

Designing Mental Wellbeing Verification Tools. In **RQ2**, we expose that the failure of tracks to align with user expectations is common in various media (e.g., Table 3). We also expose attractive titles that can mislead users (Figure 7). Thus, future efforts should design verification tools that evaluate different wellness technologies. Our spectral analysis methodology can be easily extended to other sound-based therapy (e.g., music therapy, isochronic tones, monoaural beats). However, it is crucial to note that accommodating other types of mental wellbeing technology comes with its own challenges. For instance, metadata analysis may need to extend beyond track titles alone (e.g., descriptions). Similarly, verification efforts for visual-based therapy would require mapping raw visual data to an intermediate representation and collaboration with domain experts. While preliminary work shows tools leveraging user expectations can be beneficial (Arunasalam et al. 2024), further research effort is required to produce usable tools that instill user trust.

Multi-Stakeholder Involvement. We argue that stakeholders of digital mental wellbeing technologies/apps should be proactively involved in ensuring users are given access to these tools. Here, we note that regardless of the true scientific impact of binaural beats, users have expectations surrounding binaural beats. The presence of commonly expected frequencies and titles that abide by commonly expected goals are pertinent information for users who consume binaural content to know. Thus, tools to verify binaural content is important for user trust. Additionally, these tools can be leveraged by creators to ensure they abide by users expectation.

To that end, third-party apps/services can be designed to provide information on a track. Such services should support cross-app integration, i.e., tools should have access to a high-quality version of audio. However, developers behind mental wellbeing apps may be apprehensive about granting data access to third parties. An alternative would be for app developers to embed tools in their interface, detailing what binaural beats are present throughout the track to users.

Additionally, mental wellbeing content is also often hosted on content-sharing platforms (e.g., YouTube). We argue that platforms should play an active role in implementing verification. Verification can be integrated into the framework that creators use, similar to fake news interventions such as credibility indicators (Lu et al. 2022). To illustrate, given users find titles informative, platforms can enforce title vetting to ensure outlandish claims are avoided.

6.2 Theoretical Implications

Our research applies User-Centered Design (Tenner 2015) and the Technology Trust Framework (Mcknight et al. 2011). In leveraging online communities, we discover *what* users expect and *how* these expectations are shaped. Users also ground their reasoning to external factors (e.g., self-conducted online research).

We argue that mental wellbeing solutions (e.g., verification tools) grounded in UCD and the Technology Trust Frameworks should consider expectations and *how* users develop their expectations. Understanding the *how* can aid in the design process. To illustrate, tools leveraging user expectations to verify mental wellbeing technologies can provide pointers to resources (e.g., scientific papers) to instill user trust. Similarly, solutions should also cater to novice and non-users, who may not begin using digital wellbeing tools with grounded expectations (e.g., to help novice users, content websites can include pop-up notifications indicating that a track does not conform to most users' established expectations).

6.3 Ethical Implications

It is possible to encapsulate our expectation-conformance methodology within a tool to verify binaural beats tracks. However, these expectations may not apply to everyone. For instance, a user may have niche expectations. Thus, for such tools to be deployed in the real world, great caution is needed to carefully curate user-expectation rules.

6.4 Limitations and Future Work

The first limitation of our work is that our qualitative analysis is based on publicly available forums and/or websites. Thus, the user expectations we extract may not necessarily reflect those discussed in more private communities. Second, it is reasonable to assume that new expectations may emerge amongst binaural beats listeners. Our current methodology does not capture evolving user expectations. However, our methodology can be easily updated to add user expectation rules without significant overhead.

Given our study's implications on verification tools and informing users, we hope to extend future work to add to the broader impact of our study. First, we aim to apply our methodology to additional frequency based mental wellness technology. Second, we aim to investigate design choices on how to inform users when commonly held expectations are not met. Third, we hope to investigate the impact of poor expectation conformance on user consumption of content (e.g., how reports of lack of conformance may negatively impact their consumption of mental wellbeing technology).

7 Conclusions

Our work grounds rules for what binaural beats tracks should adhere to via user expectations. We draw from online resources to conduct thematic analysis to deepen our understanding of users' expectations. We then evaluate over ~7K binaural beats tracks available on popular streaming platforms and identify deceptive tracks, applying our methodology to verify if a given track conforms to listeners' expectations. We extract a track's mental state intent and time-frequency model before applying rule validation to flag a track for deception. Our findings provide a future direction for designing systems for users of wellbeing technologies.

References

344 Audio. 2023. How to make binaural beats. https://www.344audio.com/post/article-how-to-make-binaural-beats.

Arunasalam, A.; Tong, J.; Farrukh, H.; Ozmen, M. O.; Saha, K.; and Celik, Z. B. 2024. Expectation Conformance in Online Sound Therapy: Designing Tools for Users of Mental Wellbeing Applications. In *Companion Publication of CSCW*.

Atkinson, R.; and Flint, J. 2001. Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social research update*.

Becher, A.-K.; Höhne, M.; Axmacher, N.; Chaieb, L.; Elger, C. E.; and Fell, J. 2015. Intracranial electroencephalography power and phase synchronization changes during monaural and binaural beat stimulation. *European Journal of Neuroscience*.

Boot, W. R.; Simons, D. J.; Stothart, C.; and Stutts, C. 2013. The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on psychological science*.

BrainTap. 2021. BrainTap offers daily guided meditations for unique individuals. https://braintap.com/.

Brigham, E. O. 1988. *The fast Fourier transform and its applications*. Prentice-Hall, Inc.

Bschor, T.; Nagel, L.; Unger, J.; Schwarzer, G.; and Baethge, C. 2024. Differential Outcomes of Placebo Treatment Across 9 Psychiatric Disorders: A Systematic Review and Meta-Analysis. *JAMA psychiatry*.

Carskadon, M. A.; Dement, W. C.; et al. 2005. Normal human sleep: an overview. *Principles and practice of sleep medicine*.

Colom, A. 2022. Using WhatsApp for focus group discussions: Ecological validity, inclusion and deliberation. *Qualitative Research*.

Cornet, V. P.; Toscos, T.; Bolchini, D.; Ghahari, R. R.; Ahmed, R.; et al. 2020. Untold stories in user-centered design of mobile health: practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. *JMIR mHealth and uHealth*.

Daudén Roquet, C.; and Sas, C. 2018. Evaluating Mindfulness Meditation Apps. In *Extended Abstracts of ACM CHI*.

De Choudhury, M.; and De, S. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *AAAI ICWSM*.

Digital Music News. 2020. Spotify Says Mental Health Is One of Its Fastest-Growing Genres. https://tinyurl.com/4rppxs8r.

empeek. 2020. Mobile Users Spent $195 Million in Meditation Apps Last Year, Up 52% Over 2018. https://empeek.com/insights/what-you-should-know-about-creation-of-a-meditation-application-like-calm-and-headspace-features-steps-and-costs.

Fiesler, C.; Beard, N.; and Keegan, B. C. 2020. No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. In *AAAI ICWSM*.

Fox, S.; and Fallows, D. 2003. Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access. *Pew Internet & American Life Project*.

Fugard, A. J.; and Potts, H. W. 2015. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. *International Journal of Social Research Methodology*.

Gasior, M.; and Gonzalez, J. 2004. Improving FFT frequency measurement resolution by parabolic and Gaussian spectrum interpolation. In *American Institute of Physics*.

GitHub. 2020. mvcisback/py-metric-temporal-logic. https://github.com/mvcisback/py-metric-temporal-logic.

Google Developers. 2020. YouTube Data API : Google Developers. https://tinyurl.com/4bfmuyvf.

Google Search. 2021. Google Search. https://pypi.org/project/googlesearch-python/.

Google Trends. 2021. Google Trends. https://trends.google.com/trends.

Harris, F. J. 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*.

Hinchcliffe, V.; and Gavin, H. 2009. Social and virtual networks: Evaluating synchronous online interviewing using instant messenger. *The Qualitative Report*.

Hwang, S.; and Foote, J. D. 2021. Why do people participate in small online communities? *ACM Human-Computer Interaction*.

Landau, H. 1967. Sampling, data transmission, and the Nyquist rate. *Proceedings of the IEEE*.

Larsen, M. E.; Huckvale, K.; Nicholas, J.; Torous, J.; Birrell, L.; et al. 2019. Using science to sell apps: evaluation of mental health app store quality claims. *NPJ digital medicine*.

Lu, Z.; Li, P.; Wang, W.; and Yin, M. 2022. The Effects of AI-based Credibility Indicators on the Detection and Spread of Misinformation under Social Influence. *ACM CSCW*.

Lyon, D. A. 2009. The discrete fourier transform, part 4: spectral leakage. *Journal of object technology*.

McDonald, N.; Schoenebeck, S.; and Forte, A. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *ACM CSCW*.

Mcknight, D. H.; Carter, M.; Thatcher, J. B.; and Clay, P. F. 2011. Trust in a specific technology: An investigation of its components and measures. *ACM TMIS*.

Miller, G. A. 1995. WordNet: a lexical database for English. *Communications of the ACM*.

NumPy. 2021. Discrete Fourier Transform. https://numpy.org/doc/stable/reference/routines.fft.html.

NY Times. 2023a. Wellness Products False Claims. https://www.nytimes.com/2023/07/26/well/live/wellness-products-false-claims.html.

NY Times. 2023b. Wellness Scams. https://www.independent.co.uk/life-style/wellness-scams-goop-pseudoscience-b2260871.html.

Oster, G. 1973. Auditory beats in the brain. *Scientific American*.

Palgi, Y.; Shrira, A.; Ring, L.; Bodner, E.; Avidor, S.; et al. 2020. The loneliness pandemic: Loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. *Journal of affective disorders*.

Papadamou, K.; Papasavva, A.; Zannettou, S.; Blackburn, J.; Kourtellis, N.; et al. 2020. Disturbed YouTube for kids: Characterizing and detecting inappropriate videos targeting young children. In *AAAI ICWSM*.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *EMNLP*.

Piterman, N.; and Pnueli, A. 2018. Temporal logic and fair discrete systems. *Handbook of Model Checking*.

Reese, K.; Smith, T.; Dutson, J.; Armknecht, J.; Cameron, J.; and Seamons, K. 2019. A usability study of five {two-factor} authentication methods. In *USENIX SOUPS*.

Riemer, K.; Ciriello, R.; Peter, S.; et al. 2020. Digital contact-tracing adoption in the COVID-19 pandemic: IT governance for collective action at the societal level. *European Journal of Information Systems*.

Ross, B.; Miyazaki, T.; Thompson, J.; Jamali, S.; and Fujioka, T. 2014. Human cortical responses to slow and fast binaural beats reveal multiple mechanisms of binaural hearing. *Journal of neurophysiology*.

Saha, K.; and Sharma, A. 2020. Causal factors of effective psychosocial outcomes in online mental health communities. In *AAAI ICWSM*.

Saha, K.; Torous, J.; Caine, E. D.; and De Choudhury, M. 2020. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *JMIR*.

Singh, S.; Kaushal, R.; Buduru, A. B.; and Kumaraguru, P. 2019. KidsGUARD: Fine grained approach for child unsafe video representation and detection. In *ACM/SIGAPP Symposium on Applied Computing*.

Social Life. 2014. The social life of health information. https://www.pewresearch.org/short-reads/2014/01/15/the-social-life-of-health-information/.

Spotify. 2022. Spotify Developers API. https://developer.spotify.com/.

Tenner, E. 2015. The design of everyday things by donald norman. *Technology and Culture*.

Turner, M.; Schmitz, M.; Bierey, M. M.; Khamis, M.; and Marky, K. 2023. Tangible 2FA–An In-the-Wild Investigation of User-Defined Tangibles for Two-Factor Authentication. In *Symposium on Usable Privacy and Security*.

Vial, S.; Boudhraâ, S.; and Dumont, M. 2022. Human-centered design approaches in digital mental health interventions: exploratory mapping review. *JMIR Mental Health*.

Wu, Z.; and Palmer, M. 1994. Verb semantics and lexical selection. *arXiv preprint cmp-lg/9406033*.

YouTube. 2023. How to make binaural beats under 5 minutes. https://www.youtube.com/watch?v=hfjLuqOK9Co.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, please see Ethical Statement (below) and Discussion
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, please see Section 3.1 and Section 4
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, please see Ethical Statement (below), Section 3.1 and Section 4
   (e) Did you describe the limitations of your work? Yes, please see Discussion (Section 6)
   (f) Did you discuss any potential negative societal impacts of your work? Yes, please see Discussion (Section 6)
   (g) Did you discuss any potential misuse of your work? NA
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, please see Ethical Statement (below)
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, we perform unsupervised ML (k-means) in RQ2, and thus provide code – see https://osf.io/ks4yv/
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes, see Implementation in Appendix
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, see Section 5.2. We note that K-means clustering is only used to provide qualitative insight into different types of deceptive tracks.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, existing code is cited throughout the paper and in Appendix (Implementation)
   (b) Did you mention the license of the assets? No, we use standard libraries, specifying license would be unnecessary.
   (c) Did you include any new assets in the supplemental material or as a URL? Yes, see https://osf.io/ks4yv/
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? See Ethical Statement below.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR ? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA

Ethical Statement

To minimize identification risk for online users, we take steps to preserve privacy for collected online communication data (public forums and blog entry data). We limited the collection of Personally Identifiable Information (PII) and made no efforts to collect user-specific information such as usernames, locations, and account metrics (e.g., posting frequency). Data was stored in an encrypted cloud service and quotes presented in this paper contain no PII.

We also consider the ethics of data collection and analysis from APIs (e.g., YouTube and Spotify). Here, we note that Spotify, in its terms of service, prohibits use of data for ML training. To address this, we draw on guidelines presented by Fiesler et al (Fiesler, Beard, and Keegan 2020) regarding ethical data collection. Fiesler et al. suggests that restricting data collection to situations permitted by terms of service is flawed as it implies that abiding by terms of service makes collection ethical by default. Instead, they argue that collection should abide by the Menlo and Belmont Report – which provides guidelines for ethical research practice. We extend this argument to our work. We use data for unsupervised ML and our collection and analysis closely align with ethics laid out in the report. We cluster publicly available text information to analyze data, with no commercial intent. We therefore argue that our data collection and analysis is ethical.

Our research members come from diverse backgrounds, conducting interdisciplinary research in HCI, CSS, and Security & Privacy. Our goal is to measure whether online tracks conform to user expectations. After consulting neuroscience researchers in academia and industry, we discovered that the science behind binaural beats' ability as mood enhancers in the human brain is mercurial and constantly evolving. Therefore, we do not claim to evaluate the efficacy of binaural beats to induce different mental states in the brain.

A Appendix

The subsequent sections pertain to (1) deceptive titles, (2) keywords leveraged in data collection, (3) findings in user expectation analysis, (4) implementation and (5) performance of user-expectation conformance methods, (6) detailed overview of C1 tracks and (7) fine-grained evaluation results.

A.1 Outlandish Deceptive Claims

Figure 1 shows the binaural beats track "Cleanse Infections & Get Rid Of Virus, Bacteria, Fungal- Dissolve Toxins", that makes an outlandish claim to deceive users. Similarly, Figure 2 presents an advertised binaural track for sexual arousal.

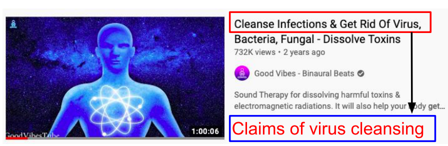


Figure 1: Deceptive claim of virus cleansing

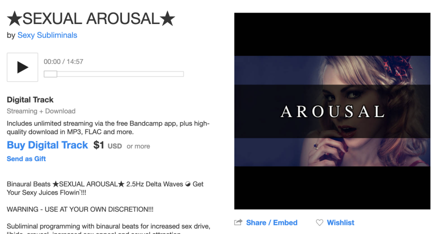


Figure 2: Sample landing page listeners are redirected to purchase a track with the claim of improving arousal.

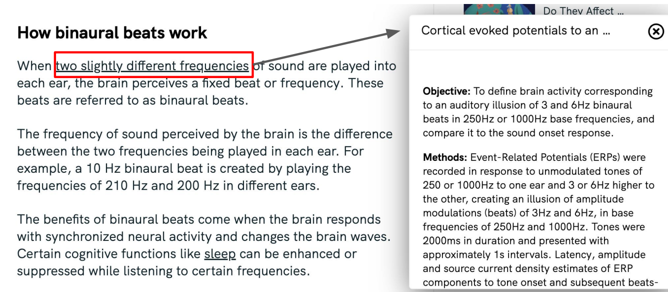
A.2 Keywords Used in Crawling

We list all keywords used in the Google Search API crawler.

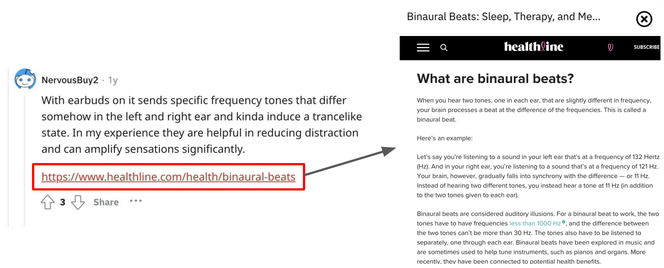
- sleep binaural beats
- binaural beats youtube
- binaural beats music
- meditation binaural beats
- download binaural beats
- binaural beats free
- best binaural beats
- healing binaural beats
- theta binaural beats
- focus binaural beats
- do binaural beats work
- binaural beats for sleep
- binaural beats app
- what are binaural beats
- alpha binaural beats
- binaural beats delta
- binaural beats sleep music
- binaural beats anxiety
- binaural beats lucid dreaming
- binaural beats meaning

A.3 Research Driven User Expectation/Beliefs

Figure 3a and Figure 3b illustrate blogs and forums that reference user-trusted resources concerning binaural beats.



(a)



(b)

Figure 3: (a) Blog content that references a peer reviewed paper. (b) Binaural beats related reddit thread that references a medical website (Healthline).

A.4 Implementation

Our implementation for expectation conformance leveraged our custom python implementation and python libraries/modules.

Intent Extraction. We use the NLTK library to process track titles. Titles are fed sequentially to `word_tokenize`, `stopwords`, and `stem.snowballstemmer` for tokenizing the text into words, removing stopwords, and stemming via the Porter Stemmer algorithm respectively. To compute word similarity with the WordNet taxonomy, we use `wordnet.synset` to generate the synset groupings of a word, and `wordnet.synset.wup_similarity` to compute the synset-mental state Wu-Palmer similarity.

Time-Frequency Extraction. We perform spectral analysis by leveraging the `rfft` function from the NumPy Fast Fourier Transform library (NumPy) to compute the Fourier Transform of the audio data. We extend the `Scipy.Signal` function to generate a Blackman-Harris window to address spectral leakage when computing the Discrete Fourier Transform. For the parabolic interpolation for peak frequencies, we use our implementation of the algorithm in (Gasior and Gonzalez 2004).

Rule Validation. Rules F1−F3 are validated using our custom Python implementation. We compare extracted intent labels for Rule F1 and F2, checking if the output label is `outlier` or `multi − label`. Rule F3 is validated on the time-frequency model. Each timestamp is checked for 0 Hz. We use the py-metric-temporal-logic python module (GitHub) to validate F4 and F5. We use the string-based API to define each stage of F4 rules (e.g., S2.1−S2.3). We then pass the string as a parameter with the track's time-frequency model and intent to validate the rules.

A.5 Performance

We study the end-to-end time latency of intent extraction, spectral analysis and rule validation. We run our experiments on a laptop computer (1.4GHz 4-core Intel i5 processor and 16GB RAM) using Python 3.8. We record the time taken to process ten 30 min and ten 60 min tracks on our laptop computer. The average time for a 30 min track is 7 sec. The time-frequency model extraction and rule validation take on average 3.8 sec and 3 sec. The average time for a 60 min track is 20 sec. Time-frequency model extraction and rule validation take on average 7.3 and 12 sec. We observe that runtime overhead increases linearly with the track duration. Intent extraction and user interface generation take on average less than a second. We found that fine-grained FFT for time-frequency model extraction and `F4`, `F5` rule validation through the py-metric-temporal-logic library are the bottlenecks. To address this, we implemented our custom validation algorithm for `F4`, `F5` rules and reduced the overhead of rule validation to an average of 2 sec.

A.6 Detailed Overview of C1 Tracks

We present a detailed overview of the `C1` tracks' statistics. More specifically, Figure 4 shows a histogram of the tracks' view count. Figure 5 shows the distribution of track categories – creators often classify binaural beats tracks in categories outside "music" (e.g., entertainment). Figure 6 shows the distribution of track durations. Lastly, we show the distribution of the fractions of likes to dislikes (Likes/(Likes+Dislikes)) in Figure 7.

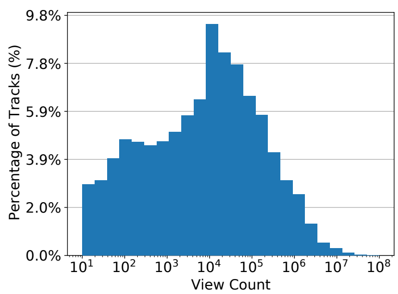


Figure 4: Track view counts (`C1` metadata)

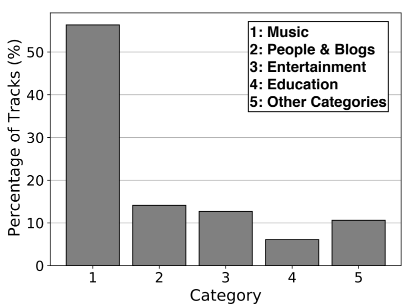


Figure 5: Category descriptors in `C1` metadata

A.7 Non-Conforming Tracks

Table 1, Table 2, Table 3 present category specific breakdowns for tracks that do not conform to commonly held user expectations.

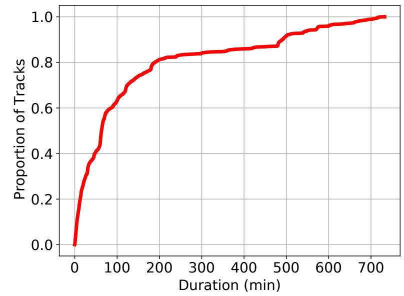


Figure 6: CDF of track duration (`C1` metadata)

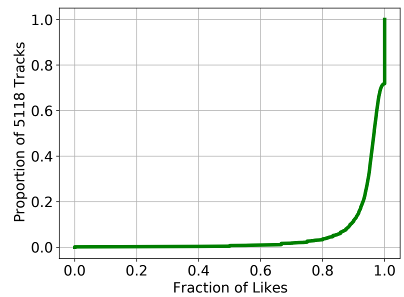


Figure 7: CDF of likes to dislikes (`C1` metadata)

Track Intent	**F1**	**F2**	**F3**	**F4**	**F5**
Outlier	100 %	✗	9.39%	✗	✗
Multi-Label	✗	100 %	2.14%	✗	✗
Sleep	✗	✗	1.46%	74.23%	2.2%
Relaxation	✗	✗	2.24%	92.44%	20.03%
Concentration	✗	✗	1.83 %	92.99%	18.6%
Cognitive Enhancement	✗	✗	2.11%	90.63%	51.66%
Total	31.4 %	18.5%	4.3%	44.07%	9.67%

Table 1: Distribution of deception detected in `C1`

Track Intent	**F1**	**F2**	**F3**	**F4**	**F5**
Outlier	100 %	✗	1.88%	✗	✗
Multi-Label	✗	100 %	3.12%	✗	✗
Sleep	✗	✗	2.41%	73.8%	4.55%
Relaxation	✗	✗	2.47 %	86.42%	14.2%
Concentration	✗	✗	✗%	68.06%	7.6%
Cognitive Enhancement	✗	✗	✗%	82.86%	48.57%
Total	28.6 %	7.36%	1.77%	47.89%	5.91%

Table 2: Distribution of deception detected in `C2`

Track Intent	**F1**	**F2**	**F3**	**F4**	**F5**
Outlier	100 %	✗	1.93%	✗	✗
Multi-Label	✗	100 %	✗	✗	✗
Sleep	✗	✗	✗	93.75%	3.12%
Relaxation	✗	✗	0.38%	98.08%	8.08%
Concentration	✗	✗	✗	100%	5.41%
Cognitive Enhancement	✗	✗	✗	100%	5.56%
Total	53.9 %	10.1%	1.14%	35.04%	2.46%

Table 3: Distribution of deception detected in `C3`