

## Three Psychometric Scales for Groove Research

Senn, Olivier; Bechtold, Toni Amadeus; Jerjen, Rafael; Kilchenmann, Lorenz; Hoesl, Florian

DOI:

[10.1177/20592043231185663](https://doi.org/10.1177/20592043231185663)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Senn, O, Bechtold, TA, Jerjen, R, Kilchenmann, L & Hoesl, F 2023, 'Three Psychometric Scales for Groove Research: Inner Representation of Temporal Regularity, Time-Related Interest, and Energetic Arousal', *Music & Science*, vol. 6, pp. 1-16. <https://doi.org/10.1177/20592043231185663>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Three Psychometric Scales for Groove Research: Inner Representation of Temporal Regularity, Time-Related Interest, and Energetic Arousal

Olivier Senn<sup>1</sup> , Toni Amadeus Bechtold<sup>1,2</sup> , Rafael Jerjen<sup>1</sup>,  
Lorenz Kilchenmann<sup>1</sup> and Florian Hoesl<sup>1</sup> 

## Abstract

This study develops three composite psychometric scales for the use in listening experiments across music psychology. The new scales measure the following three psychological constructs: (1) *Inner representation of temporal regularity*: This scale allows listeners to assess to what extent they experience a subjective feeling of temporal regularity and predictability while listening to music. The scale consists of four items, and it measures the underlying construct with good reliability (Cronbach's  $\alpha = .88$ ). (2) *Time-related interest*: This scale measures how much listeners feel that the rhythm of the music captures their attention. The scale consists of three items and shows good reliability ( $\alpha = .85$ ). (3) *Energetic arousal*: This scale measures how much the listeners feel energized while they are listening to music. The scale uses four items and has excellent reliability ( $\alpha = .95$ ). The development of the scales was motivated by the psychological model of musical groove in which the three underlying psychological constructs play an important role; the new scales serve to test the model hypotheses. The three scales can also be administered separately and might prove to be useful in research contexts within music psychology that investigate subjective experiences of temporal regularity, interest, or energetic arousal in music listeners.

## Keywords

Energetic arousal, groove, mood, music psychology, psychometric scale, rhythm, temporal regularity, time-related interest

Submission date: 14 October 2022; Acceptance date: 12 June 2023

## Introduction

Groove research, as a subdiscipline of music psychology, studies the relationship between music and body motion. Specifically, it investigates how music triggers the groove experience, understood as listeners' pleasurable urge to move in response to the music (Janata et al., 2012; Madison, 2006). In the last two decades, groove research has considered many factors that affect the groove experience. These factors may be connected to the music, to the listener, or to the listening situation (Senn et al., 2021).

Senn et al. (2019) proposed a psychological model of musical groove that intends to explain the urge to move in response to music in the context of musical, personal, cognitive, situational, and behavioral factors. The model

incorporates many findings of earlier studies and formulates hypotheses about the relationships between the music and listeners' responses. Figure 1 shows the core relationships of the model in a slightly updated version compared to Senn et al. (2019). The model and its hypotheses can be summarized as follows: the music and its properties (left box)

<sup>1</sup> School of Music, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland

<sup>2</sup> Department of Music, University of Birmingham, Birmingham, UK

## Corresponding author:

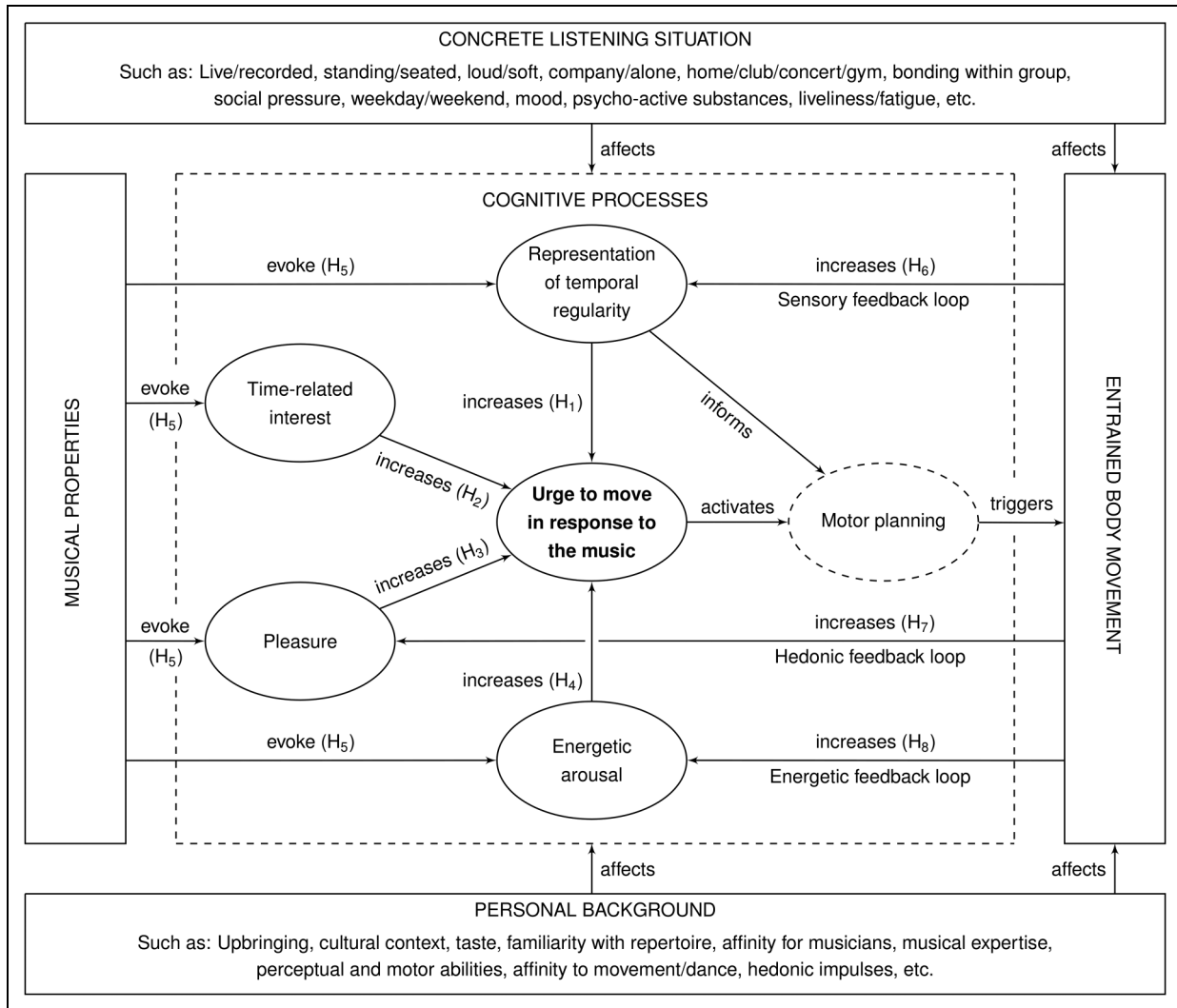
Olivier Senn, School of Music, Lucerne University of Applied Sciences and Arts, Arsenalstrasse 28a, 6010 Luzern-Kriens, Switzerland.

Email: [olivier.senn@hslu.ch](mailto:olivier.senn@hslu.ch)



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page

(<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



**Figure 1.** Groove model (version 2, Senn et al., 2023). For the first version, see Senn et al. (2019).

evoke a series of cognitive processes in the listener (central box) that are hypothesized to induce an *urge to move*, which in turn motivates body movement (right box). These cognitive processes are influenced by the concrete listening situation (top) and the personal background of the listener (bottom).

The model claims that listeners' *urge to move* is affected by a variety of cognitive processes that are linked to music perception and that can be understood as mediators for the *urge to move* (note that the psychological constructs relating to these cognitive processes will be introduced more comprehensively below):

1. The first hypothesis states that listeners are more likely to experience an urge to move if the music also triggers an *inner representation of temporal regularity* (H<sub>1</sub>). According to Merker (2014), the experience of temporal regularity (e.g., an isochronous pulse) is a necessary precondition for the synchronization of body movement with the music.
2. The model further hypothesizes that the urge to move increases with the *time-related interest* (understood as attention to the rhythm) that the music triggers in the listener (H<sub>2</sub>). This hypothesis derives from the discussion on syncopation and groove, specifically from the idea that listeners' body movements "fill in" beats of the overarching meter that are silent due to syncopation (Witek, 2017). Music with many syncopations requires that listeners pay attention to the rhythm in order to fill in the missing beats using their own bodies (see also Spiech et al., 2022).
3. The third hypothesis states that greater listening *pleasure* increases the urge to move (H<sub>3</sub>). We base this hypothesis on the results of an earlier study which suggest that pleasure is a precondition for the urge to move (Senn et al., 2020, p. 58). The hypothesized causal relationship between pleasure and the urge to move contrasts with Janata et al.'s non-hierarchical definition, which describes the

experience of groove paratactically as a “pleasant sense of wanting to move” (2012, p. 56).

4. The model hypothesizes that the urge to move increases when the music raises *energetic arousal* in the listener, i.e., a feeling of excitement and activation ( $H_4$ ). The energy felt by the listener may be released in the form of body movement. This agrees with pupillometry studies that indicate a connection between arousal and groove: Bowling et al. (2019) have shown that high-groove music triggered pupil dilation in the listener, which is interpreted as a physiological marker of arousal (for the connection between movement energy and groove music, see also Dotov et al., 2021).
5. The groove model assumes that the urge to move is a “late” response in the sequence of cognitive processes leading up to body movement. This means that the urge to move is not caused directly by the music, but mediated through other cognitive processes such as the experience of *regularity*, *interest*, *pleasure*, and *energetic arousal* ( $H_5$ ). This provides a criterion to assess the completeness of the model: as long as musical stimuli affect the urge to move without mediation through any other cognitive process, important causal pathways must be missing from the model, and the model is incomplete.

The model further hypothesizes that actual entrained *body movement* (right box) acts back into the cognitive domain:

6. The model states that synchronized body movement increases the *inner representation of temporal regularity* (sensory feedback loop,  $H_6$ ), because the regularities are experienced in a variety of sensory modalities, besides hearing. Dancers may experience the regularities more strongly if they don’t just hear the temporal regularities, but also feel the periodic change of pressure under their feet, or the relative movement of the environment in front of their eyes.
7. Body movement is enjoyable to many people; therefore listening *pleasure* might be enhanced through synchronized body movement (hedonic feedback loop,  $H_7$ ).
8. Finally, people experience a feeling of power during body motion, at least as long as fatigue is low (Terry et al., 2020). This effect might in turn increase the energetic arousal in the listener/dancer (energetic feedback loop,  $H_8$ ).

The psychological groove model – centering on the *urge to move*, but also taking musical, cognitive, personal, and situational aspects into account – potentially provides a unifying framework for the study of the groove experience. However, until recently, the model has only been a

theory, and the hypotheses have not yet been verified. One major obstacle to test these hypotheses is the lack of psychometric scales that are capable of measuring the intensity of the cognitive processes involved. The *urge to move* and *pleasure* scales with three items each have been published in English (Senn et al., 2020) and German (Düvel et al., 2021) language versions. The current study aimed at developing three additional composite psychometric scales that measure to what extent listeners experience *temporal regularity*, *time-related interest*, and *energetic arousal* while listening to music. This development followed the methodology of McCoach et al. (2013) and unfolded in three phases: firstly, candidate questionnaire items for the scales were collected, and their content validity was evaluated by experts in the field of music psychology. Secondly, the factor structure of the items was investigated using data from a first listening experiment and exploratory factor analysis. The best items were then selected for the final questionnaire. Thirdly, the reliability of the selected items was studied based on data from a second listening experiment and confirmatory factor analysis. The goal of this procedure was to develop scales that are both valid and reliable.

## Definition of the Constructs

We understand the *inner representation of temporal regularity* as a cognitive process in the music listener that, during the listening process, consciously or subconsciously analyses the temporal organization of the music up to the current moment and predicts future musical events on the basis of this analysis. The *inner representation of temporal regularity* is strong if listeners easily recognize patterns within the time organization and effortlessly predict the future. Learning a rhythmic pattern as it unfolds has been successfully described as an instance of predictive coding (Senn, 2023; Vuust et al., 2009, 2018; Vuust & Witek, 2014). To our knowledge, no psychometric scale is currently available that measures the strength of listeners’ subjective experience of *temporal regularity* while listening to the music.

*Time-related interest* is a state of mind in which listeners focus their attention on the time organization of the music, such as its rhythm. We expect complex rhythms (e.g., syncopated rhythms, see Witek et al., 2014) to capture listeners’ interest and attention more than simple rhythms. No psychometric scale that measures the intensity of *time-related interest* appears to be available today.

We understand *energetic arousal* in response to music as a listener’s subjective experience of their own excitement and alertness, an inner feeling of being lively and energetic, which is triggered or at least enhanced by the music. The experience of energetic arousal may be associated with physiological responses in listeners, such as pupil dilation (Bowling et al., 2019; Spiech et al., 2022), increased heart and respiration rates, and others (Salimpoor et al., 2009), but it is not identical with these physiological processes.

We also differentiate energetic arousal (as listeners' experience of their own state of feeling energized, relevant to the current construct) from listeners' perception that the music is energetic, or that the musicians play the music with high energy (for the difference between perceived and induced emotion, see Gabrielsson, 2001; Song et al., 2016). We derived the *energetic arousal* scale from the *vigor* sub-scale of the Profile of Mood States (POMS) questionnaire (McNair et al., 1971, 1992; Heuchert & McNair, 2012) that addresses a very similar experience.

## Development of Candidate Items

### Inner Representation of Temporal Regularity and Time-Related Interest

The *inner representation of temporal regularity* and *time-related interest* scales needed to be developed from scratch. We proceeded as follows: the first and second authors composed a list of 34 sentences. These sentences formulate statements that aim to address one of the two constructs according to the definitions given above. In a survey within the team, each of the five authors judged the

statements whether they address the temporal regularity or/and the time-related interest construct and whether they should be retained or discarded. With respect to 28 items, at least three authors voted for retaining the item; these items were selected for the pretest (see items in Table 1, below).

### Energetic Arousal

The *energetic arousal* scale was based on the Profile of Mood States (POMS), which is a self-report questionnaire that measures a person's current mood in six different dimensions: tension, anger, depression, vigor, fatigue, confusion (McNair et al., 1971, 1992; Heuchert & McNair, 2012, Terry et al., 1999). The POMS consists of 65 items, each presenting a single adjective. Respondents are asked to rate how well the adjective describes their current mood using a five-point Likert scale ("Not at all", "A little", "Moderately", "Quite a lot", "Extremely"). Several short forms of the POMS questionnaire exist (Shacham, 1983; Grove & Prapavessis, 1992; Curran et al., 1995; Mendoza et al., 1999; Baker et al. 2002; Terry et al., 2003). In their abbreviation and adaptation of POMS scales, Hewston et al. (2008) target mood effects

**Table 1.** Candidate questionnaire items with mean (standard deviation) content validity ratings for the perceived temporal regularity and rhythmic interest constructs. Items No. 16–28 were subsequently excluded due to low content validity ratings or substantial criticism by the participants of the pretest. The content validity ratings of the dominant concept are in bold print.

No.	Item	Content Validity	
		Regularity	Interest
1	I recognize a regular beat in this music.	<b>0.926</b> (0.141)	0.476 (0.400)
2	The rhythm of this music is predictable.	<b>0.926</b> (0.169)	0.583 (0.347)
3	I recognize a clear pulse in this music.	<b>0.889</b> (0.185)	0.524 (0.389)
4	The rhythm of this music is repetitive.	<b>0.778</b> (0.226)	0.571 (0.338)
5	The rhythm of this music is steady.	<b>0.753</b> (0.286)	0.393 (0.315)
6	The rhythm of this music sounds disorganized.	<b>0.654</b> (0.339)	0.548 (0.330)
7	The rhythm of this music is chaotic.	<b>0.630</b> (0.350)	0.405 (0.306)
8	This music has an even rhythm.	<b>0.630</b> (0.297)	0.381 (0.311)
9	The rhythm of this music catches my attention.	0.370 (0.350)	<b>0.869</b> (0.228)
10	This music has a captivating rhythm.	0.272 (0.262)	<b>0.833</b> (0.248)
11	This music has a fascinating rhythm.	0.235 (0.304)	<b>0.833</b> (0.264)
12	The rhythm of this music is interesting.	0.235 (0.290)	<b>0.821</b> (0.279)
13	The rhythm of this music is exciting.	0.210 (0.247)	<b>0.786</b> (0.260)
14	The rhythm of this music is boring.	0.235 (0.290)	<b>0.762</b> (0.325)
15	This music has a surprising rhythm.	0.420 (0.301)	<b>0.679</b> (0.294)
16	This music has a regular beat.	<b>0.914</b> (0.149)	0.488 (0.390)
17	This music has a uniform rhythm.	<b>0.691</b> (0.260)	0.333 (0.272)
18	The basic underlying rhythm of this music stays unchanged, regardless of what happens on top of it.	<b>0.691</b> (0.319)	0.440 (0.363)
19	The rhythm of this music gives me an idea of how it will continue.	<b>0.654</b> (0.352)	0.476 (0.293)
20	This music consists of fixed rhythmic patterns.	<b>0.667</b> (0.277)	0.357 (0.255)
21	The rhythm of this music feels static.	<b>0.519</b> (0.311)	0.417 (0.309)
22	The rhythm of this music is fascinating.	0.198 (0.249)	<b>0.750</b> (0.347)
23	I get surprised by the rhythm of this music.	0.519 (0.325)	<b>0.655</b> (0.307)
24	I look forward to hearing what comes next in this music.	0.222 (0.277)	<b>0.655</b> (0.294)
25	I quickly lose interest in this music.	0.235 (0.274)	<b>0.607</b> (0.385)
26	I would like to hear how this music continues.	0.235 (0.274)	<b>0.595</b> (0.356)
27	I want to hear more of this music.	0.160 (0.267)	<b>0.560</b> (0.375)
28	Listening to this music is boring.	0.222 (0.261)	<b>0.440</b> (0.385)

of listening to music in the context of sport (Music Mood-Regulation Scale, MMRS).

The construct measured by the POMS vigor scale “is typified by feelings of excitement, alertness and physical energy” (Terry et al., 1999, p. 863), which agrees very closely with our definition of *energetic arousal* above. The eight items of this scale consist of one mood-related adjective (or expression) each: “lively”, “active”, “energetic”, “cheerful”, “alert”, “full of pep”, “carefree”, “vigorous”. The vigor construct is quasi-identical with the energetic arousal dimension from Thayer’s multidimensional model of mood and arousal: Thayer described energetic arousal as “subjective sensations of energy, vigor, or peppiness” (Thayer, 1989, p. 6), which anticipated three of the concepts that operationalize the POMS vigor scale (“energetic”, “vigorous”, “full of pep”).

The *energetic arousal* scale developed in this study adapts the POMS vigor scale so that it can easily be used in listening experiments as a response to music. The first change consisted in identifying the music as the cause of energetic arousal. Thus, instead of just asking whether listeners feel energized in general, the new items ask whether the music in particular makes them feel energized. The second change consisted in formulating questionnaire items that are full statements and contain the POMS vigor scale’s adjectives. The reformulation as full statements allows to use the same 7-point agreement scale for respondent feedback that is already in use in the Experience of Groove Questionnaire for

the *urge to move* and *pleasure* scales (Senn et al., 2020), thus simplifying the administration of the questionnaire. The adapted candidate items for the *energetic arousal* scale are presented in Table 2 (lower block, items 32–39). The content validity of the items was not tested in the pretest, since content validity is already covered by the use of the core adjectives in the POMS vigor scale.

## Pretest: Content Validity of 28 Questionnaire Items

In the course of the pretest, groove research experts rated the 28 candidate questionnaire items on their content validity and suitability to measure the constructs *inner representation of temporal regularity* and *time-related interest*. The goal of the pretest was to assess which of the items best capture either of the constructs (content validity), how the formulation of the items could be improved, and which items were ambiguous, or otherwise inadequate, and should therefore be discarded (suitability). The candidate items for the *energetic arousal* scale, derived from the POMS, were not subjected to the pretest.

## Method

**Participants.** A total of 28 individuals (7 female) participated in the pretest. They were selected on the basis of

**Table 2.** Pattern matrix of the 26 questionnaire items (three correlated factors, promax rotation). Factor coefficients in bold print indicate items that were subsequently chosen to be used in *Experiment 2*.

No.	Item	Temporal Regularity	Time-rel. Interest	Energetic Arousal
29	The rhythm of this music gives a strong sense of regularity.	<b>0.905</b>	−0.005	0.062
5	The rhythm of this music is steady.	<b>0.885</b>	−0.022	0.011
8	This music has an even rhythm.	0.870	0.058	0.024
2	The rhythm of this music is predictable.	<b>0.867</b>	−0.209	0.142
1	I can hear a regular beat in this music.	0.855	0.093	0.061
4	The rhythm of this music is repetitive.	0.845	−0.230	0.100
6	The rhythm of this music sounds disorganized.	<b>−0.811</b>	−0.274	0.189
3	I can hear a clear pulse in this music.	0.811	0.102	0.097
7	The rhythm of this music is chaotic.	−0.774	−0.181	0.262
12	I find the rhythm of this music to be interesting.	−0.067	<b>0.993</b>	−0.120
11	I find the rhythm of this music to be fascinating.	−0.098	<b>0.945</b>	−0.078
30	I find the rhythm to be an interesting aspect of this music.	0.023	0.893	−0.019
14	The rhythm of this music bores me.	0.163	<b>−0.778</b>	0.049
10	I find the rhythm of this music to be captivating.	0.123	0.742	0.119
31	The rhythm of this music catches my ear.	0.211	0.667	0.143
9	The rhythm of this music catches my attention.	0.084	0.593	0.239
13	The rhythm of this music is exciting.	−0.107	0.558	0.400
15	The rhythm of this music is full of surprises.	−0.717	0.444	0.047
33	This music makes me feel active.	0.042	−0.056	0.958
34	This music makes me feel energetic.	0.035	−0.022	<b>0.931</b>
37	This music makes me feel full of pep.	0.048	−0.083	<b>0.884</b>
32	This music makes me feel lively.	−0.002	0.069	<b>0.874</b>
39	This music makes me feel vigorous.	0.039	−0.009	<b>0.826</b>
36	This music makes me feel alert.	−0.196	−0.042	0.655
35	This music makes me feel cheerful.	0.099	0.228	0.638
38	This music makes me feel carefree.	0.165	0.393	0.386

their expertise in rhythm research, groove research, and the psychology of music. All participants were active in academia (thirteen professors, six post-doctoral researchers, nine PhD students). They were affiliated with academic institutions in Germany ( $n = 7$ ), Norway ( $n = 5$ ), the United States ( $n = 5$ ), Canada ( $n = 4$ ), the United Kingdom ( $n = 3$ ), Sweden ( $n = 2$ ), Japan ( $n = 1$ ), and the Netherlands ( $n = 1$ ).

**Procedure.** Participants were invited individually by email. The survey was implemented online on the SosciSurvey platform (www.socisurvey.de). Participants logged into the survey using an access code and gave informed consent. They were presented 28 candidate items twice in two blocks. In one block, they rated to what extent these 28 items were relevant to measure a music listener's subjective impression of *temporal regularity*. In another block, they rated the relevance of the same items to measure the impression of *time-related interest*. The ratings were Likert-type answer scales with four different ordinal levels ("not relevant" = 0, "slightly relevant" = 1/3, "relevant" = 2/3, "very relevant" = 1). Both the sequence of the two blocks and the sequence of items within a block were randomized. Participants could comment on each item in a free text field. At the end of each block, participants had the option to propose new items to measure the *inner representation of temporal regularity* or *time-related interest* constructs.

## Results and Discussion

Table 1 shows the mean content validity ratings on each of the two constructs for each of the 28 candidate items; standard deviations are given in parentheses. Participants judged items 1–8 to be relevant to measure the subjective impression of *temporal regularity*. They further judged items 9–15 to be relevant to measure the subjective impression of *time-related interest*.

Participants had the opportunity to comment on each of the items. In several cases, the comments motivated us to reformulate the items. In other cases, the comments convinced us that items were problematic and needed to be discarded altogether.

- *Items 1 and 3:* Participants criticized the use of the word "recognize", because of its meaning ("to know from memory"). We replaced this by "hear" for future uses of these items.
- *Item 2:* A participant proposed to change the wording of item 2 to: "The rhythm of this music seems predictable to me." This amendment improved the item by accentuating the subjective aspect and was retained.
- *Item 7:* Some participants commented that "chaotic" was a vague term. We left the item as it is, since the word is used in everyday parlance and addresses disorganization or irregularity.
- *Items 10–12:* One participant proposed to change the wording of these items to: "I find the rhythm of this

music to be...". This amendment emphasizes the subjectivity of the statement and was retained.

- *Item 14:* One participant proposed an alternative wording for item 14 ("The rhythm of this music bores me."), which we retained.
- *Item 15:* Participants commented that the concept of surprise is relevant to the interest construct, but the wording of the item might be difficult to understand for laypeople. We changed this to "The rhythm of this music is full of surprises", which we assumed is easier to understand.
- *Items 16–28:* The participants judged these items to be either confusing to the musical layperson, awkwardly worded, too long, or they flagged them as inferior duplicates of other items. We dropped them from the list of candidate items and did not use them in *Experiment 1*.

Participants proposed three promising new items:

- *Item 29:* "The rhythm of this music gives a strong sense of regularity."
- *Item 30:* "I find the rhythm to be an interesting aspect of this music."
- *Item 31:* "The rhythm of this music catches my ear."

These items were retained for the next phase of development.

One participant pointed out that an isochronous beat or pulse is not the only way in which music can establish a regular rhythmic pattern: in several musical cultures, there are stable rhythmic patterns that are not based on isochrony (see Polak et al., 2016). Items that refer to isochrony (such as "beat" or "pulse" used in items 1, 3, or 16) may not be suitable to measure the impression of rhythmic regularity in such musical contexts. We decided to include the items in *Experiment 1*, because they were likely to be strongly correlated with the underlying *temporal regularity* dimension, but subsequently we excluded them from the final item selection in order to maximize the applicability of the scale in the largest possible range of musical/cultural settings.

The feedback of the expert participants allowed us to select items with good content validity as candidate items for the *inner representation of temporal regularity* or *time-related interest* scales. It also permitted to discard items that failed to properly address either of the constructs or that were unsuitable for some reason, awkwardly worded, or redundant. A total of nine items relating to *temporal regularity* (items 1–8 from Table 1 plus item 29) and nine items pertaining to *time-related interest* (items 9–15 plus items 30 and 31) were selected to be used in *Experiment 1*.

## Experiment 1: Exploration of 26 Questionnaire Items

In this part of the study, participants used 26 candidate questionnaire items to respond to eight musical stimuli in

the course of a listening experiment (*Experiment 1*); the items are listed in Table 2. Of these candidate items, nine were associated with *temporal regularity* (upper block of Table 2), nine with *time-related interest* (central block) – these items were selected on the basis of the pretest. Eight items were associated with *energetic arousal* (lower block); they were adapted from the POMS vigor scale.

One goal of this experiment was to use exploratory factor analysis in order to investigate the factor structure of the 26 items, based on participants' rating data. A second goal was to reduce the number of candidate items to 3–4 items per construct, which were best aligned with the underlying factors.

## Method

**Stimuli.** In *Experiment 1*, eight musical excerpts of 30 s duration were used as audio stimuli (for titles and acts, see Table 3). They were selected from commercially available recordings and represent a variety of styles and musical cultures. With this selection, we aimed at triggering strong positive or negative subjective reactions in the *temporal regularity*, *time-related interest*, and *energetic arousal* dimensions.

The stimuli were selected as follows: the second author compiled a list of 45 musical excerpts from his personal music collection. In an internal poll, the five authors assigned the stimuli to high or low *temporal regularity*, *time-related interest*, and *energetic arousal* categories. We selected a set of stimuli for which the authors agreed most in terms of their categorization and which covered many different combinations of categories (see Table 3). A complete discography of the stimuli is given in the Appendix (Table 1).

The stimuli were excerpts from the following recorded tracks:

1. “Baya Baya” (2001) from the Safri Duo’s album “Episode II” is a dance music track based on a pre-programmed electronic drum beat. The musicians add rhythms on various percussion instruments (drums, marimbaphone) that evoke music from the

Caribbean. We hypothesized that this music would be perceived as regular, rhythmically interesting, and energizing.

2. “The Astounding Eyes of Rita” is a track from the eponymous 2009 album by a quartet led by Tunisian oud player Anouar Brahem, together with bass clarinetist Klaus Gesing, bass guitarist Björn Meyer, and darbouka player Khaled Yassine. The oud is a short-necked lute used in countries of Northern Africa and the Near East. The darbouka is a drum instrument with a goblet-shaped body, mostly used in Egypt. The selection from the track is quiet and relaxed. It has a clear beat, but the darbouka and the offbeat oud/clarinet melody add plenty of rhythmic detail. We judged that this music would trigger impressions of high temporal regularity and rhythmic interest, but low energetic arousal in listeners.
3. “We Will Rock You” is a 1977 song by British rock band Queen. The selected section only features *a capella* singing and an iconic stomp-stomp-clap-(pause) rhythm recorded by the band members in many overdubs to simulate a crowd. The rhythm creates a strong sense of regularity and is highly energetic, but it is repetitive, and we did not believe it would be perceived as particularly interesting.
4. “Waterfall” (2011) by Rose (also known as Seth Horvitz) is a techno music track that starts with nothing more than a regular electronic bass drum beat plus an electronic hi-hat on the offbeats. We supposed that participants would hear the temporal regularity in this initial section of the track, but not experience much interest or energy.
5. “Fast City” was recorded live in 1980 by the jazz-rock band Weather Report and published on the compilation album “Live and Unreleased”. The selected passage is an extract from Joe Zawinul’s solo on electric piano, but the accompaniment played by bassist Jaco Pastorius, drummer Peter Erskine, and percussionist Robert Thomas Jr. is equally improvised; the passage could best be described as a collective improvisation. It is difficult to detect a beat in this music, but the rhythmic

**Table 3.** Stimuli for listening *Experiments 1* and 2. The categorical *regularity*, *interest*, and *energetic arousal* variables specify authors' assumptions regarding the ratings (high vs. low ratings) these stimuli were likely to obtain in a listening experiment.

No.	Title	Act/Composer	Regularity	Interest	Energetic Arousal
1	Baya Baya	Safri Duo	high	high	high
2	The Astounding Eyes of Rita	Anouar Brahem	high	high	low
3	We Will Rock You	Queen	high	low	high
4	Waterfall	Rrose	high	low	low
5	Fast City	Weather Report	low	high	high
6	Vilambit Laya	Ustad Alla Rakha & Ustad Zakir Hussein	low	high	low
7	Mars	John Coltrane & Rashied Ali	low	low	high
8	Lux Aeterna	György Ligeti	low	low	low



- interaction is infectious and energy levels are high. So we assumed that listeners would perceive this passage as non-regular, interesting, and energetic.
6. The sixth stimulus is drawn from the repertoire of classical Hindustani music. “Vilambit Laya” is the first track of the 2004 album “Shared Moments” by tabla players Ustad Alla Rakha and Ustad Zakir Hussain, together with sarangi player Ustad Sultan Khan. The tabla are high-pitched hand drums, and the sarangi is a bowed string instrument. The music is part of a slow introductory section preceding a raga. It does not have a strong regular beat, and the general atmosphere is quiet. Yet the tabla playing is rhythmically diverse and virtuosic. We categorized this extract as non-regular, high in rhythmic interest, and low in energetic arousal.
  7. The album “Interstellar Space” by saxophonist John Coltrane and drummer Rashied Ali was recorded in 1967, shortly before Coltrane’s death, and published in 1974. The album is a free jazz improvisation that was cut in one single take. The selected passage from the section entitled “Mars” is dense, wild, and loud. It is punctured by Coltrane’s multiphonic shrieks and Ali’s relentless drumming. It does not show any obvious structure in any musical dimension. We classified it as irregular, and highly energetic. Since the rhythm is very dense without having any discernable structure, we classified it as not interesting to the average listener (anticipating that many free jazz fans and experts would disagree).
  8. “Lux Aeterna” is a 1966 composition for 16-parts mixed a capella choir by Hungarian composer György Ligeti. It was famously used by Stanley Kubrick in his 1968 movie “2001 – A Space Odyssey”. The selection was taken from a 2008 recording with the Cappella Amsterdam under Daniel Reuss. The music is quiet, the choir voices change notes slowly and seemingly independently, which results in beautiful dissonant sounds and a lack of directionality. We classified this music as metrically irregular, rhythmically amorphous (thus uninteresting) and not triggering much energetic arousal.

The selection from each piece was 30 s long. In order to give participants enough time to rate the 26 candidate items, we repeated the passages in a loop, so participants heard them several times during a trial. The stimuli were chosen such that they did not change in terms of tempo, instrumentation, dynamics and mood during their 30 s runtime. The original audio files were purchased on iTunes or obtained from CD. The loudness of each excerpt was adjusted to an average-weighted RMS of –22 dB using *Audacity* (version 2.2.2) in order to obtain approximately equal loudness levels across stimuli. We added short fade-ins (1 s) and fade-outs (3 s) and exported the music in stereo to mp3 format (192 kbit/s).

**Participants.** In total,  $n = 59$  participants were recruited (29 via email at the Lucerne *University of Applied Sciences and Arts*; 30 through *Amazon’s MTurk*). They had a mean age of 38 years (ranging from 17 to 71 years,  $SD = 12$  years); 30 participants were female, 28 male, and one did not disclose gender information. Of the participants, 14 self-identified as professional musicians, 12 as music students, 12 as amateur musicians, 20 as music listeners, and one person indicated not to be interested in music. Participants had a mean score of 28.90 on the Gold-MSI training scale, so their musical training was non-significantly ( $t_{(57)} = 1.236$ ,  $p = 0.222$ ) greater than the UK population norm of 26.52 reported in Müllensiefen et al. (2014, p. 10). Participants lived in Switzerland ( $n = 26$ ), the USA ( $n = 23$ ), India ( $n = 4$ ), Italy ( $n = 1$ ), and the Netherlands ( $n = 1$ ). Four participants did not disclose information on their country of residence. The vast majority had at least an independent language level (level B, C or native speaker) in the language of the survey (English), three participants indicated, their language skills were basic (level A). The most preferred musical styles among participants were rock/rock’n’roll, classical, and jazz.

**Procedure.** The listening experiment was run online on the *SosciSurvey* platform ([www.soscisurvey.de](http://www.soscisurvey.de)). Participants gave informed consent and were asked to provide demographic information (age, gender, country of residence, musical expertise, musical taste). They were instructed to seek a quiet environment and use quality headphones to carry out the experiment. Participants were presented a test stimulus, and they adapted the playback loudness to a loud, but comfortable level. They were asked not to change the loudness level during the experiment.

Participants then listened to the eight stimuli (presented in a randomized order) and rated the degree to which they agreed with the 26 candidate items presented on the screen (see items in Table 2). The items were nine statements relating to the temporal regularity of the music (upper block of Table 2), nine statements relating to time-related interest (central block), and eight statements pertaining to energetic arousal (lower block). While listening to a stimulus, participants used a 7-point Likert scale to indicate how much they agreed with each statement with respect to the stimulus currently playing (*strongly disagree* = 0, *disagree* = 1, *slightly disagree* = 2, *neither agree nor disagree* = 3, *slightly agree* = 4, *agree* = 5, *strongly agree* = 6). This multi-point response scale is identical to the scale used in the Experience of Groove Questionnaire (Senn et al., 2020) for compatibility.

Each stimulus played on a loop during a trial. When the 26 ratings were complete, participants pressed the “Next” button. Then the music stopped, the data was transmitted to the *SosciSurvey* database, and participants moved on to the next trial. The sequence of trials/stimuli was randomized; and the order of the list of items on the online survey page was also randomized for each trial.

After the last trial, participants filled the Gold-MSI training scale (Müllensiefen et al., 2014); and they were informed about the purposes of the study in a debriefing. Participants took a median of 20 min to complete the survey; MTurkers received a remuneration of USD 6.25; the other participants received a CHF 10 voucher for consumption at the cafeteria of the music department of the Lucerne University of Applied Sciences and Arts (with the possibility to opt out); no partial course credit was offered to music students.

**Analysis.** A total of 472 ratings (59 participants  $\times$  8 stimuli) were collected on each of the 26 candidate questionnaire items. This number of observations was sufficiently large to carry out an exploratory factor analysis (Gorsuch, 2015, p. 350). The items had specifically been chosen to measure how strongly participants experienced *temporal regularity*, *time-related interest*, and *energetic arousal* while listening to the stimuli. Accordingly, we expected a factor analysis with three factors to have a good fit with the data. Inspection of a scree plot, parallel analysis (McCoach et al., 2013, p. 123), and Kaiser's criterion (Kaiser, 1958) confirmed the number of factors. We expected the factors to be correlated. Accordingly, models with an oblique *promax* rotation were fitted to the data. The analysis was carried out in *R* (version 4.1.2) and *R-Studio* (version 2021.09.2 Build 382) using the *factanal* function from the *stats* package (version 3.6.2).

## Results

**Exploratory factor analysis.** Table 2 (columns 3–5) shows the pattern matrix of the exploratory factor analysis (EFA). The table is ordered in three blocks that reflect our initial association of an item with a particular construct: *inner representation of temporal regularity* (top block), *time-related interest* (central block), or *energetic arousal* (bottom block). Items are listed in descending order according to the pattern coefficient of the dominant construct within each block.

The three factors accounted for 70% of the variance across all items. Most items had only one single pattern coefficient with a high absolute value, so most items can unambiguously be assigned to one of the three factors. The exceptions were:

- *Item 15* (“The rhythm of this music is full of surprises.”), which we expected to have a high positive loading on the *time-related interest* factor, but which instead loaded very negatively on the *temporal regularity* factor.
- *Item 38* (“This music makes me feel carefree.”), derived from the POMS vigor scale, which did not have a high loading on any of the three factors.

The inner representation of temporal regularity and time-related interest factors were quasi-orthogonal ( $r = 0.147$ );

temporal regularity and energetic arousal showed a weak positive correlation ( $r = 0.326$ ); time-related interest and energetic arousal showed a positive correlation of medium strength ( $r = 0.674$ ).

The next step was to select items for the final questionnaire scales. In order to provide sufficient degrees of freedom for the use of the scales in structural equation modelling, we aimed at choosing at least three items per scale that reliably measure the underlying construct. Items should have high content validity ratings (if available, see Table 1) and high pattern matrix coefficients (Table 2). Items should further address different semantic aspects of the respective constructs.

For the *inner representation of temporal regularity* scale, we selected items 29, 5, 2, and 6:

- *Item 29* (“The rhythm of this music gives a strong sense of regularity”) addresses the temporal regularity topic in a direct and straightforward way. The item has a very high pattern matrix coefficient for the temporal regularity factor (0.905), but no content validity rating, because it was proposed by one of the participants during the pretest.
- *Item 5* (“The rhythm of this music is steady”): According to the Cambridge Dictionary (Cambridge University Press, n.d.), the adjective “steady” characterizes processes that happen gradually and regularly, not suddenly or unexpectedly. These meanings agree with our definition of temporal regularity quite closely. The item received high content validity ratings in the pretest (0.753), and had a high pattern matrix coefficient (0.885) in the EFA. We preferred item 5 to the similar, but slightly more ambiguous item 8.
- *Item 2* (“The rhythm of this music is predictable”). This item captures the predictability of a temporally regular stimulus. If music shows temporal regularity, some aspects will repeat over time and become predictable to the listener (content validity: 0.926; pattern matrix coefficient: 0.867).
- *Item 6* (“The rhythm of this music sounds disorganized”). This is an inverted item which relies on the fact that “disorganized” is an antonym of “regular”. If the rhythm is perceived as disorganized, it will also sound as if it was irregular (content validity: 0.654; pattern matrix coefficient:  $-0.811$ ).

We did not consider items 1 (“I can hear a regular beat in this music.”) and 3 (“I can hear a clear pulse in this music.”) for the *inner representation of temporal regularity* scale, because they refer to isochrony, which does not capture all aspects that make rhythm regular.

For the *time-related interest* scale, we chose items 12, 11, and 14:

- *Item 12* (“I find the rhythm of this music to be interesting”): This item addresses time-related interest in a direct way. The item had high content validity

(0.821) in the pretest, and loaded very strongly on the *time-related interest* factor (0.993).

- *Item 11* (“I find the rhythm of this music to be fascinating”): Fascination implies that the rhythm attracts the attention of the listener; it adds a more emotional aspect to the interest topic (content validity: 0.833; pattern matrix coefficient: 0.945).
- *Item 14* (“The rhythm of this music bores me”): This inverted item uses the concept of boredom as an antonym to interest (content validity: 0.762; pattern matrix coefficient: −0.778).

The items for the *energetic arousal scale* were derived from the vigor sub-scale of the Profile of Mood States (POMS) questionnaire (McNair et al., 1971, 1992; Heuchert & McNair, 2012); they were not rated on content validity during the pretest. From this scale, we chose numbers 34, 37, 32, and 39:

- *Item 34* (“This music makes me feel energetic”): This item directly addresses the energy topic (pattern matrix coefficient: 0.931).
- *Item 37* (“This music makes me feel full of pep”): This item offers a more informal and colloquial perspective on feeling energized (pattern matrix coefficient: 0.884).
- *Item 32* (“This music makes me feel lively”): This item addresses the energetic arousal aspect from a holistic perspective associating energy with life and liveliness (pattern matrix coefficient: 0.874).
- *Item 39* (“This music makes me feel vigorous”): This item (pattern matrix coefficient: 0.826) has a biologicistic/health-related connotation that is less apparent in the other items.

We explicitly excluded item 33 (“This music makes me feel active”) from the scale, because it is directly associated with body movement as corporeal activity and belongs to the domain of content of the Experience of Groove Questionnaire’s *urge to move* scale.

## Experiment 2: Validation of the Scales

In the final development phase, we investigated whether the factor structure found in the first experiment replicates in a second experiment with an independent sample and with the reduced set of the 11 selected items (Table 4).

### Method

**Stimuli and procedure.** *Experiment 2* used the same stimuli and the same procedure as *Experiment 1* with respect to the online format (*SosciSurvey*), informed consent, and administration of the demographics, music expertise, music preference, and *MSI Training* questionnaires. In each trial, participants used the selection of 11 items to assess the stimuli. The sequence of trials/stimuli was randomized, and the item list was also presented in a randomized order on the survey page for each trial.

**Table 4.** Items selected for *Experiment 2*.

No.	Label	Item
29	R1	The rhythm of this music gives a strong sense of regularity.
5	R2	The rhythm of this music is steady.
2	R3	The rhythm of this music is predictable.
6	R4	The rhythm of this music sounds disorganized.
12	I1	I find the rhythm of this music to be interesting.
11	I2	I find the rhythm of this music to be fascinating.
14	I3	The rhythm of this music bores me.
34	E1	This music makes me feel energetic.
37	E2	This music makes me feel full of pep.
32	E3	This music makes me feel lively.
39	E4	This music makes me feel vigorous.

**Participants.** A total of  $n = 82$  valid datasets were collected, 24 of them via personal invitation, and the remaining 58 through *Amazon MTurk*. Participants had a mean age of 36 years (ranging from 20 to 70 with  $SD = 10.8$  years); 27 of them were female, and 55 male. They self-identified as professional musicians (20), amateur musicians (12), music listeners (49), or not interested in music (1). Their *MSI Training* scores had a mean of 25.43, which is nominally lower ( $t_{(81)} = -0.735$ ,  $p = 0.465$ ) than the UK population mean of 26.52 reported by Müllensiefen et al. (2014, p. 10). The majority of participants lived in the USA (52) or Switzerland (18). The remainder reported living in India (3), Norway (3), South Sudan (2), Germany, Italy, and Sri Lanka (1 each); one participant did not provide information about their country of residence. Most participants were native English speakers (53), the rest declared their use of English was competent (level C, 20 participants), independent (level B, 6), or basic (level A, 3). Their preferred musical styles were rock/rock’n’roll, pop, or classical. Participants used a median time of 18 min to fill the survey. *MTurkers* received USD 6.25 for their participation, other participants received a voucher worth CHF 10 for the university cafeteria (with the possibility to opt out).

**Analysis.** We collected 656 ratings ( $82 \text{ participants} \times 8 \text{ stimuli}$ ) for each of the 11 questionnaire items that subjectively measure the *inner representation of temporal regularity* (4 items), *time-related interest* (3 items), and *energetic arousal* (4 items). Based on the results of *Experiment 1*, we expected a confirmatory factor model with three correlated factors to have a good fit with the data. The number of factors was confirmed by a scree plot, parallel analysis, and Kaiser’s criterion. The *lavaan* (v. 0.6–3) package was used in *R* for confirmatory factor analysis and Cronbach’s  $\alpha$  estimates were calculated using the “alpha” function from the *psych* library.

### Results

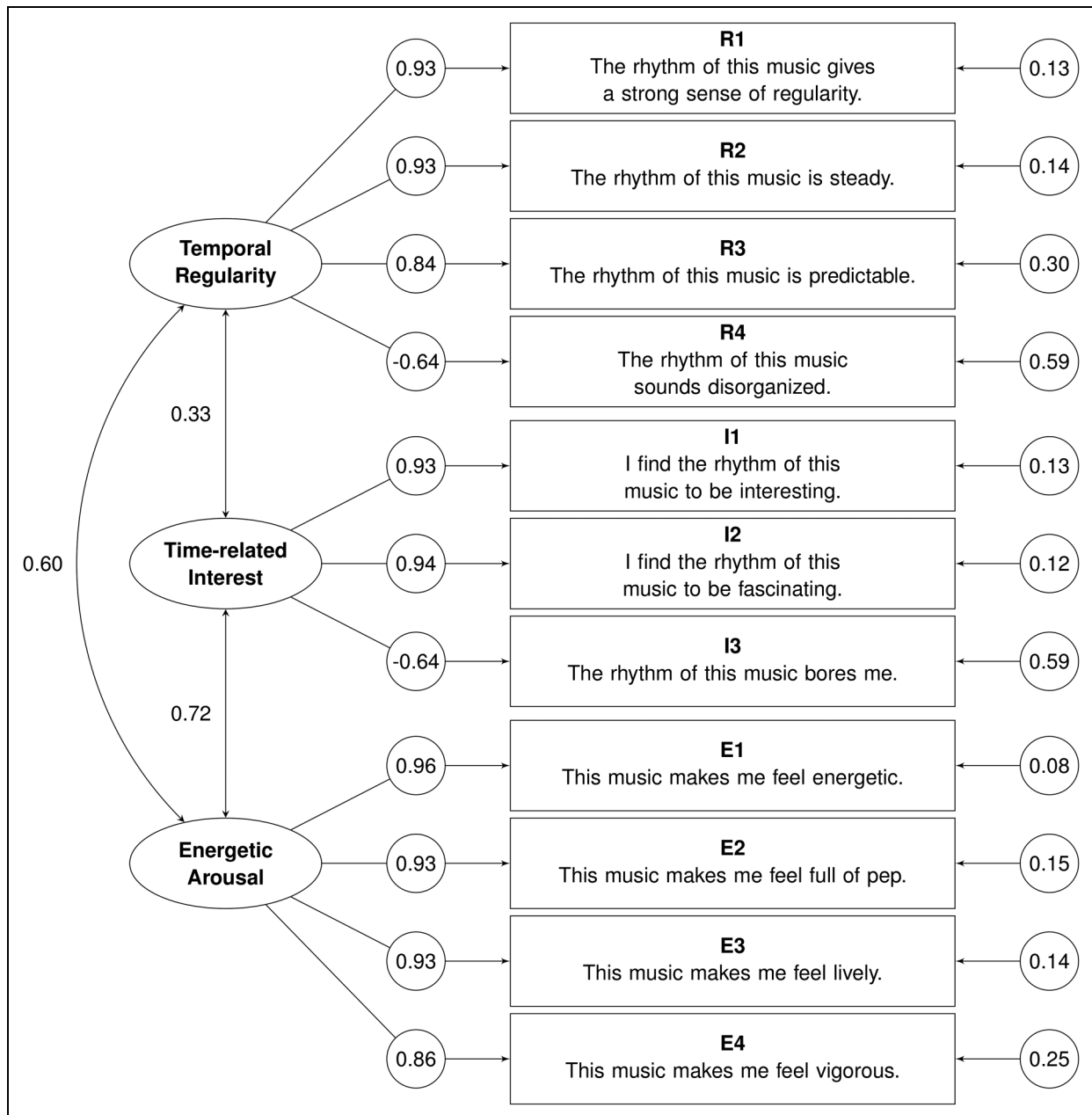
The target model consisted of three correlated latent variables, *inner representation of temporal regularity*

(measured by items R1–R4), *time-related interest* (measured by items I1–I3), and *energetic arousal* (measured by items E1–E4; for items and item labels, see Figure 2 and Table 4). The absolute standardized regression coefficients were high for all items (ranging from 0.64 to 0.96), and standardized uniquenesses were low (ranging from 0.13 to 0.59). CFI = 0.998 indicated a good fit, but RMSEA = 0.080 (90% CI: 0.069, 0.090) indicated only an adequate fit.

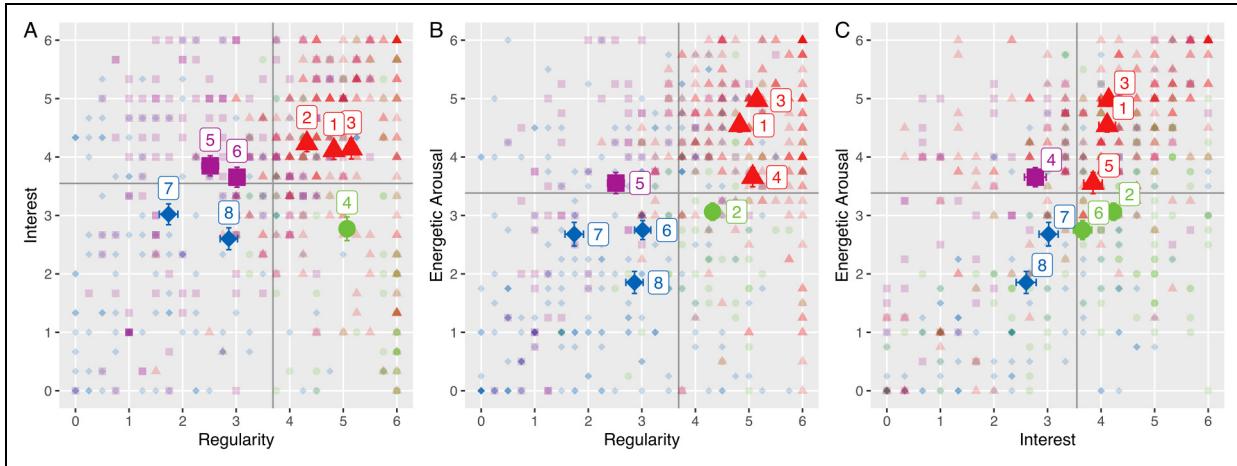
Figure 3 shows scatterplots of the mean ratings for each stimulus and each combination of scales (large symbols with error bars in the foreground) and individual ratings

(semi-transparent small symbols in the background). We observe that the mean *temporal regularity* ratings of the stimuli were in line with our expectations: stimuli 1–4 (see Table 3) had been selected as representants of music with regular rhythm, and stimuli 5–8 as examples with irregular rhythms. This categorization manifests itself in Figure 3 (A, B). The mean *temporal regularity* ratings per stimulus were widely dispersed with a standard deviation of 1.31.

The stimuli did not receive extreme mean ratings with respect to *time-related interest*: the means of the stimuli were close together with a standard deviation of only 0.66; this indicates that the stimuli were not clearly profiled



**Figure 2.** Confirmatory factor analysis model with the three latent variables *inner representation of temporal regularity* (measured by R1–R4, Cronbach's  $\alpha = .88$ ), *time-related interest* (I1–I3,  $\alpha = .85$ ), and *energetic arousal* (E1–E4,  $\alpha = .95$ ). Double arrows: Pearson correlation coefficients. Arrows from the left: factor coefficients. Arrows from the right: uniquenesses.



**Figure 3.** Scatterplots of mean scale values for each stimulus (large numbered symbols with error bars) and single ratings (semi-transparent small symbols) for each combination of scales: (A) *Temporal regularity* vs. *Time-related interest*; (B) *Temporal regularity* vs. *Energetic Arousal*; (C) *Time-related interest* vs. *Energetic Arousal*. Color and symbol shape depend on the quadrant of the stimulus mean. Stimuli: (1) Baya Baya (SAF); (2) The Astounding Eyes of Rita (BRA); (3) We Will Rock You (QUE); (4) Waterfall (RRO); (5) Fast City (WEA); (6) Vilambit Laya (UST); (7) Mars (COL); (8) Lux Aeterna (LIG). Grey lines indicate the mean of each scale across all ratings.

with respect to *time-related interest*. The categorization into high and low *time-related interest* stimuli (Table 3) worked for most of the stimuli, but misclassified Queen’s “We Will Rock You” (3) as a stimulus that raises low rhythmic interest (due to its repetitiveness). Listeners judged this stimulus to be above-average with respect to *time-related interest*.

The participants experienced clear differences of *energetic arousal* across the stimuli; the mean ratings of the eight stimuli had a standard deviation of 1.02. Yet, we had misclassified two stimuli: we had judged “Waterfall” (4) by Rose to be a stimulus that triggered little *energetic arousal*, but this was not reflected in the ratings. Participants judged this stimulus to be above-average in terms of experienced energy. Conversely, we thought the intense performance on “Mars” (7) by Coltrane and Ali would be heard as energizing, but participants rated it low on energetic arousal.

*Inner representation of temporal regularity* (Cronbach’s  $\alpha = .88$ ) and *time-related interest* ( $\alpha = .85$ ) showed good reliability (Tavakol & Dennick, 2011). The third scale, *energetic arousal*, even had excellent reliability ( $\alpha = .95$ ). All three scales appear to be positively correlated. The strongest correlations were observed between *energetic arousal* and *time-related interest* ( $r = .72$ ) and between *energetic arousal* and *temporal regularity* ( $r = .60$ ). The correlation between *time-related interest* and *temporal regularity* was weaker ( $r = .33$ ). The strong positive correlations raise the question whether the three scales (or subsets thereof) measure the same underlying psychological construct. A model in which the three factors were collapsed into one single factor had a significantly worse fit than the target model ( $\chi^2(3) = 3536$ ,  $p < .001$ ). But also two-factor models, in which the *Regularity* and *Interest* ( $\chi^2(1) = 2144$ ,  $p < .001$ ), *Regularity* and *Energy* ( $\chi^2(1) = 2021$ ,  $p < .001$ ) or *Interest* and *Energy* ( $\chi^2(1) = 943$ ,  $p <$

.001) dimensions were collapsed into one factor had a worse fit than the target model. We can thus assume that the three scales measure three different constructs.

The correlations of the three scales were much stronger than the factor correlations observed in *Experiment 1*. Since the factors and scales are supposed to measure the same underlying constructs across the two experiments, the increased strengths of the correlations observed in *Experiment 2* need to be further investigated. The strong correlations might at least partly have their roots in a lack of attention by the participants. We can explore this by analyzing the data of two participant sub-samples separately. Some participants were recruited at the authors’ university and within their personal networks (among them many musicians and music students). They can be expected to be interested in music in general and to carry out the trials attentively. The remaining participants were recruited through Amazon MTurk. They represent a general population; we may assume that their participation was primarily motivated by financial interest, and that they carried out the trials with less attention than the other participant sample.

The two sub-samples indeed show considerable differences in their rating behavior: the *inner representation of temporal regularity* scale was less reliable in MTurkers ( $\alpha = .84$ ) compared to the other participants ( $\alpha = .95$ ). The same is true for the *time-related interest* scale ( $\alpha = .84$  for MTurkers,  $\alpha = .90$  for others), but not for *energetic arousal* (MTurk:  $\alpha = .94$ , others:  $\alpha = .95$ ). The reduced reliabilities are caused to a considerable extent by inverted questionnaire items: each of the regularity and interest scales contains one inverted item (R4, I3, see Figure 2), whereas the energetic arousal scale only consists of items with the same orientation. In many cases, participants from the MTurk sample did not notice the change of

**Table 5.** Correlations between scales in two participant sub-samples.

	MTurk (n = 58)			Other participants (n = 24)		
	Regularity	Interest	Energy	Regularity	Interest	Energy
Regularity	—	0.553	0.667	—	−0.097	0.543
Interest	0.553	—	0.789	−0.097	—	0.495
Energy	0.667	0.789	—	0.543	0.495	—

directionality and rated the items as if they were not inverted, thus indicating inattentive rating behavior. This lack of attention can also be observed in the scale correlations (Table 5). In the MTurk sub-sample, all three composite scales show significant medium to large positive correlations. In the other sub-sample, the correlations were generally smaller, the *regularity* and *interest* scales were plausibly independent ( $r = -0.097$ ,  $z = -1.558$ ,  $p = .119$ , Wald approximation) and resembled the factor correlations observed in *Experiment 1*.

## Discussion

The three new scales have been shown to be reliable, and they appear to be valid operationalizations of the corresponding latent constructs outlined in the groove model (Senn et al., 2019, 2023). The content validity of the composite *inner representation of temporal regularity* and *time-related interest* scales was established in a pretest with music psychologists and rhythm and groove researchers. The scales' validity was further developed in *Experiment 2*, with stimuli that we considered to be likely to trigger low or high ratings with respect to the three scales (Table 3). This was confirmed by the data of *Experiment 2* in most cases. However, participants considered the *time-related interest* of the “We Will Rock You” (3) stimulus to be greater than we expected. We thought that participants would judge the repetitive stomp-stomp-clap-(pause) rhythm to be of little interest. Yet, we did not consider that the repetitive rhythm explicitly invites audience participation and this alone might attract participants' interest and attention.

The pretest validation by the expert panel was not required for the items of the *energetic arousal* scale, because it uses item stems of the POMS vigor scale, which targets a very similar experience to the *energetic arousal* scale (Terry et al., 1999, p. 863). Participants' responses in *Experiment 2* agreed with our expectations (Table 3) in six out of eight stimuli, further bolstering the validity of the scale. One exception was “Waterfall” (4), where the simple bass drum/hi-hat motif energizes listeners more than we expected. Potentially we underestimated that this motif is an essential element of electronic dance music which might be sufficient to trigger energetic arousal. Conversely, we had expected that the energetic performance of “Mars” (7) would also energize listeners, which proved to be wrong.

Of the overall 24 predictions we made in Table 3 about the effects of the stimuli on the listeners in the three

dimensions, 21 turned out to be correct, and three were wrong. The discrepancies or misplaced stimuli may be explained within the frame of the groove model (Senn et al., 2019, 2023): we authors judged the stimuli on the basis of our own personal background as music researchers and musicians with our own ideas on which kind of music is interesting or energizing, which in some cases might not agree with the opinions of members from other populations. In our assessment of “Mars” we obviously confounded perceived and induced energetic arousal (see Gabrielsson, 2001; Song et al., 2016).

The analysis of the data from *Experiment 2* showed that the scales reliably measure three distinct psychological constructs. However the substantial positive correlations between the three scales are a source of concern. The analyses of scale reliability and inter-scale correlations on the level of sub-samples revealed substantial differences with respect to rating behavior. In the responses of the participants who were personally invited by the authors (many of them with a professional musical background), scale reliabilities were generally high, and the scales showed some independence. In the responses of the MTurkers, scale reliabilities were reduced for scales with inverted items, and the composite scales showed high positive correlations. These observations provide evidence that MTurk participants were either less attentive during the experiment than the personal invitees, they read items less carefully, or were more confused by the task.

Our general assessment that the three scales are valid and reliable is not contradicted by the differences of the rating behaviors across participant groups. Yet, the (in-)dependence of the psychological constructs underlying the three scales is unclear as of today. Cumulatively, future uses of the scales will reveal to what extent the three scales depend on each other or are independent. However, we can give first recommendations about the administration of the scales in order to improve the data quality and avoid spurious positive correlations:

1. In *Experiment 2*, the 11 items of the three scales were presented together in one single block and in a randomized order. It appears that this form of item presentation is not ideal when participants are not sufficiently focused on their task, and a lack of attention affects their responses. The situation can potentially be improved by an item presentation on the survey page that bundles the items of each scale in one distinct block, maybe even using the

scale names as headers. This procedure was used in the first validation of the groove model, yet positive correlations between scales remained strong (see Senn et al., 2023, p. 295).

2. Ideally, each scale is measured in a separate listening trial. We can expect this procedure to reduce unwarranted positive correlations, because items from different scales do not appear on the same survey page, and participants will be able to better focus on the topic of one scale. On the downside, the decoupling of scales extends the duration of any survey considerably. This reduces the number of experimental conditions/stimuli that can be tested, and increases the number of times participants need to rate the same stimuli. For these reasons, the one scale=one trial approach might not be feasible in every experiment, because the prolongation represents an additional strain on participants' already scarce attention resource.
3. Finally, participant selection and data collection should be designed and implemented such that the attention of the participants and their dedication to the rating task can be maximized. This is a truism, since all empirical behavioral research relies on participants' readiness to engage with a survey or an experiment. As a general rule, we should vary methods of data collection (e.g., online vs. on site) in order to avoid systematic biases that arise as a consequence of the collection method. For the same reason, we should vary the pools from which we sample our participants and try to find incentives that properly motivate the participants apart from financial remuneration.

## Conclusions

This study aimed at the development of three psychometric scales for groove research that measure how strongly music listeners experience an *inner representation of temporal regularity*, *time-related interest*, and *energetic arousal* while listening to music. The development was successful: the three scales are concise (3–4 items per scale), show good content validity (as established in a pretest with the help of researchers from the groove, rhythm perception, and music psychology fields), good to excellent reliability (Cronbach's  $\alpha$  between 0.85 and 0.95), and their form is compatible with the *urge to move* and *pleasure* scales from the Experience of Groove Questionnaire (Senn et al., 2020; Düvel et al., 2021). The whole set of five scales (two from Senn et al., 2020; three developed by this study) will allow for the testing of the hypotheses formulated in the introduction of this paper and in Senn et al. (2019), and have already been used for this purpose (Senn et al., 2023).

The three scales are tailor-made for the concrete application of testing the groove model. However, they add to the general repertoire of psychometric instruments in music

psychology that might be used in a variety of settings. The *energetic arousal* scale might become relevant to the study of music as a motivator in sports and exercise (Terry et al., 2020), by directly connecting the effect of music to the subjective impression of energy. The *energetic arousal* scale might also be used to assess the vitalizing effect of music in the therapy of depression, Parkinson's, or burnout. The *temporal regularity* scale can be applied to subjectively assess the most generic forms of musical time organization (such as a regular beat). This may be combined with beat detection test batteries (e.g., Fujii & Schlaug, 2013) in order to study to what extent beat deafness sufferers are aware of their deficit. The *time-related interest* scale is likely to be useful in the study of rhythmic complexity and syncopation by indicating to what extent a conscious analysis of time and meter takes place in a listener.

## Acknowledgments

The authors would like to thank the following colleagues for kindly participating in the pretest: Birgitta Burger, Guilherme Schmidt Câmara, Daniel Cameron, Nina Düvel, Anders Friberg, Ronald Friedman, David Hammerschmidt, Fred Hosken, Petr Janata, Thomas Kaplan, Satoshi Kawase, Reinhard Kopiez, Douglas Kowalewski, Philippe Labonde, Daniel Levitin, Guy Madison, Rainer Polak, George Sioros, Dana Swarbrick, Maria Witek, Clemens Wöllner, Agata Zelechowska, and six colleagues who chose to remain anonymous.

## Action Editor

Jessica Grahn, Western University, Brain and Mind Institute & Department of Psychology.

## Peer Review

Haley Kragness, Bucknell University, College of Arts & Sciences, Department of Psychology. Alex Hofmann, University of Music and Performing Arts Vienna, Department of Music Acoustics.

## Author Contributions

OS: study design, data analysis, data visualization, draft of the manuscript, final manuscript. TB: preparation of audio stimuli, preparation of online survey, revision of the manuscript. RJ: selection of questionnaire items, revision of the manuscript. LK: selection of questionnaire items, revision of the manuscript. FH: preparation of questionnaire items, preparation of online survey, revision of the manuscript.

## Data Availability Statement

The experimental data is available from the *Supplemental Material* section of this article. The stimuli were excerpted from commercial recordings and cannot be published for reasons of intellectual property rights. Readers interested in the stimuli may contact the first author (olivier.senn@hslu.ch).

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.




## Ethics Statement


The design of this study was approved by the ethics commission of the Lucerne University of Applied Sciences and Arts on June 12, 2020 (decision letter EK-HSLU 002 M 20). All participants provided informed consent prior to taking part in the study.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, (grant number 100016\_192398).

## ORCID iDs

Olivier Senn  <https://orcid.org/0000-0002-4944-3115>

Toni Amadeus Bechtold  <https://orcid.org/0000-0002-4768-5997>

Florian Hoesl  <https://orcid.org/0000-0002-6700-5354>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Baker, F., Denniston, M., Zabora, J., Polland, A., & Dudley, W. N. (2002). A POMS short form for cancer patients: Psychometric and structural evaluation. *Psycho-Oncology*, 11(4), 273–281. <https://doi.org/10.1002/pon.564>
- Bowling, D. L., Graf Ancochea, P., Hove, M. J., & Fitch, W. T. (2019). Pupillometry of groove: Evidence for noradrenergic arousal in the link between music and movement. *Frontiers in Neuroscience*, 12(1039), 1–12. <https://www.frontiersin.org/article/10.3389/fnins.2018.01039>. <https://doi.org/10.3389/fnins.2018.01039>
- Cambridge University Press (n.d.). Steady. *Cambridge Dictionary*. Retrieved July 3, 2023, from <https://dictionary.cambridge.org/dictionary/english/steady>
- Curran, S. L., Andrykowski, M. A., & Studts, J. L. (1995). Short form of the profile of mood states (POMS-SF): Psychometric information. *Psychological Assessment*, 7(1), 80–83. <https://doi.org/10.1037/1040-3590.7.1.80>
- Dotov, D., Bosnyak, D., & Trainor, L. J. (2021). Collective music listening: Movement energy is enhanced by groove and visual social cues. *Quarterly Journal of Experimental Psychology*, 74(6), 1037–1053. <https://doi.org/10.1177/1747021821991793>
- Düvel, N., Labonde, P., Bechtold, T., Senn, O., & Kopiez, R. (2021). Experience of groove questionnaire: German translation and validation. *Music Perception: An Interdisciplinary Journal*, 39(1), 83–99. <https://doi.org/10.1525/mp.2021.39.1.83>
- Fujii, S., & Schlaug, G. (2013). The Harvard Beat Assessment Test (H-BAT): A battery for assessing beat perception and production and their dissociation. *Frontiers in Human Neuroscience*, 7(771), 1–16. <https://doi.org/10.3389/fnhum.2013.00771>
- Gabrielsson, A. (2001). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae*, 5(1 suppl.), 123–147. <https://doi.org/10.1177/10298649020050S105>
- Gorsuch, R. L. (2015). *Factor analysis: Classic edition*. Routledge.
- Grove, J. R., & Prapavessis, H. (1992). Preliminary evidence for the reliability and validity of an abbreviated profile of mood states. *International Journal of Sport Psychology*, 23(2), 93–109.
- Heuchert, J. P., & McNair, D. M. (2012). *Profile of Mood States (2nd Edition)*. Multi-Health Systems. /doiLanding?doi=10.1037%2F05057-000
- Hewston, R., Lane, A., & Karageorghis, C. I. (2008). Development and initial validation of the Music Mood-Regulation Scale. *E-Journal of Applied Psychology*, 4(1), 15–22. <https://doi.org/10.7790/ejap.v4i1.130>
- Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, 141(1), 54–75. <https://doi.org/10.1037/a0024208>
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200. <https://doi.org/10.1007/BF02289233>
- Madison, G. (2006). Experiencing groove induced by music: Consistency and phenomenology. *Music Perception: An Interdisciplinary Journal*, 24(2), 201–208. <https://doi.org/10.1525/mp.2006.24.issue-2>
- McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument Development in the Affective Domain: School and Corporate Applications* (3rd ed. 2013). Springer.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Profile of Mood States (POMS) Manual*. Educational & Industrial Testing Service.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). *Profile of Mood States (POMS) Revised Manual*. Educational & Industrial Testing Service.
- Mendoza, T. R., Wang, X. S., Cleeland, C. S., Morrissey, M., Johnson, B. A., Wendt, J. K., & Huber, S. L. (1999). The rapid assessment of fatigue severity in cancer patients. *Cancer*, 85(5), 1186–1196. [https://doi.org/10.1002/\(SICI\)1097-0142\(19990301\)85:5<1186::AID-CNCR24>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0142(19990301)85:5<1186::AID-CNCR24>3.0.CO;2-N)
- Merker, B. (2014). Groove or swing as distributed rhythmic consonance: Introducing the groove matrix. *Frontiers in Human Neuroscience*, 8(454), 1–4. <https://doi.org/10.3389/fnhum.2014.00454>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Polak, R., London, J., & Jacoby, N. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of Malian jembe drumming. *Frontiers in Neuroscience*, 10(285), 1–11. <https://doi.org/10.3389/fnins.2016.00285>
- Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening



- are related to degree of emotional arousal. *PLOS ONE*, 4(10), e7487. <https://doi.org/10.1371/journal.pone.0007487>
- Senn, O. (2023). A predictive coding approach to modelling the perceived complexity of popular music drum patterns. *Heliyon*, 9(4), e15199. <https://doi.org/10.1016/j.heliyon.2023.e15199>
- Senn, O., Bechtold, T., Hoesl, F., Jerjen, R., Kilchenmann, L., Rose, D., Baldassarre, A., Sigrist, C., & Alessandri, E. (2023). An SEM approach to validating the psychological model of musical groove. *Journal of Experimental Psychology: Human Perception and Performance*, 49(3), 290–305. <https://doi.org/10.1037/xhp0001087>
- Senn, O., Bechtold, T., Rose, D., Câmara, G. S., Düvel, N., Jerjen, R., Kilchenmann, L., Hoesl, F., Baldassarre, A., & Alessandri, E. (2020). Experience of groove questionnaire: Instrument development and initial validation. *Music Perception*, 38(1), 46–65. <https://doi.org/10.1525/mp.2020.38.1.46>
- Senn, O., Bechtold, T. A., Hoesl, F., & Kilchenmann, L. (2021). Taste and familiarity affect the experience of groove in popular music. *Musicae Scientiae*, 25(1), 45–66. <https://doi.org/10.1177/1029864919839172>
- Senn, O., Rose, D., Bechtold, T., Kilchenmann, L., Hoesl, F., Jerjen, R., Baldassarre, A., & Alessandri, E. (2019). Preliminaries to a psychological model of musical groove. *Frontiers in Psychology*, 10(1228), 1–5. <https://doi.org/10.3389/fpsyg.2019.01228>
- Shacham, S. (1983). A shortened version of the profile of mood states. *Journal of Personality Assessment*, 47(3), 305–306. [https://doi.org/10.1207/s15327752jpa4703\\_14](https://doi.org/10.1207/s15327752jpa4703_14)
- Song, Y., Dixon, S., Pearce, M. T., & Halpern, A. R. (2016). Perceived and induced emotion responses to popular music: Categorical and dimensional models. *Music Perception*, 33(4), 472–492. <https://doi.org/10.1525/mp.2016.33.4.472>
- Spiech, C., Sioros, G., Endestad, T., Danielsen, A., & Laeng, B. (2022). Pupil drift rate indexes groove ratings. *Scientific Reports*, 12(1), 11620. <https://doi.org/10.1038/s41598-022-15763-w>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Terry, P. C., Karageorghis, C. I., Curran, M. L., Martin, O. V., & Parsons-Smith, R. L. (2020). Effects of music in exercise and sport: A meta-analytic review. *Psychological Bulletin*, 146(2), 91–117. <https://doi.org/10.1037/bul0000216>
- Terry, P. C., Lane, A. M., & Fogarty, G. J. (2003). Construct validity of the profile of mood states—adolescents for use with adults. *Psychology of Sport and Exercise*, 4(2), 125–139. [https://doi.org/10.1016/S1469-0292\(01\)00035-8](https://doi.org/10.1016/S1469-0292(01)00035-8)
- Terry, P. C., Lane, A. M., Lane, H. J., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. *Journal of Sports Sciences*, 17(11), 861–872. <https://doi.org/10.1080/026404199365425>
- Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal*. Oxford University Press.
- Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, 1423(1), 19–29. <https://doi.org/10.1111/nyas.13622>
- Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music – brain responses to rhythmic incongruity. *Cortex*, 45(1), 80–92. <https://doi.org/10.1016/j.cortex.2008.05.014>
- Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5(1111), 1–14. <https://doi.org/10.3389/fpsyg.2014.01111>
- Witek, M. A. G. (2017). Filling in: Syncopation, pleasure and distributed embodiment in groove. *Music Analysis*, 36(1), 138–160. <https://doi.org/10.1111/musa.12082>
- Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, Body-Movement and Pleasure in Groove Music. *PLoS ONE*, 9(4). <https://doi.org/10.1371/journal.pone.0094446>

## Appendix

**Table 1.** Discographic information about the stimuli.

No.	Title	Artists (Composer)	Album, Label (Release Year)	Minutage	Siglum
1	Baya Baya	Safri Duo	Episode II, Universal (2001)	00:55–01:25	SAF
2	The Astounding Eyes of Rita	Anouar Brahem (Anouar Brahem)	The Astounding Eyes of Rita, ECM (2009)	01:25–01:55	BRA
3	We Will Rock You	Queen (Brian May)	News of the World, EMI (1977)	00:00–00:30	QUE
4	Waterfall	Rose	Merchant of Salt, Sandwell District (2011)	00:00–00:30	RRO
5	Fast City	Weather Report (Joe Zawinul)	Live and Unreleased, Columbia (2002)	04:30–05:00	WEA
6	Vilambit Laya	Ustad Alla Rakha & Ustad Zakir Hussain	Shared Moments, Navras (2004)	13:02–13:32	UST
7	Mars	John Coltrane & Rashied Ali (John Coltrane)	Interstellar Space, Impulse (1974)	05:45–06:15	COL
8	Lux Aeterna	Daniel Reuss & Cappella Amsterdam (György Ligeti)	Ligeti: Lux Aeterna, Harmonia Mundi (2008)	05:07–05:37	LIG