Preprint, Under Review at *The Journal of the Acoustical Society of America* (2021). Copyright © 2021 Arctop Inc. All Rights Reserved.

Differences In The Effects On Human Focus Of Music Playlists And Personalized Soundscapes, As Measured By Brain Signals.

Aia Haruvi,[1] Ronen Kopito,[1] Noa Brande-Eilat,[1] Shai Kalev,[1] Eitan Kay,[1] Dan Furman.[1]

[1] *Arctop, Research & Development, Kaufmann St. 4, Tel Aviv-Yafo, 6801296, Israel.*

ABSTRACT

The goal of this study was to learn what properties of sound affect human focus the most. Participants (N=62, 18-65y) performed various tasks while listening to either no background sound (silence), popular music playlists for increasing focus (pre-recorded songs), or personalized soundscapes (audio composed in real-time to increase a specific individual's focus). While performing tasks on a tablet, participants wore headphones and brain signals were recorded using a portable electroencephalography headband. Participants completed four one-hour long sessions, each with different audio content, at home. We successfully generated brain-based models to predict individual participant focus levels over time and used these models to analyze the effects of various audio content during different tasks. We found that while participants were working, personalized soundscapes increased their focus significantly above silence (p=0.008), while music playlists did not have a significant effect. For the young adult demographic (18-36y), silence was significantly less effective at producing focus than audio content of any type tested (p=0.001-0.009). Personalized soundscapes enhanced focus the most relative to silence, but professionally crafted playlists of pre-recorded songs also increased focus during specific time intervals, especially for the youngest audience demographic. We also found that focus levels can be predicted from physical properties of sound, enabling human and artificial intelligence composers to test and refine audio to produce increases or decreases in listener focus with high temporal (millisecond) precision. Future research includes real-time adjustment of sound for other functional objectives, such as affecting listener enjoyment, calm, or memory.

I. INTRODUCTION

> "Now I will do nothing but listen …
> I hear all sounds running together,
> combined, fused or following,
> Sounds of the city and sounds out of the city,
> sounds of the day and night …
> I hear the key'd cornet, it glides quickly in through my ears,
> It shakes mad-sweet pangs through my belly and breast.
> I hear the chorus, it is a grand opera,
> Ah this indeed is music - this suits me."[1]
>
> WALT WHITMAN, *Song of Myself*

For most of human history our habitat has been defined sonically by silence punctured only by sounds of the natural world: babbling brooks, bluebirds, thunder, human voices and music-making tools.[2,3] The soundscape, or aural landscape i.e. the acoustical environment our human ancestors perceived and lived in, was relatively consistent from generation to generation. Since the 19th century however – with the advent of industrial machines – that has changed dramatically. We still live amongst silence and natural sounds, but also ambient noise, amplified electronic music and an abundance of digital audio content available "on-demand."[4] With such an expansion in humanity's modulation of the auditory world, it is fair to say that as a species we have begun to cause a "shift in the sensorium."[5]

A. Effects of sound on human experience

Most people can speak from experience to the fact that certain sounds and arrangements of sounds (like music or sound effects) can be pleasant, reduce stress, increase motivation, and more.[6,7,8] Sounds can, of course, do the opposite as well.[9,10,11] Many scientific studies have explored the relationship between sound, music and humans from an objective perspective that seeks to analyze properties of audio that correlate with specific emotions or particular attentional response in humans. For example, Cheung et al[12] found that pleasure from music depends on states of expectation, such as a skipped rhythmic beat which can be pleasurable or discomforting depending on the listener's circumstance. *Sweet Anticipation*[13] similarly maps how music evokes emotions within a theory of expectation and describes psychological mechanisms responsible for our mixture of responses to auditory stimuli.

Within the field of affective computing,[14] sound has been studied increasingly for its ability to rapidly affect people's emotional and attentional state[15,16,17,18,19] A finding that reappears across studies is that the difference between 'noisy' and 'beautiful' sounds to human is, indeed like other aesthetic preferences, subjective and largely in the 'ear of the listener.' But only up to a point: psychophysical thresholds exist and there are clearly natural laws governing much of the way humans hear and experience sound.[20,21,22]

One of the most promising approaches to studying the impact of sound on humans has been combining audio content feature analysis (based on a sound's physical properties), with measures of human experience (such as Self-Assessment Manikin surveys on arousal and valence after listening).[23,24] Using properties of sound as features and experience measures as labels for those features, several groups have attempted to build machine learning algorithms that can predict emotional responses based on the sound properties alone, commonly according to a valence-arousal circumplex model.[25,26,27,28,29] However results in this area remain mixed due to lack of sufficiently high dimensional measurement and modeling tools suitable for capturing the fast changes in human experience that accompany changes in sound.[30,31]

B. Attention and emotion decoding from brain signal

Neuroscientific research into the basis of human attention and emotion traditionally has relied on functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies which enjoy high spatial resolution, but a low temporal resolution on the order of seconds.[32,33,34,35] Because processes within the brain, along with human behavior itself, occur at the sub-second range, this magnitude of time resolution has been inadequate for addressing the most pressing questions. Additionally, the field has been riddled by the notion that stimuli can be assumed to have the same affective valence for all participants; furthermore, that subjecting participants to artificial, laboratory conditions in order research these "same" valences does not fundamentally affect listener experience and skew results.[36,37,38,39]

Recent developments in portable brain sensors and wearable devices that are affordable, comfortable, and easy to use have enabled neuroscientific data acquisition to be done "in-the-wild" at a mass scale, making it possible for the first time to measure brain responses seamlessly from diverse audiences within their natural, dynamic real-world environments.[40,41,42] This has led to an outpouring of studies that use machine learning algorithms to analyze brain signal data for emotion and attention recognition goals.[43,44,45] Verifying the reliability of decoder algorithms that classify emotions, attention, valence, arousal, stress and other attributes of human experience at a high temporal resolution has remained a persistent challenge. There is no consensus today about where the information boundary exists in noninvasively measurable brain signals which are known to change at the order of milliseconds and contain a wide variety of meaningful data.[46,47,48]

As the brain's role in attentional processes remains an evolving area of study, at least seven regions are known to contribute: the frontal lobe, posterior parietal lobe, cingulate cortex, thalamus, superior colliculus, reticular activating system, and the claustrum.[49,50]

These are fundamental regions for orchestrating not only focus, but multisensory integration generally,[51,52] evidenced in the broad interhemispheric circuitry.[53,54,55,56] In our fast-paced world full of distractions, sound provides a safe, noninvasive means of modulating brain activity at both short and long timescales, and it will be beneficial for the field to better understand the anatomical and physiological mechanisms the brain uses to produce focus. Nevertheless, that topic, and the related topic of localization of brain function are beyondthe scope of this paper.

It is important to note also that this a multidisciplinary study aiming to contribute to a diverse literature. Accordingly, readers may have difficulty finding common ground and common definitions for psychological constructs or instrumentally derived brain signals that relate to focus. The word 'focus' itself is a conceptual pigeonhole, since the brain state of 'focus' is not necessarily distinct from 'interest,' 'concentration,' 'attention' or any other word meant to describe it. Describing phenomenological states through words or numbers remains a persistent challenge far outside the scope of this work: here we kept to an operational definition of focus based on subjective and self-reported assessments made by participants after each experience, as described in Methods.

C. Combining brain decoding with sound testing to optimize for focus

The motivation for this study was to learn what properties of sound affect human focus the most. We ran a large-scale, naturalistic neuroscience experiment where participant brain signals were measured at home (i.e. their natural habitat) to enable the data analyzed to be as close to real world phenomena as possible. Focus was taken as a toy problem representing part of the larger case, which is learning what properties of sound affect human experience comprehensively. Sound is also taken as a microcosm for study here, to be a single physical construct representing a bigger picture, where the goal is understanding the totality of sensory inputs to humans, including how changes in visual, ambient, olfactory, tactile and other perceptions affect emotions and attention.

Currently, the field has a limited ability to gauge focus responses to sound. The primary method for doing so involves surveying and self-report, despite the well-known issue that surveys inevitably interrupt the listening process and interfere with listener experience. This interference has significant implications not only on the experience of sound acutely, right after the listener is disrupted, but throughout the remainder of the sound as well because auditory inputs have a natural cadence that is rooted in continuous time. Music in particular has an inherent momentum, it ebbs and flows and has a certain *sturm und drang* that is inexorably tied to time and lost whenever a stream of sound stops.

Although the effects of sound on measurable human brain signals are faint in terms of physical amplitudes (microvolts), due to the high temporal resolution of the signals (milliseconds), the information in decoded data is rich and well-suited to the timescales of changes in human experience. Sound is a uniquely plastic medium also, conducive to a variety of programmatic manipulations, which makes it an especially fruitful experimental stimulus. In the audio industry today there are several companies (including Endel, Brain.fm, Mubert, Enophone, Focus@Will, Melodia, AIVA, etc.) who offer automated, AI-generated sound for a variety of commercial purposes. Similarly, we see video, gaming, and learning content created by human and AI teams to be areas that will blossom in upcoming years. Across all use-cases, focus is consistently a key parameter of experience and so the current study was designed to be as generally applicable as possible.

II. METHODS

A. Participants

Sixty-two (62) participants (22 female), 18-65 years, completed four (4) sessions over a single (1) week at their own home. Adult participants were recruited from an opt-in screening panel and came from all five (5) major regions of the continental United States (Northeast, Southwest, West, Southeast, and Midwest). Only participants who reported normal hearing, normal vision, or vision that was corrected to normal with contact lenses were included. We excluded volunteers who reported using medication that might influence the experiment or other neurological or psychiatric conditions that could influence the results. All participants were native English speakers. A written informed consent was obtained from each participant prior to their participation and all participants were compensated equally at a rate of $30 USD per hour.

B. Paradigm

1. Tasks

Participants performed various tasks in a designated app (neuOS™/Arctop) while listening to one of three types of sound and wearing a headband recording their brain activity. Each participant received a kit to their home that included headphones, headband and a tablet with the app pre-installed. Participants recorded four one hour long sessions, while listening to different types of sounds. Sessions included 30 minutes of a "Preferred Task" — selected by the participant, followed by short tasks, such as Tetris (a video game), Arithmetics (math problems) and Creativity (word problems). Randomization was used to assign subjects to groups according to a pseudorandom schedule that controlled for the order of different audio types. (Fig. 1).

Participants were instructed to choose a Preferred Task which they could perform in a seated position while listening to music, and which they would be happy repeating in all 4 sessions. For example, knitting, working, reading, sudoku, or puzzles were all valid Preferred Task options. At the end of each task the participants self-reported on their experience through a survey which used linearly-scaled slider buttons to quantify their experience (e.g. focus level, enjoyment, stress, motivation). For the Preferred Task, the survey included reporting on their focus level during the first and second half of the task, resulting in 6 self-reported focus levels per session (Preferred Task: 2, arithmetics: 1, creativity: 1, tetris: 2).

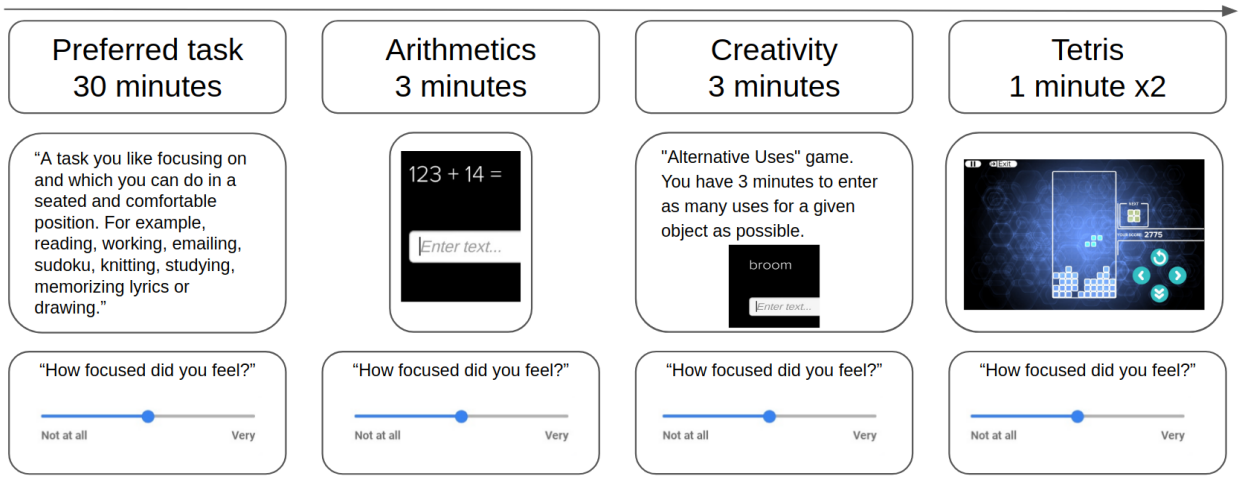


Fig. 1. Schematic illustration of the paradigm in each recording session. Each session started with 30 minutes of a task selected by the participant ("Preferred Task"), followed by 3 minutes of arithmetics exercises, 3 minutes of a creativity task, and two levels of tetris (one minute each). After each task, participants answered a survey where they reported on aspects of their experience (e.g. focus, enjoyment, stress) using linear scale sliders from "Not at all" (0) to "Very" (1).

2. Sounds

Each participant experienced four audio conditions over the four days of the study: two music playlists by leading digital service providers (Spotify and Apple, downloaded 9/2020), one personalized soundscape (Endel), and silence (no audible sounds). We selected Spotify's 'Focus Flow' playlist and Apple Music's 'Pure Focus' playlist to represent the category of pre-recorded sounds designed to increase focus. For personalized soundscapes we selected the mobile application Endel to represent the category of real-time, custom-made audio. The Endel app 'Focus' soundscape was used by each participant on their own device. For the condition of silence, participants still wore headphones, but no music or audible sounds of any kind were played and no soundscape was generated - participants simply completed the session in a quiet environment where they expected to be free from disruptions.

C. Data Processing

1. Data Acquisition

While participants were listening to audio and engaging in a variety of tasks, their electrical brain activity was recorded using InteraXon's Muse-S device, a portable, noninvasive electroencephalograph (EEG) weighing 41 grams. The headband includes four dry fabric EEG sensors (sampling rate: 256 Hz), photoplethysmography (PPG) sensors (for heart rate) and motion sensors (gyroscope, accelerometer). The EEG sensors are located on the scalp, two frontal channels and two temporal channels, with the reference channel at Fpz. The headbands were put on by participants themselves with the assistance of a quality control screen that started each session by giving participants real-time feedback on the signal quality, making it easy to adjust the headband for optimal signal quality (Fig. 2).

2. *Preprocessing*

A band-pass filter (0.5-70Hz) was applied on each channel together with a notch filter (60Hz) to remove line noise. From within the performed tasks, 5 seconds of brain data segments were extracted from the filtered signal, using a sliding window with a stride of 200ms (5Hz).

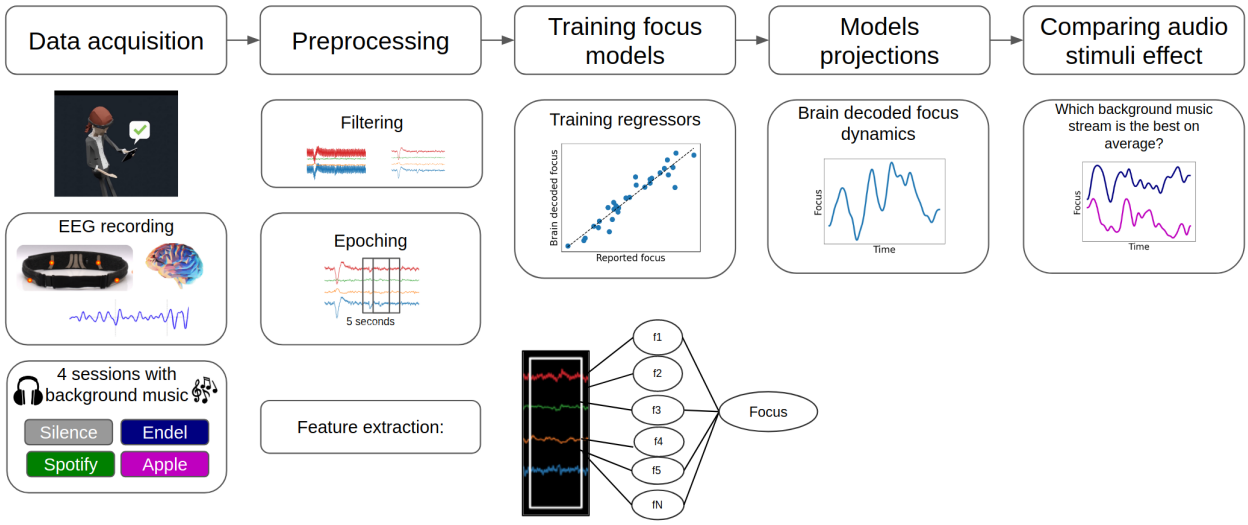


Fig. 2. Schematic illustration of the processing pipeline. Data acquisition included at home EEG recordings of 4 sessions, each with a different background audio stream. EEG processing included filtering the signal, feature extraction and training machine learning models to map between brain features and reported focus. Obtaining the brain decoded focus dynamics enables comparison of focus levels during different types of audio streams.

3. *Feature extraction*

From each EEG segment (epoch), relevant features were calculated: power spectrum features - each segment was transformed to the frequency domain using Welch method, and for each channel, the average power in different frequency bands was calculated. Power spectrum interactions - the power spectrum ratio between bands and engagement index. Time domain features - for each channel, the first four moments, entropy and number of zero-crossing points. Pairwise correlations between channels in the different frequency bands were calculated as well. For each epoch, a total of 124 features were extracted and to avoid extreme values, a programmatic trimming procedure was performed for high and low values.

4. *Brain-based focus models*

Eleven (11) participants were excluded from further analysis due to excessive noise in their recorded brain data and/or unreliable survey responses, leaving a total of 51 participants (mean age= 36, SD=8, 17 females) in the experimental analysis. Average EEG features were calculated for all valid participants (N=51) in each subtask (e.g. creativity, tetris), resulting in 1224 focus ranked events (51 participants x 4 sessions x 6 ranked events per session). Then, in a cross validation procedure, multiple random forests regression models were trained on random subsets of participants (80%) to predict the self-reported focus based on the EEG features.

For each participant, from the subset of regression models not used during training, the best fit model was selected. To control for overfitting a shuffle analysis was performed. For each participant, the selected model was chosen after random permutations of the self-reported focus. Fig. 5A shows the correlations distribution obtained for the selected models compared to shuffle. The selected regression model was then applied on all EEG segments to get a continuous brain-decoded gradient of focus dynamics. A Gaussian filter was used to smooth the dynamics. Fig. 3 shows the resulting brain decoded focus levels of a single participant across all four sessions and all tasks. Supp. Fig. 2 shows the dynamics for two additional example participants for further comparison.

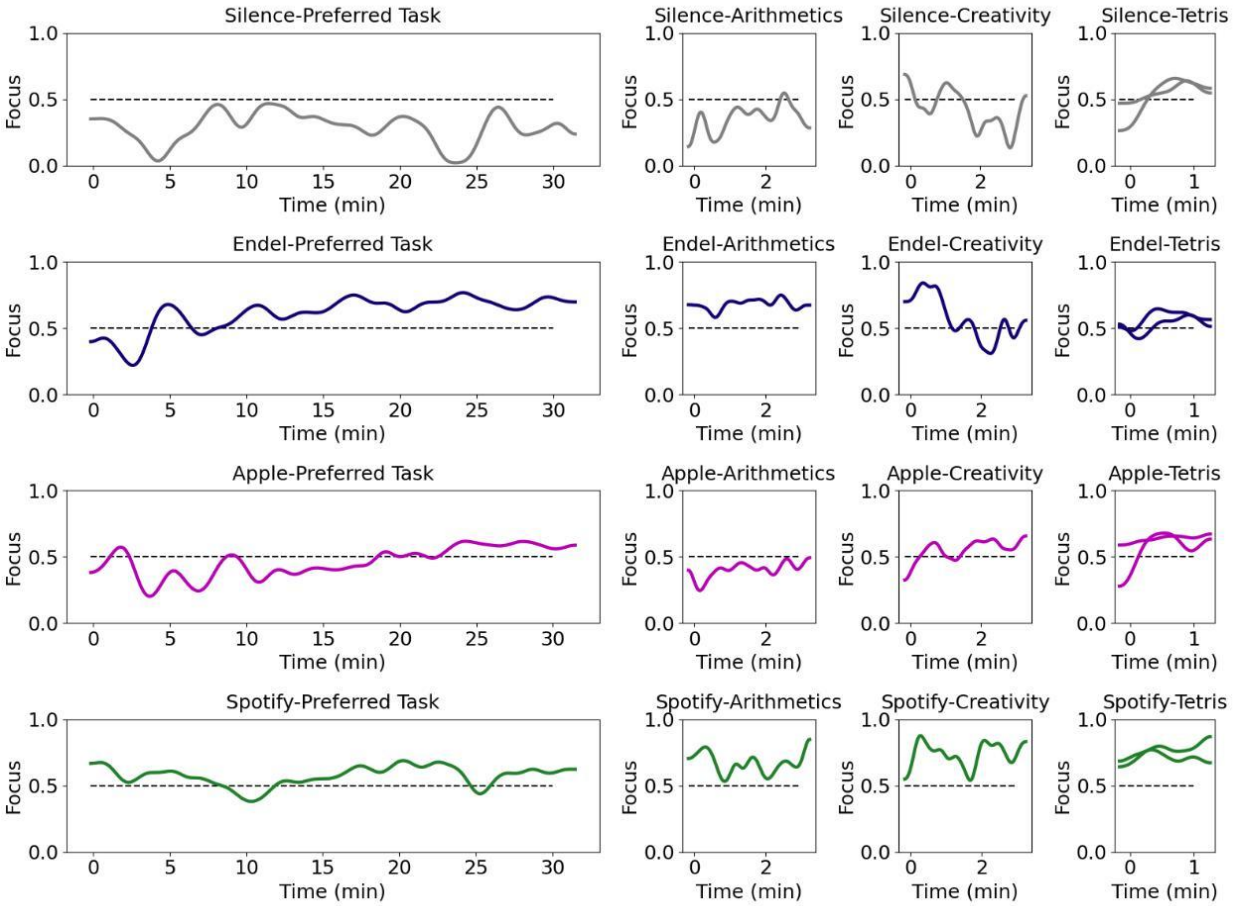


Fig. 3. Brain based Focus model dynamics of a single participant during four recorded sessions. Each row represents a session with a different audio stream. Each session included various tasks: 30 minutes of a "Preferred Task," followed by 3 minutes of an arithmetic task, 3 minutes of a creativity task and 2 tetris levels (1 minute each).

5. Statistical methods

For comparisons between average focus levels during the different audio streams, we calculated for each participant (N=51) the median focus level within each task. For each task, we conducted a one-way repeated measures ANOVA (Analysis of variance) test. Then, if $p<0.05$, paired t-tests were applied post hoc to compare between pairs of audio streams using Holm-Bonferroni correction.

Time series statistical tests were applied to compare focus level dynamics and discover specific time periods of significant difference. A paired t-test was applied at each second between focus levels of two audio streams. The p values were then corrected for multiple comparisons with setting a threshold for a minimum significant sequential time samples. The threshold was determined by random permutations (1000 iterations) of participants' conditions and repeating the statistical test, resulting in a distribution of significant sequential time samples. The threshold was set as the 95% percentile of the resultant distribution.

6. Audio signal decomposition and feature extraction

The raw audio files of Apple and Spotify playlists were used to obtain audio features dynamics in the time and frequency domain. The features were calculated using Python's library pyAudioAnalysis[57] (e.g. energy, spectral entropy, chroma coefficients). The features were calculated in short-time windows of 50 ms with a sliding window of 25 ms. Then, basic statistics of the audio features were calculated in windows of 30 seconds (e.g. mean and std), resulting in 136 features. To enable mapping to the brain model, the brain decoded focus levels were also averaged in the corresponding 30 seconds windows (Fig. 4).

To obtain the threshold for significant correlations between audio features and focus levels ($p<0.05$), a shuffle analysis was performed. Random permutations (1000 iterations) of the brain decoded focus levels were applied across songs (to preserve the time dependency of focus levels within a song). The correlations of each audio feature was calculated with the permuted focus level. The threshold was set as the 95% percentile of the resulting correlation's distribution.

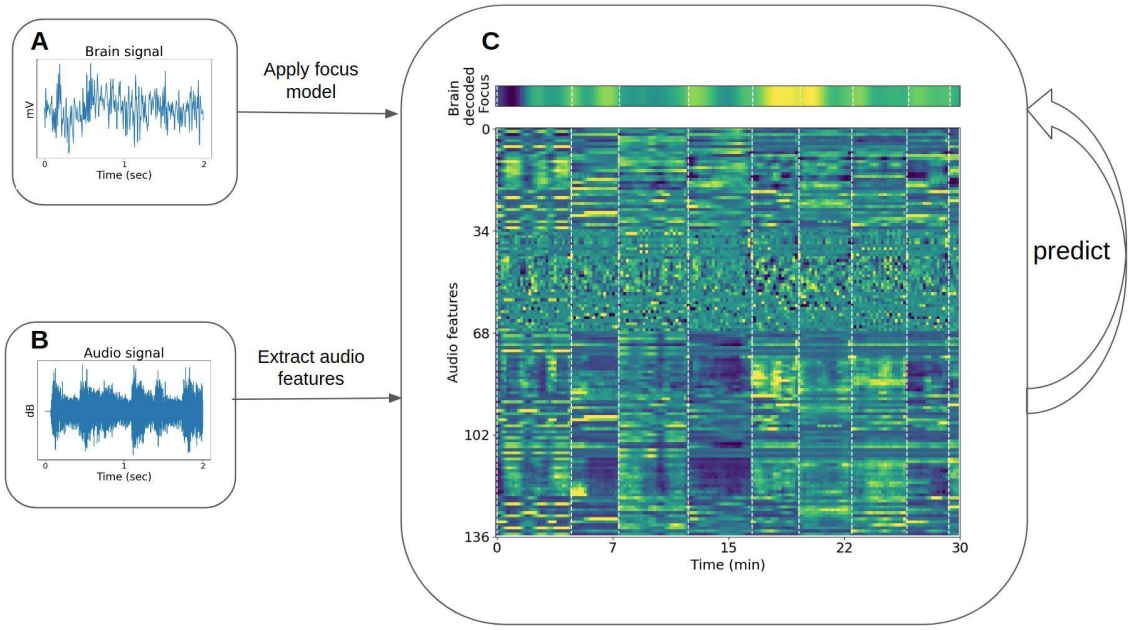


Fig. 4. Diagram demonstrating the framework for reverse correlation of time-series focus values with audio features. A. Example of a recorded EEG segment, which after applying the preprocessing and trained models on 30 minutes of recordings, transforms to the brain decoded focus dynamics (top C). B. Examples of an audio segment taken from one of the songs. Bottom C. The audio features dynamics during 30 minutes of recordings.

7. *Obtaining the audio decoded focus model*

To map the relation between the calculated audio features and the averaged brain decoded focus, we applied principal component analysis (PCA) to reduce audio features dimensionality. We then trained regression models between the transformed audio features and the brain decoded focus (via cross validation with 70% of the songs in each iteration) for the significant audio features only. The presented audio decoded focus model is a linear model based on the first PCA component of the features (shifted and rescaled), calculated using.

III. RESULTS

A. Brain-based focus models predict self-reported focus

Fig. 5 shows the Pearson correlations between the brain decoded focus model predictions (median across task) and the self reported focus. Aggregating all tasks from all participants, our focus models' performance is Corr(416)=0.6, $p<10^{-4}$ (Fig. 5B). The average correlation per participant is <Corr(24)>=0.543, $p<10^{-4}$, while the average for the shuffled control is <Corr(24)>=0.26, p=0.34 (Fig. 5A). Averaging the results across the tasks and the audio conditions, yielded a correlation of Corr(16)=0.8, $p<5*10^{-4}$ (Fig. 5C).

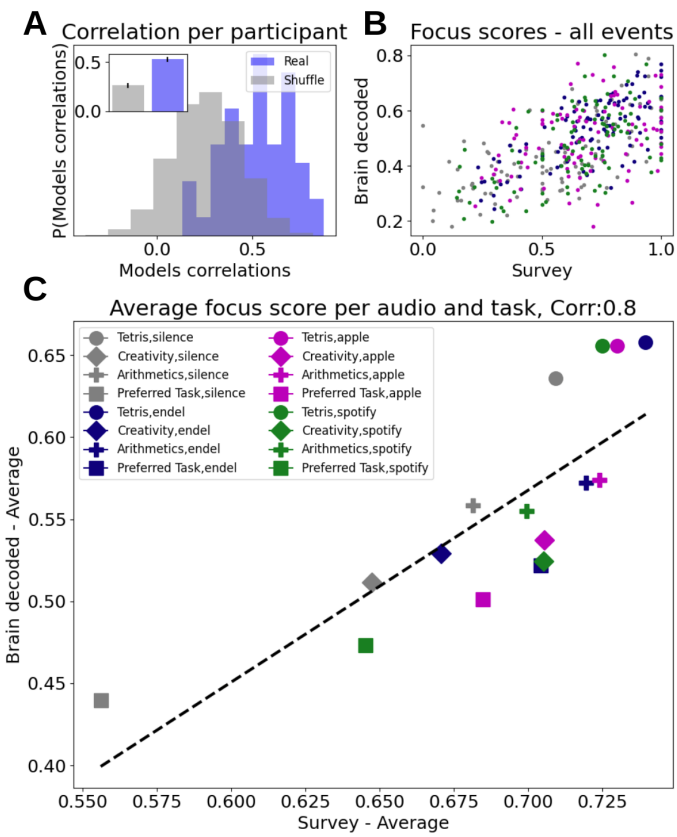


Fig. 5. Focus models performance. A. Histograms of focus models Pearson correlations per participant (N=51, blue), relative to shuffled control (gray). Inset shows average values (Real=0.54, Shuffled=0.26). B. Average focus levels per event vs. self-reported focus (survey) - Resulted in Pearson correlation of 0.6. C. Average focus levels across task type and audio type vs. average reported focus (average survey), resulted in Pearson correlation of 0.8.

B. Music had an effect on focus levels only during the Preferred Task.

Using our validated focus models, we then compared between the average focus levels elicited by the audio streams in each task. The background audio stream had an effect only on the Preferred Task (F(3,150)=4.144, p=0.008, statistical methods for details). We have not found differences in focus levels for the arithmetic, tetris or creativity task (Table 1).

Task	F(3,150)	p
Tetris	0.349	0.78
Arithmetics	0.148	0.93
Creativity	0.155	0.93
Preferred task	**4.144**	**0.008**

Table 1. Results of a one-way repeated measures ANOVA, applied separately on each task. Significant difference between focus levels while listening to different audio streams is found only in the Preferred Task (p=0.008).

C. Soundscapes induce a higher focus level compared to silence.

To find focus differences between the audio streams, we ran post hoc tests with Holm-Bonferroni correction and found that on average, streaming soundscapes (with Endel app) was significantly higher compared to silence (Fig. 6A, table 2, p=0.01), while streaming music using Apple or Spotify did not have an effect (p=0.12, p=0.74 for Apple and Spotify respectively). In addition, for 35.3% of the participants the Endel session was the one with the highest focus level, while for 27.5% Apple was the highest, for 19.6% Spotify and for 17.6% silence (Fig. 6B, the details sorted focus levels per participant are shown in Supp. Fig. 1).

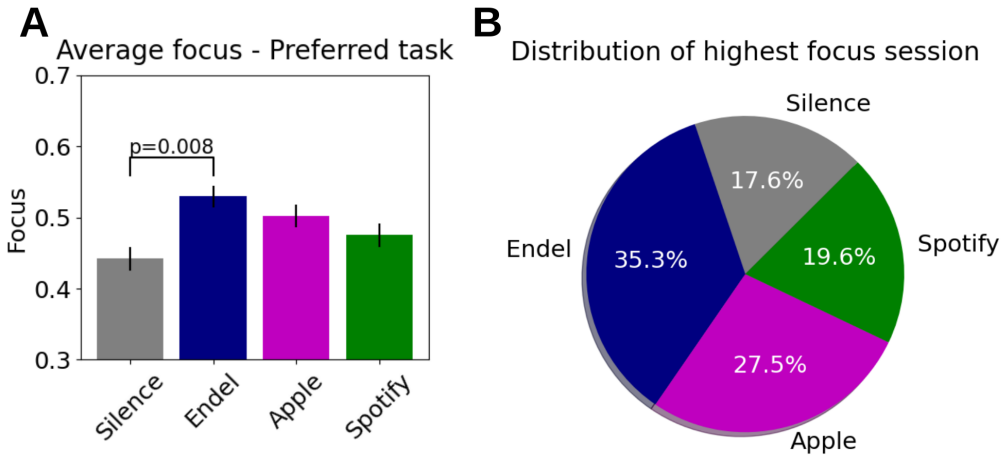


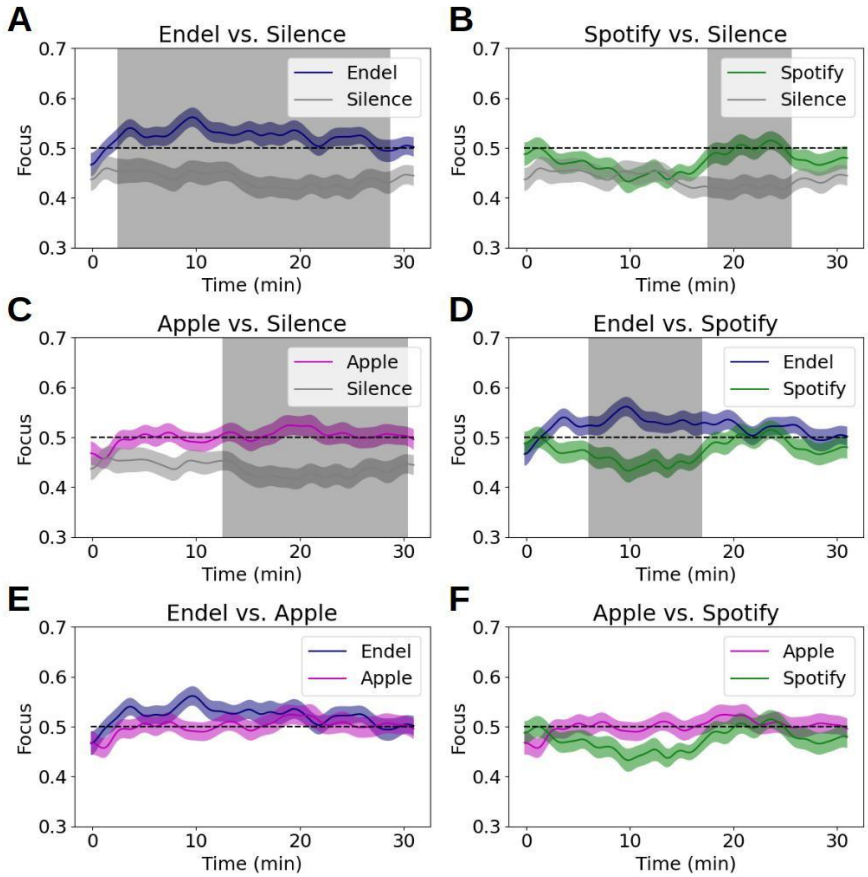
Fig. 6. Comparison of the brain decoded focus during the Preferred Task while listening to different audio streams. Average focus levels for each audio stream during the Preferred Task, including statistical results showing focus levels during listening to personalized soundscapes is higher than silence (p=0.008). B. Distribution of the best session (highest focus on average) for each participant.

Pair	Avg	p (holm)
Endel-silence	**0.088**	**0.01**
Apple-silence	0.060	0.12
Spotify-silence	0.033	0.74
Endel-apple	0.028	0.74
Endel-spotify	0.054	0.15
Spotify-apple	-0.02	0.74

Table 2. Results of post-hoc statistical tests, comparing all pairs of audio streams' focus levels. P values are corrected using Holm-Bonferroni method. Endel was found to most significantly affect focus relative to the baseline of silence.

D. Time series analysis of the focus dynamics reveal differences between all audio streams and silence.

Exploiting the temporal resolution for the focus dynamics, we then compared the focus dynamics of the audio streams during the 30 minutes of the Preferred Task (Fig. 7, table 3). When comparing Endel's soundscapes vs. Silence (Fig. 7A), we found that the focus level elicited by Endel's soundscape was higher 87% of the time, starting after 2.5 minutes. In addition, although, on average there wasn't a significant difference, the focus level elicited by Apple's playlist was higher than Silence 60% of the time, starting at 12.5 minutes (Fig. 7C), and the focus level elicited by Spotify's playlist was higher than Silence 27% of the time, starting at 17 minutes (Fig. 7B). In addition, we found that focus elicited by Endel's soundscape was higher than Spotify's playlist in 37% of the time, starting at 6 minutes (Fig. 7D).



Pair	Significant difference (% of session)	Significant segments (minutes)
Endel-silence	87%	2.5-28
Apple-silence	60%	12.5-30
Spotify-silence	27%	17.5-25.5
Endel-apple	0%	
Endel-spotify	37%	6-17
Spotify-apple	0%	

Fig. 7. (Left) Comparing brain decoded focus dynamics during the 30 minutes of the Preferred Task. Each subfigure shows a comparison between two audio streams, while the gray areas are the timings with a significant difference ($p<0.05$ corrected, see statistical methods for details). **Table 3** (Right). **Summary of focus time dynamics comparison**, showing for each pair the percentage of time and time segments with significant difference (100% = 30 minutes).

E. Focus level differences between Endel and Silence is task dependent

During the Preferred Task, 51% of the participants (26) chose to work, while the rest (49%) read a book (29.4%), played games (9.8%) or did other various tasks (e.g. knitting, 9.8%). To assess the audio effect on focus levels during these different tasks, we split the participants to the ones who worked and those that did other tasks. We found that for the "working" group, the focus level elicited by Endel's soundscapes was higher compared to silence ($p=0.017$), while for the "not-working" group there was no difference (Fig. 8, table 4).

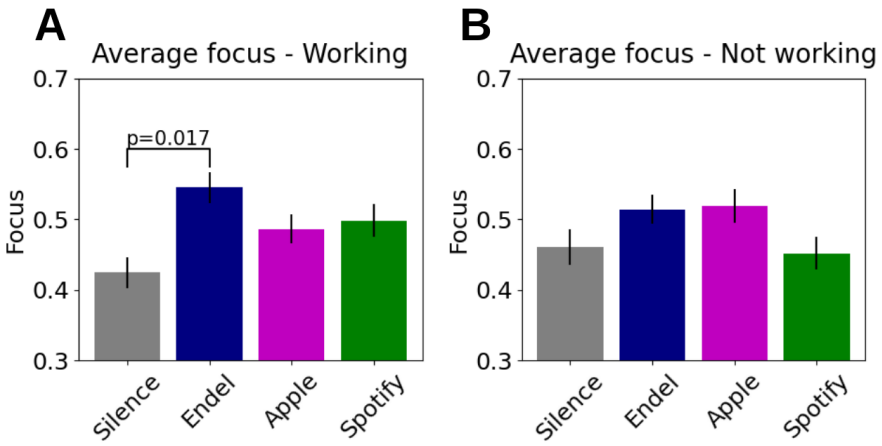


Fig. 8. Average brain decoded focus scores during the Preferred Task split by participants who worked during the 30 minutes, and the ones which did not work (read, played, etc). Only for the working subset of participants, we found that listening to soundscapes (Endel) was significantly higher than silence.

F. Focus level differences between audio and silence is age dependent

We next split the participants into two age groups according to the median (36). We found that for the younger participants (age<36), all audio streams were better than silence (Fig. 9, Table 4, $p<0.01$) while for the older participants (age>36), there was no difference between audio and silence.

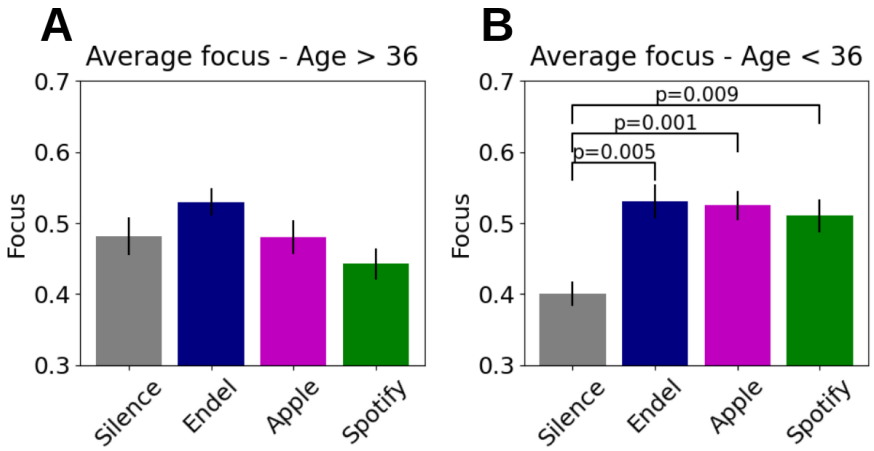


Fig. 9. Average brain decoded focus scores during the Preferred Task, split by participants' median age (th=36). Only for the younger subset of participants, we found that listening to any music was significantly higher than silence.

Group	F	p
Working	3.86 (3,75)	0.01
Not-working	1.72 (3,72)	0.17
Age<36	6.27 (3,72)	<0.001
Age>36	1.83 (3,75)	0.15

Table 4. Results of a one-way repeated measures ANOVA performed on each subgroup, comparing the average brain decoded focus levels of each audio stream during the Preferred Task.

G. Brain decoded focus can be predicted by the audio signal decomposition

Having evidence that background music has an effect on focus levels, we go further and ask whether music and soundscapes can be composed according to a formula to increase focus levels. Meaning, are we able to understand audio signal characteristics that drive focus well enough to predict focus levels based on audio signals alone.

Leveraging the high temporal resolution of the noninvasive brain measurements, we next generated a prediction model which will predict the brain-based focus level based on the audio features extracted from the audio signal. Raw audio files streamed in Apple and Spotify sessions were used to extract different audio features with a running sliding window of 30 seconds. Each feature was checked for its contributory power to the measured average focus level. Fig. 10 shows the resulting correlations between each audio feature and the brain based focus level. 20 features of the 136 evaluated were found to have significant correlations ($p<0.05$, see Statistical methods).

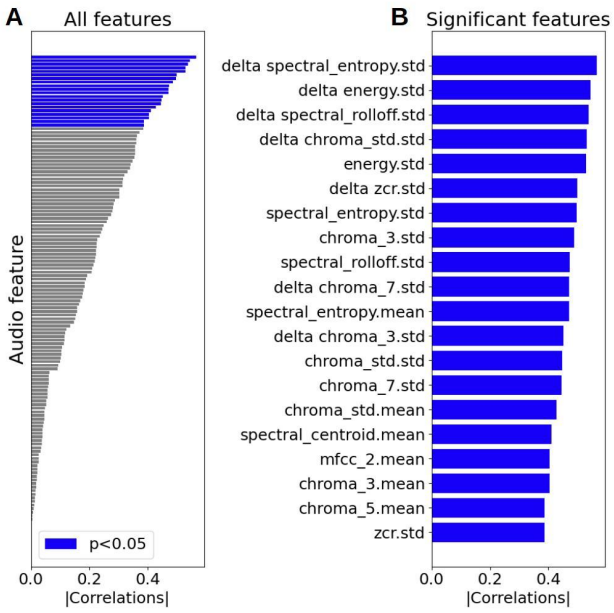


Fig. 10. Single features correlations (absolute value) between audio features extracted from music playlists and listener brain decoded focus. A. Correlations for all extracted features (see Methods). Significant features are colored in blue. B. Only significant features ($|Corr|>0.39$) are named at right, see Statistical methods for details).

We next combined multiple audio features to generate an audio based model which predicts focus level (see methods). Fig. 11 shows the dynamics of the audio decoded focus, together with the brain decoded focus (Corr=0.7). Fig. 11D shows that if we threshold our dynamics to output a binary prediction (low/high focus), the audio model reaches 88% accuracy in predicting the brain based focus (AUC=0.93).

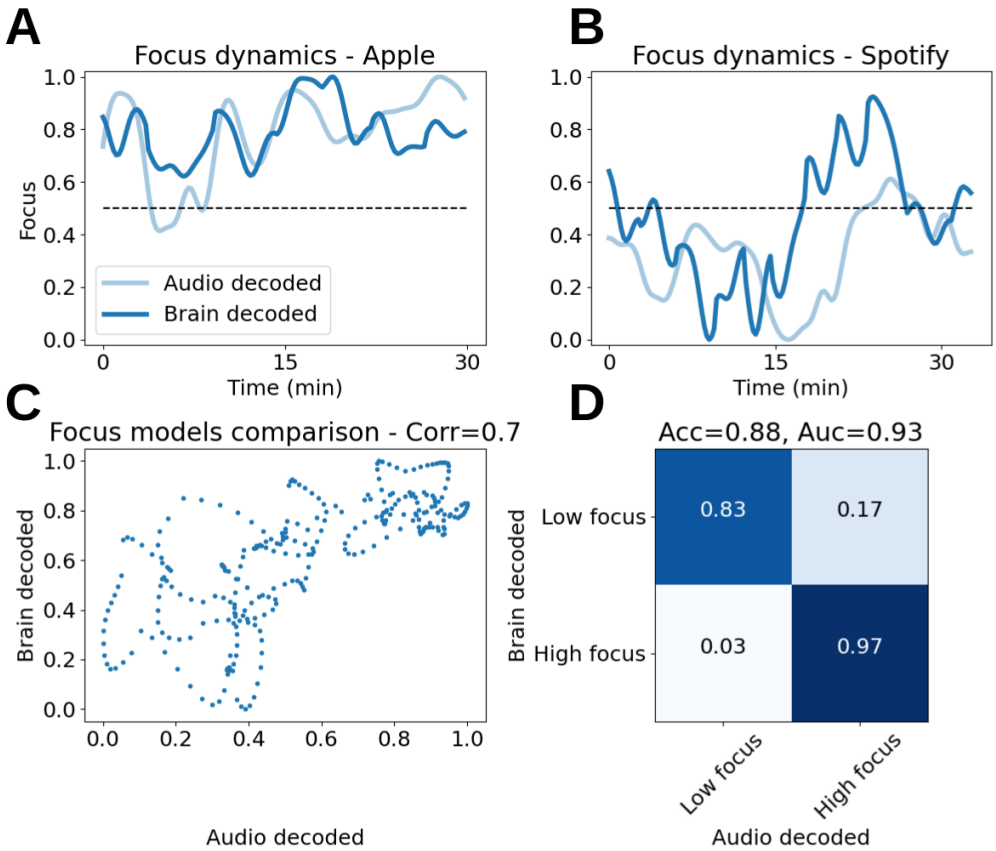


Fig. 11. Results of predicting brain decoded focus from audio features. A+B. Dynamics of brain decoded focus (dark blue) and audio decoded focus (light blue), during 30 minutes of the Preferred Task for Apple (A) and Spotify (B). C. Brain decoded focus (y-axis) vs. Audio decoded focus (x-axis) for both playlists (Apple + Spotify). D. Confusion matrix after thresholding the focus predictions to classify between low and high focus. Classification accuracy obtained: 88% (Area under ROC curve: 0.93).

Beyond composing soundscapes for focus, we can also use these prediction models to rate the focus level of a song and assemble successful playlists based on existing songs. To demonstrate this, we compared the song average of the audio decoded output to the brain decoded output. As can be seen in Fig. 12B, there is a correlation of 0.74 between the focus models at the song level. Fig. 12A shows these averages sorted by the brain-based model.

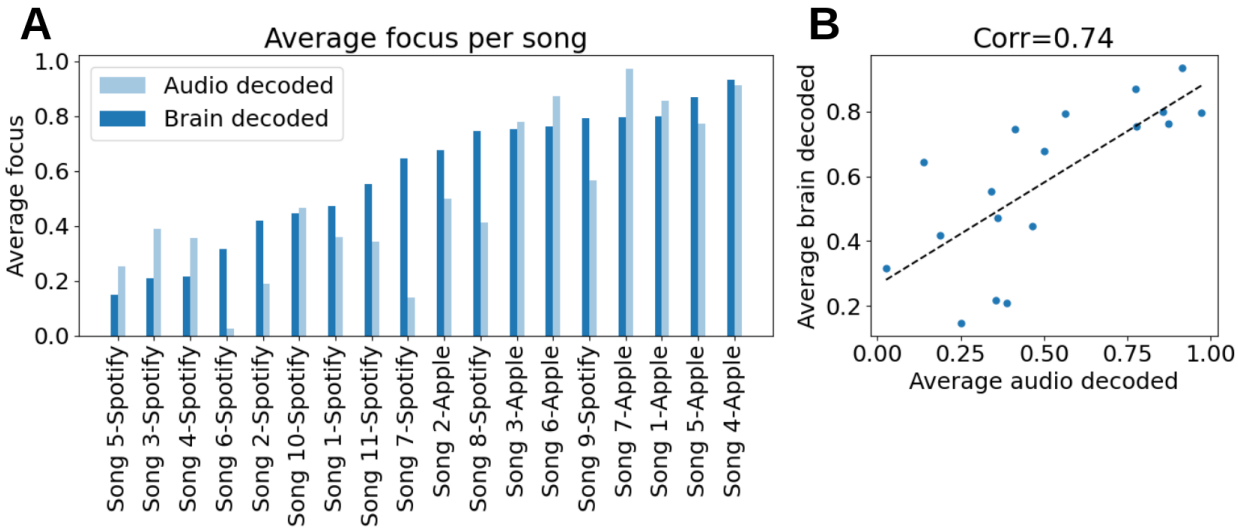


Fig. 12. Averaging focus scores for each song. A. Sorted focus scores per song obtained by the brain model (brain decoded - blue), next to the focus obtained by the audio model (audio decoded- light blue). B. Focus scores per song - brain decoded (y axis) vs. audio decoded (x axis). Pearson correlation between them: Corr(18)=0.74, p=0.0004.

IV. DISCUSSION

"The soundscape of the world is changing. Modern man is beginning to inhabit a world with an acoustical environment radically different from any he has hitherto known." R. Murray Schafer.

As the sounds available to us continue to multiply and we have an increasing number of options to modulate our auditory lives by, a handful of take-aways from this study standout:

A. Brain-based measurement of focus is possible "in-the-wild."

Although the effects of sound and music on the human brain can be subtle in measured brain signals, in terms of the changes produced in raw electromagnetic currents, they are robust and highly quantifiable with effectively-trained algorithms as shown here. Classifying emotional and attentional responses is particularly useful when done at the sub-second temporal resolution to track dynamics continuously over time at the same timescale as brain functions impact perception and behavior. In this study we demonstrated that noninvasive brain decoding technology is able to deliver this needed resolution.

A key benefit of the current approach is that this method of high temporal resolution brain measurement can be performed reliably outside of traditional laboratories. In this current study not a single laboratory or facility was used for data acquisition. Instead, 18-65 year olds across the U.S. received a technology kit in the mail and experienced music playlists and personalized soundscapes from the comfort of their own homes while they recorded their own brain signals from the chair or desk of their choice, and at the time of their choosing. In other words, in their natural habitat, at their own pace.

B. Focus is increased most by personalized soundscapes.

Within the at-home environment of this study, personalized soundscapes were found to be the best at increasing participant's focus levels. After 2.5 minutes, on average, listeners of personalized soundscapes experienced a meaningful increase in focus level, while for music playlists it took approximately 15 minutes. The audio effect on focus levels was found to be task dependent as well, suggesting that willful orientation of attention towards work tasks may have created a brain context especially suited to modification by sound. While engaged in work, participants may also have been more prone to distraction and thus more impacted by the positive uplift of audio relative to other contexts such as gaming, where the visual content was very immersive and required fast-paced decisions.

C. Sound preferences and focus effects vary between people.

Similar to Mehr et al. who found a variety of human experiences in response to sound and concluded that, "music does appear to be tied to specific perceptual, cognitive, and affective faculties, including language (all societies put words to their songs), motor control (people in all societies dance), auditory analysis (all musical systems have signatures of tonality), and aesthetics (their melodies and rhythms are balanced between monotony and chaos)..."[58] we found an astonishing diversity in focus dynamics. Due to the variety, a next step will include closed-loop selections of sounds, where iterative sound testing is used per person to identify the significant parameters for maximizing focus. One limitation of the current study is its inability to reach conclusions regarding gender-dependent effects, which was at least partially due to this study's imbalanced data set. Despite efforts to recruit a balanced group of participants, enrollment was done on a rolling basis and in the end the female subgroup was underpowered statistically. In future research, especially for closed loop, real-time testing, balanced participant groups will be important for reaching more detailed conclusions.

D. Personalized sound is uniquely functional.

Personalized soundscapes specifically, and personalized audio in general, should be investigated further for their capacity to increase productivity, creativity and well-being as these attributes of human experience are associated with one's ability to focus. It is possible that the seamlessness of the personalized soundscapes tested here, which played continuously without gaps in the sound like the music playlists had between songs, was a critical part of the observed effect on focus. At every juncture of the experience there is more to be learned, but at a high level, a main lesson of this study is that there is a strong need for personalization of sound in order to most effectively achieve functional goals like increasing focus.

Since the current study did not allow for a comparison of personalized soundscapes to personalized music playlists, where audiences either made their own playlist for focus or were allowed to skip songs whenever they wanted, follow up research will incorporate this variable. Equivalently, tests of pre-recorded soundscapes will be helpful in disentangling the effects of personalization on the listener, and will likely contribute to a fuller understanding of how sound properties correlate with emotion and attention changes.

V. CONCLUSION

Here we studied the effects of sound on human focus levels using noninvasive brain decoding technology. To gain a better understanding of the optimal acoustical environment for increasing focus levels in listeners, we combined a custom app, portable brain measuring headbands, and machine learning algorithms to successfully obtain high temporal resolution focus dynamics from participants at home. Using the brain decoded focus dynamics, we then analyzed how various properties of sound affected focus levels in different tasks.

We found that while performing a self-paced task for a long period of time (such as working), personalized soundscapes increased focus the most relative to silence. Professionally curated playlists of pre-recorded songs also increased focus during specific time intervals, especially for the youngest audience demographic. Large variance in response profiles across participants, together with task and age dependent effects, suggest that personalizing sounds in real-time may be the best strategy for producing a focus in the listener.

The approach taken here can be adapted to include other emotions (e.g. enjoyment, calm, happiness, etc.), attentional parameters ('The Zone,' memory, etc.) and be used to assess additional content (e.g. visual, ambient, olfactory, etc.), including interactive gaming and e-learning where personalization and high temporal resolution experience measures may be particularly beneficial.

The ancient Ionian Greek philosopher Pythagoras, who first identified the mathematical connection between a string's length and it's pitch, believed the whole cosmos was a form of musical composition.[59] We too see the rich mathematical models obtained in this study, by mapping sound properties to human experience, as a glimpse into the natural laws governing how we feel and think. The better these laws can be understood, the more empowered individuals will be to modulate their sound environments to suit their goals and states of mind.

There remains much to figure out. While we as a species continue to cause a "shift in the sensorium," we simultaneously experience that shift all over daily life and it is not clear where we as a society are headed. This study showed that sounds have a distinct effect on focus, and paves the way for designing sounds to help us focus better in the future.

ACKNOWLEDGEMENTS

This work was supported by Warner Music, Sony, Endel, and Universal Music who provided sounds, data, financial support and through communication with the authors at Arctop, a for-profit technology company, helped advance theoretical and practical aspects of this research.

REFERENCES

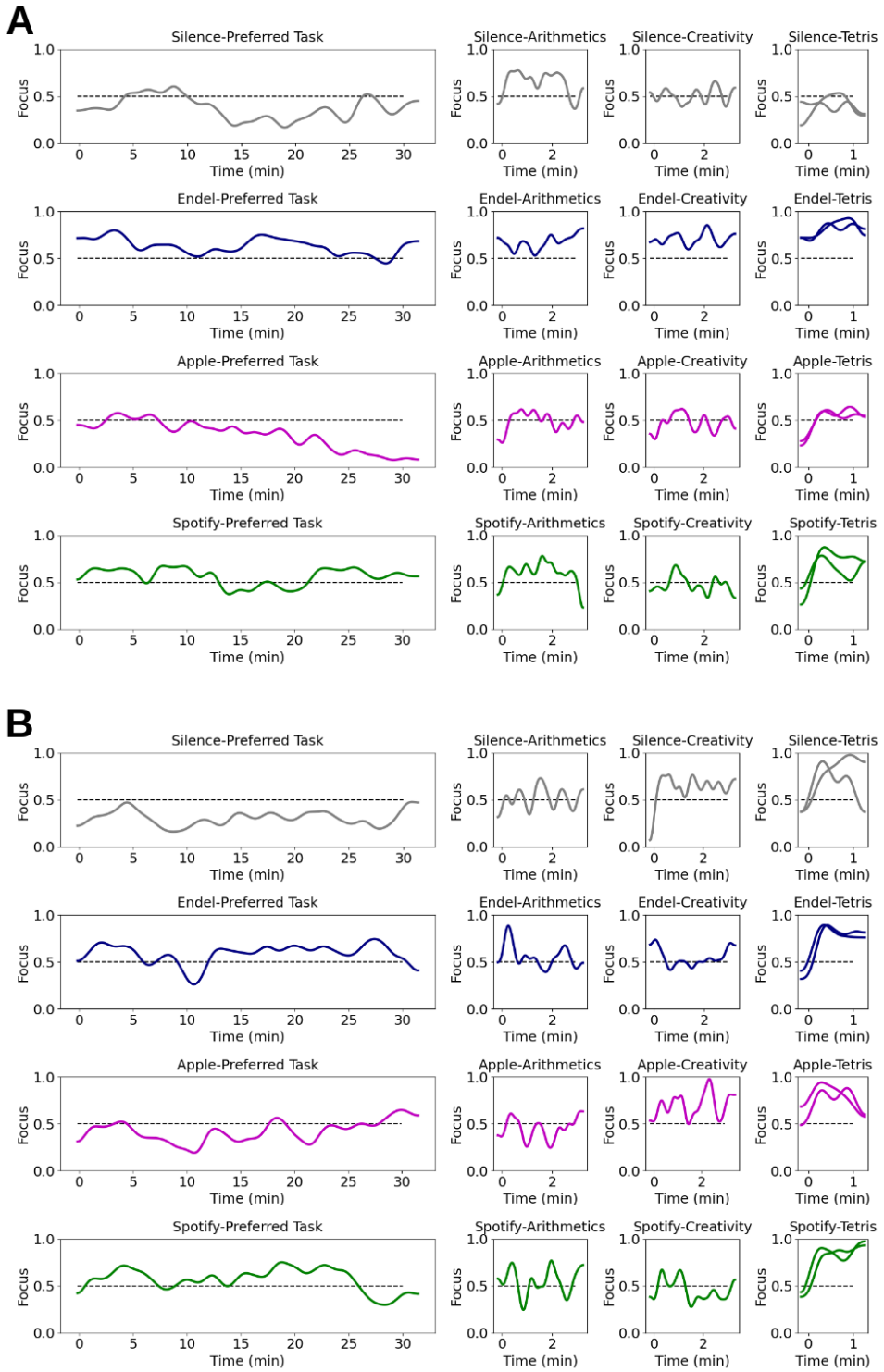
1. Whitman, Walt. Song of Myself. Roycrofters, 1904.
2. Bryson, Bill. The Body: A Guide for Occupants. Random House, 2019.
3. Harari, Yuval N. Sapiens: A Brief History of Humankind. New York: Harper, 2015.
4. Shelemay, Kay Kaufman, Vasili Chazir, Odetta, Myra Burgess, Mattie Connolly, Deidre Connolly, Mary Morrison et al. Soundscapes: Exploring music in a changing world. New York: Norton, 2001.
5. McLuhan, Marshall, and Bruce R. Powers. The global village: Transformations in world life and media in the 21st century. Communication and Society, 1989.
6. Sacks, Oliver W. Tales of music and the brain. London, UK: Picador, 2007.
7. Levitin, Daniel J. This is your brain on music: The science of a human obsession. Penguin, 2006.
8. Ofek, E., and H. Pratt. "Neurophysiological correlates of subjective significance." Clinical neurophysiology 116, no. 10 (2005): 2354-2362.
9. Schreiber, Charles A., and Daniel Kahneman. "Determinants of the remembered utility of aversive sounds." Journal of Experimental Psychology: General 129, no. 1 (2000): 27.
10. Kumar, Sukhbinder, Katharina von Kriegstein, Karl Friston, and Timothy D. Griffiths. "Features versus feelings: dissociable representations of the acoustic features and valence of aversive sounds." Journal of Neuroscience 32, no. 41 (2012): 14184-14192.
11. Zald, David H., and José V. Pardo. "The neural correlates of aversive auditory stimulation." Neuroimage 16, no. 3 (2002): 746-753.
12. Cheung, Vincent KM, Peter MC Harrison, Lars Meyer, Marcus T. Pearce, John-Dylan Haynes, and Stefan Koelsch. "Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity." Current Biology 29, no. 23 (2019): 4084-4092.
13. Huron, David. Sweet anticipation: Music and the psychology of expectation. MIT press, 2008.
14. Picard, Rosalind W. Affective computing. MIT press, 2000.
15. Cunningham, Stuart, Harrison Ridley, Jonathan Weinel, and Richard Picking. "Audio emotion recognition using machine learning to support sound design." In Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound, pp. 116-123. 2019.
16. Hu, Xiao, and Yi-Hsuan Yang. "Cross-dataset and cross-cultural music mood prediction: A case on western and chinese pop songs." IEEE Transactions on Affective Computing 8, no. 2 (2017): 228-240.
17. Roda, Antonio, Sergio Canazza, and Giovanni De Poli. "Clustering affective qualities of classical music: Beyond the valence-arousal plane." IEEE Transactions on Affective Computing 5, no. 4 (2014): 364-376.
18. Saari, Pasi, György Fazekas, Tuomas Eerola, Mathieu Barthet, Olivier Lartillot, and Mark Sandler. "Genre-adaptive semantic computing and audio-based modelling for music mood annotation." IEEE Transactions on Affective Computing 7, no. 2 (2015): 122-135.
19. Yang, Yi-Hsuan, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen. "A regression approach to music emotion recognition." IEEE Transactions on audio, speech, and language processing 16, no. 2 (2008): 448-457.
20. Levitin, Daniel J., Parag Chordia, and Vinod Menon. "Musical rhythm spectra from Bach to Joplin obey a 1/f power law." Proceedings of the National Academy of Sciences 109, no. 10 (2012): 3716-3720.
21. Nia, Hadi T., Ankita D. Jain, Yuming Liu, Mohammad-Reza Alam, Roman Barnas, and Nicholas C. Makris. "The evolution of air resonance power efficiency in the violin and its ancestors." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 471, no. 2175 (2015): 20140905.
22. Washburne, Christopher. "“More Cowbell”: Latin Jazz in the Twenty-First Century." In Latin Jazz, pp. 141-174. Oxford University Press.
23. Bradley, Margaret M., and Peter J. Lang. "The International Affective Digitized Sounds (; IADS-2): Affective ratings of sounds and instruction manual." University of Florida, Gainesville, FL, Tech. Rep. B-3 (2007).
24. Yang, Yi-Hsuan, Yu-Ching Lin, Ya-Fan Su, and Homer H. Chen. "A regression approach to music emotion recognition." IEEE Transactions on audio, speech, and language processing 16, no. 2 (2008): 448-457.
25. Russell, James A. "A circumplex model of affect." Journal of personality and social psychology 39, no. 6 (1980): 1161.

26. Hizlisoy, Serhat, Serdar Yildirim, and Zekeriya Tufekci. "Music emotion recognition using convolutional long short term memory deep neural networks." Engineering Science and Technology, an International Journal (2020).
27. Brotzer, J. M., E. R. Mosqueda, and K. Gorro. "Predicting emotion in music through audio pattern analysis." In IOP Conference Series: Materials Science and Engineering, vol. 482, no. 1, p. 012021. IOP Publishing, 2019.
28. Vempala, Naresh N., and Frank A. Russo. "Predicting emotion from music audio features using neural networks." In Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR), pp. 336-343. London, UK: Lecture Notes in Computer Science, 2012.
29. Cunningham, Stuart, Harrison Ridley, Jonathan Weinel, and Richard Picking. "Supervised machine learning for audio emotion recognition." Personal and Ubiquitous Computing (2020): 1-14.
30. Larsen, Randy J., and Edward Diener. "Promises and problems with the circumplex model of emotion." (1992).
31. Cowen, A.S. and Keltner, D., 2017. Self-report captures 27 distinct categories of emotion bridged by continuous gradients. Proceedings of the National Academy of Sciences, 114(38), pp.E7900-E7909.
32. George, Mark S., Terence A. Ketter, Priti I. Parekh, Barry Horwitz, Peter Herscovitch, and Robert M. Post. "Brain activity during transient sadness and happiness in healthy women." American Journal of Psychiatry 152, no. 3 (1995): 341-351.
33. Elliott, Rebecca, Judy S. Rubinsztein, Barbara J. Sahakian, and Raymond J. Dolan. "The neural basis of mood-congruent processing biases in depression." Archives of general psychiatry 59, no. 7 (2002): 597-604.
34. Gündel, Harald, Mary-Frances O'Connor, Lindsey Littrell, Carolyn Fort, and Richard D. Lane. "Functional neuroanatomy of grief: an FMRI study." American Journal of Psychiatry 160, no. 11 (2003): 1946-1953.
35. Maddock, Richard J., Amy S. Garrett, and Michael H. Buonocore. "Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task." Human brain mapping 18, no. 1 (2003): 30-41.
36. Heisenberg, Werner. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik." In Original Scientific Papers Wissenschaftliche Originalarbeiten, pp. 478-504. Springer, Berlin, Heidelberg, 1985.
37. Fox, Andrew S., Regina C. Lapate, Alexander J. Shackman, and Richard J. Davidson, eds. The nature of emotion: Fundamental questions. Oxford University Press, 2018.
38. Bremner, J. D., R. Soufer, G. McCarthy, R. Delaney, L. H. Staib, J. S. Duncan, and D. S. Charney. "Gender differences in cognitive and neural correlates of remembrance of emotional words." Psychopharmacology bulletin 35, no. 3 (2001): 55-78.
39. Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. Heuristics and biases: The psychology of intuitive judgment. Cambridge university press, 2002.
40. Faller, Josef, Jennifer Cummings, Sameer Saproo, and Paul Sajda. "Regulation of arousal via online neurofeedback improves human performance in a demanding sensory-motor task." Proceedings of the National Academy of Sciences 116, no. 13 (2019): 6482-6490.
41. Asif, Anum, Muhammad Majid, and Syed Muhammad Anwar. "Human stress classification using EEG signals in response to music tracks." Computers in biology and medicine 107 (2019): 182-196.
42. Lu, Yun, Mingjiang Wang, Wanqing Wu, Yufei Han, Qiquan Zhang, and Shixiong Chen. "Dynamic entropy-based pattern learning to identify emotions from EEG signals across individuals." Measurement 150 (2020): 107003.
43. Suhaimi, Nazmi Sofian, James Mountstephens, and Jason Teo. "EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities." Computational intelligence and neuroscience 2020 (2020).
44. Wagh, Kalyani P., and K. Vasanth. "Electroencephalograph (EEG) based emotion recognition system: A review." Innovations in Electronics and Communication Engineering (2019): 37-59.
45. Luo, Yuling, Guopei Wu, Senhui Qiu, Su Yang, Wei Li, and Yifei Bi. "EEG-based Emotion Classification Using Deep Neural Network and Sparse Autoencoder." Frontiers in Systems Neuroscience 14 (2020): 43.
46. da Silva, Fernando Lopes. "EEG and MEG: relevance to neuroscience." Neuron 80, no. 5 (2013): 1112-1128.
47. Erickson, Molly A., Emily S. Kappenman, and Steven J. Luck. "High temporal resolution measurement of cognitive and affective processes in psychopathology: What electroencephalography and magnetoencephalography can tell us about mental illness." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 3, no. 1 (2018): 4-6.
48. Zamm, Anna, Stefan Debener, Ivana Konvalinka, Natalie Sebanz, and Günther Knoblich. "The sound of silence: an EEG study of how musicians time pauses in individual and joint music performance." Social cognitive and affective neuroscience 16, no. 1-2 (2021): 31-42.

49. Crick, Francis C., and Christof Koch. "What is the function of the claustrum?." Philosophical Transactions of the Royal Society B: Biological Sciences 360, no. 1458 (2005): 1271-1279.
50. Schmahmann, Jeremy D., and David Caplan. "Cognition, emotion and the cerebellum." Brain 129, no. 2 (2006): 290-292.
51. Remedios, Ryan, Nikos K. Logothetis, and Christoph Kayser. "Unimodal responses prevail within the multisensory claustrum." Journal of Neuroscience 30, no. 39 (2010): 12902-12907.
52. Puryear, Corey B., and Sheri JY Mizumori. "Reward prediction error signals by reticular formation neurons." Learning & Memory 15, no. 12 (2008): 895-898.
53. Smith, Jared B., and Kevin D. Alloway. "Functional specificity of claustrum connections in the rat: interhemispheric communication between specific parts of motor cortex." Journal of Neuroscience 30, no. 50 (2010): 16832-16844.
54. Koubeissi, Mohamad Z., Fabrice Bartolomei, Abdelrahman Beltagy, and Fabienne Picard. "Electrical stimulation of a small brain area reversibly disrupts consciousness." Epilepsy & Behavior 37 (2014): 32-35.
55. Koch, Christof. "The quest for consciousness." Engineering and Science 67, no. 2 (2004): 28-34.
56. Searle, John R., Daniel Clement Dennett, and David John Chalmers. The mystery of consciousness. New York Review of Books, 1997.
57. Giannakopoulos, Theodoros. "pyaudioanalysis: An open-source python library for audio signal analysis." PloS one 10, no. 12 (2015): e0144610.
58. Mehr, Samuel A., Manvir Singh, Dean Knox, Daniel M. Ketter, Daniel Pickens-Jones, Stephanie Atwood, Christopher Lucas et al. "Universality and diversity in human song." Science 366, no. 6468 (2019).
59. James, Jamie. The music of the spheres: Music, science, and the natural order of the universe. Springer Science & Business Media, 1995.

SUPPLEMENTARY FIGURES

Supp. Fig. 1. Average Brain decoded focus levels for all participants in all sessions' Preferred Task. For each participant, the sessions are sorted from their highest average focus level (4) to the lowest (1), with colors representing the different experimental audio streams. The distribution of highest sessions is presented in Fig. 6B.



Supp. Fig. 2. Brain based Focus model dynamics of two example participants during four recorded sessions. Each figure (A/B) is a single participant. Each row represents a session with a different audio stream and each column represents a task. The dotted line at 0.5 (y-axis) represents each participant's individual neutral or baseline focus level, with '1' being the highest focus level attainable (very concentrated) and '0' the lowest (completely unfocused).