SONIFYING WALKING: A PERCEPTUAL COMPARISON OF SWING PHASE MAPPING SCHEMES

Prithvi Ravi Kantan, Sofia Dahl, Erika G. Spaich

Aalborg University
Denmark
prka,sof@create.aau.dk
espaich@hst.aau.dk

Roberto Bresin

KTH Royal Institute of Technology
Sweden
roberto@kth.se

ABSTRACT

Past research on the interactive sonification of footsteps has shown that the signal properties of digitally generated or processed footstep sounds can affect the perceived congruence between sensory channel inputs, leading to measurable changes in gait characteristics. In this study, we designed musical and nonmusical swing phase sonification schemes with signal characteristics corresponding to high and low 'energy' timbres (in terms of the levels of physical exertion and arousal they expressed), and assessed their perceived arousal, valence, intrusiveness, and congruence with fast (5 km/h) and slow (1.5 km/h) walking . In a web-based perceptual test with 52 participants, we found that the nonmusical high energy scheme received higher arousal ratings, and the musical equivalent received more positive valence ratings than the respective low energy counterparts. All schemes received more positive arousal and valence ratings when applied to fast walking than slow walking data. Differences in perceived movement-sound congruence among the schemes were more evident for slow walking than fast walking. Lastly, the musical schemes were rated to be less intrusive to listen to for both slow and fast walking than their nonmusical counterparts. With some modifications, the designed schemes will be used during walking to assess their effects on gait qualities.

1. INTRODUCTION

Multimodal interactive systems can alter human perception of motor behavior, opening avenues for technological applications for the rehabilitation of movement impairments [1]. This may cater to patients suffering from orthopedic ailments (e.g. fractures, ligament tears) and neurological conditions (e.g. stroke, traumatic brain injury). Walking is an important activity of daily life and key determinant of longevity in older adults [2]. It is a highly complex movement whose kinematic properties are mediated by visual, auditory, proprioceptive, and tactile sensory feedback [3, 4, 5].

In recent years, many researchers have designed and tested interactive paradigms aimed at providing gait-related auditory feedback, either by sonifying movement parameters [6, 7] or generating/altering footstep sounds [8, 9, 10]. Studies have found even simple manipulations of footstep sounds (such as time delays [10] and spectral modification [9]) to lead to significant changes in gait parameters and emotional experiences, specifically pertaining to arousal and perceived agency over the sound. Other studies [8, 11] have applied more complex physics-based synthesis models to generate footstep sounds corresponding to various firm and aggregate surfaces (e.g. wood, gravel, snow) while users walked on asphalt and wood, finding that semantic and temporal incongruences between the haptic and auditory feedback led to slower walking speeds and greater deviation from normal gait parameters [8]. Comparable effects of inter-modality feedback incongruence were seen during a surface tapping task [1], where users exhibited inferior tapping ability and unpleasant arousal experiences as the auditory feedback became increasingly incongruent with tactile information. The authors concluded that inter-modality feedback congruence is an important criterion in determining how actions are modulated by multimodal interactive systems [1].

Temporal gait parameters have been shown to be modulated by the emotional intention (in terms of arousal and valence) of the walker [11], with significant differences depending on whether the walking style is happy, sad, tender, or aggressive. Another study showed that walking sounds and music shared commonalities in terms of their emotionally expressive features, specifically highlighting features related to sound intensity, tempo, and tempo regularity [12]. These align well with those identified in earlier work on musical expression and communication of emotion over the past decades [13, 14, 15]. In terms of sound-related emotion, the determinants of perceived and induced arousal and valence have also been explored in a machine learning analysis [16], which listed signal features related to dynamics, spectral flux, spectral roughness, roll-off, brightness, etc. as major factors influencing perceived and induced arousal and valence. Related signal features have also been identified as key differentiators between 'activating' and 'relaxing' music, which were found to elicit different gait speeds in a spontaneous walking experiment [17]. Here, we refer to these timbral qualities in terms of a high-level attribute called 'energy' which represents physical exertion and arousal expressed in the sound [18], which humans decode from sound through inverse modelling processes driven by motor mimetic mechanisms [18, 19].

There is clear untapped potential for the development of interactive sonification systems that can help modulate human gait to fit rehabilitation and exercise goals. The majority of existing sonification schemes have focused on the portion of the gait cycle where the foot is in contact with the ground (stance phase), with very little attention paid to the non-contact portion (swing phase), which accounts for roughly 40% of the duration of a gait cycle as well as the entirety of the forward limb movement, and is intrinsically linked with walking speed [20]. Proprioceptive feedback is known to be an important component of motor control during the swing phase [5]. Given the known interplay between haptic and auditory feedback during the stance phase [8, 9, 10], we posit that by manipulating the arousal and valence properties of a swing

DOI: 10.5281/zenodo.7552228

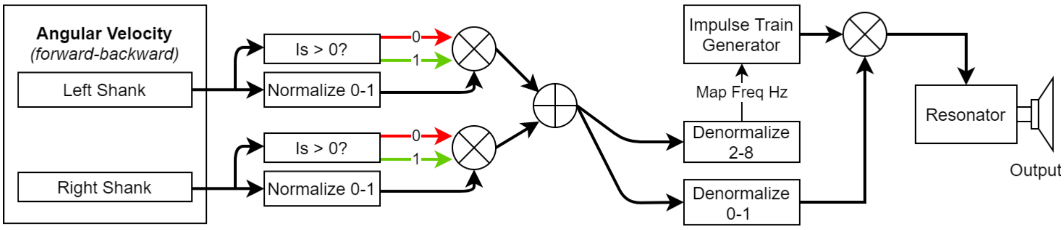

Figure 1: Block diagram for the creaking scheme (nonmusical-low energy).

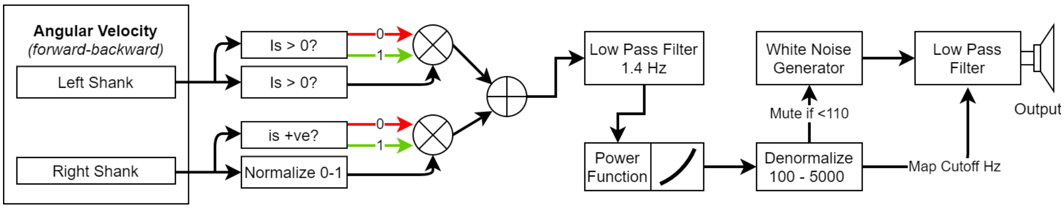

Figure 2: Block diagram for the whooshing scheme (nonmusical-high energy).

phase sonification, it is possible to alter gait parameters by mediating the perceived congruence between sensory inputs from the involved modalities. In addition, there is evidence that sonification designs integrating musical pitch structures and instrumental textures can elicit superior user experience to comparable nonmusical designs [21], and music has been known to readily induce and motivate movement [22]. However, we could not find any studies that directly compared musical and nonmusical schemes for gait sonification.

In this study, we designed and perceptually evaluated a set of musical and nonmusical swing sonification schemes with distinct arousal and valence properties - a necessary step prior to a full-fledged evaluation procedure with real-life walkers. The subsequent sections outline details of the sonification design, technical implementation, and our web-based perceptual evaluation.

2. DESIGN AND IMPLEMENTATION

2.1. Technical Setup

We built a gait sonification framework by modifying and upgrading the real-time system built in [23]. The hardware component for this study comprises two ESP32-based M5Stack Grey microcontrollers [1] equipped with 9-DoF MPU9250 IMU chips. These transmit IMU data as OSC packets over WiFi using the UDP protocol. The packets are received and processed by a JUCE-built software application [2] running on a Windows laptop. A TP-Link Archer C20 wireless router is used to provide a dedicated 2.4 GHz wireless network with 300 Mbps bandwidth for data transmission. To generate the movement measurements and sonified audio material for the current study, two IMUs (mounted on the outer side of each shank) were used.

The JUCE software provides functionality for sensor configuration, movement visualization, movement/audio parameter mapping, and audio mix configuration in real-time. The audio DSP functionality was implemented in the FAUST programming language [3] and compiled as a JUCE-compatible class. It includes a range of melodic and percussive physics-based instrument models (e.g. djembe, guitar, flute, voice) whose synthesis parameters can be mapped to movement parameters computed from the IMU data (e.g. body segment orientations and angular velocities, joint angles, gait events) through a mapping matrix. Once assigned to synthesis parameters, movement parameters are normalized within their user-defined bounds and can undergo smoothing, polarity inversion, nonlinear transformation, and quantization before being denormalized to the configured range of the mapped audio parameter. This process is carried out for all mapped movement and audio parameters (see [23] for more information). The software can also stream pre-recorded IMU data from a collection of log files and 'play back' the data in real-time for sonification and visualization purposes.

2.2. Mapping Design

Using the mapping matrix, we designed sonification schemes under two sonic categories: *nonmusical (N)* and *musical (M)*. Each category comprised two mapping schemes - *high-energy (H)* and *low-energy (L)*. To be clear, the *nonmusical* category refers to relatively broadband unpitched synthetic sounds, whilst the *musical* category represents pitched instrument sounds playing musical note frequencies. We chose sounds having excitation signals that were continuous in the time domain (friction and blowing respectively) so as to directly mirror and capture the continuous nature of forward limb swing during walking. Waveform and spectrogram plots are shown in Fig. 4, and audiovisual demonstrations at two walking speeds are available online [4].

Nonmusical Category: This category was inspired by friction sound models used in past work [24] and is characterized by a

[1] `https://shop.m5stack.com/products/grey-development-core?variant=16804796006490`

[2] `https://juce.com/`

[3] `https://faust.grame.fr/`

[4] `https://drive.google.com/drive/u/1/folders/1PfG2ZjgFt-wY9tbNO64_mnUkXN7LYeXF`

DOI: 10.5281/zenodo.7552228

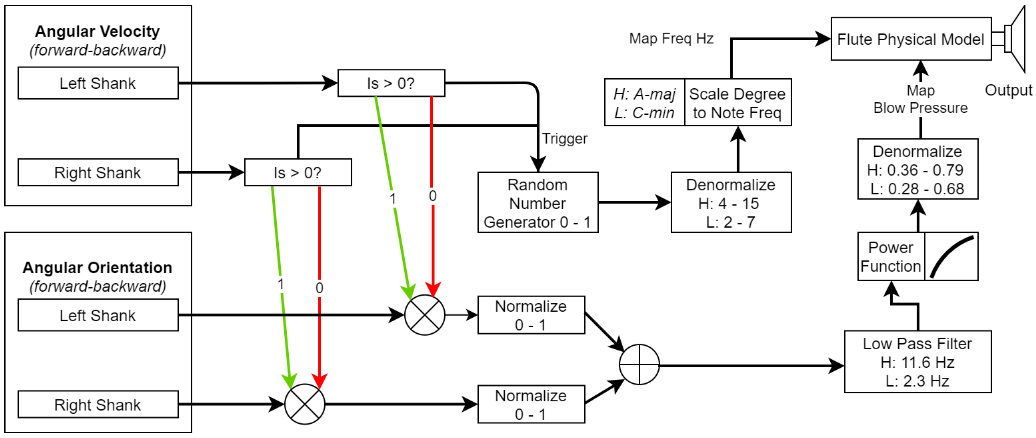

Figure 3: Block diagram for the flute-based musical schemes. The 'H' values are those used for the high-energy design and 'L' represent the low-energy design.

broadband sound without a predominant perceived pitch. We designed them in such a way that the sounds were only generated during the swing phase when the lower extremity was moving forward relative to the torso (positive rotation direction). The sound properties were determined by the angular velocity of the respective shank in the sagittal plane (side view) as shown in Figs. 1 and 2.

- Low-Energy (Creaking sound): The shank angular velocities were mapped to the frequency of an impulse train signal which was fed to a vocal tract-inspired resonator and rendered as a monophonic audio output as shown in Fig. 1. The auditory result was a simulation of a creaking sound whose intensity and timbre were controlled by the forward swing (top-left panel of Fig 4).

- High-Energy (Whooshing sound): This scheme comprised a mapping of shank angular velocities to the cutoff frequency of a low-pass filter applied to a white noise signal (see Fig. 2). This resulted in a limb-controlled whooshing sound with greater spectral bandwidth and a brighter timbre than the creaking scheme (top right panel of Fig 4).

Musical Category: Here, we used the physics-based model of a flute as the basis of both the low- and high-energy schemes because it has been found to have relatively neutral emotional associations in the context of music performance [15]. Similar to the nonmusical schemes, the flute was only audible during forward swing, and played *one* randomly chosen scale note per swing phase (notes were randomly selected from the A major or C minor tonalities for H and L schemes respectively). The *H* and *L* schemes differed in terms of several arousal- and valence-related signal properties based on [12, 15], specifically (a) envelope attack time (manipulated using a Butterworth smoothing filter), (b) blowing excitation intensity, and (c) note register and scale (see Fig. 3 and bottom panels of Fig 4). A digital reverberation effect from the FAUST libraries was applied to the model output to enhance its realism.

3. PERCEPTUAL EVALUATION

We carried out a web-based perceptual evaluation of the sonification schemes with two aims:

- To assess whether the sounds would be perceived to have the distinct arousal and valence properties that were intended during their design, and whether user perceptions varied between the musical or nonmusical schemes.
- To evaluate the extent to which arousal and valence properties affected the perceived congruence between the sound and the corresponding visually observed walking movement at two different walking speeds.

At the outset, we formulated the following hypotheses:

- **H1:** High-energy (H) sounds will receive higher perceived arousal ratings and more positive valence ratings than low-energy sounds (L) for each given sonic category and walking speed.
- **H2:** Sounds generated from fast walking will receive higher perceived arousal ratings and more positive valence ratings than those generated from slow walking for each given sonic category and energy level.
- **H3:** The energetic qualities of high-energy sounds will be rated as more congruent with those of fast walking than slow walking (and vice versa for low-energy sounds) for each given sonic category.
- **H4:** Musical sounds will be rated to be less intrusive (mentally disturbing) during walking than nonmusical sounds for each given walking speed and energy level.

3.1. Stimuli

Shank IMU recordings from one healthy 30 y/o male walker were captured over two minutes of treadmill walking at two fixed speeds - 1.5 km/h *(slow)* and 5 km/h *(fast)* and sampled at 100 Hz. Videos were also recorded simultaneously at 30 fps from a sideways angle using a Moto G8 camera phone. Using our software, we generated sonified sequences from the IMU data. The sequences corresponded to each of the four mappings (2 sonic categories × 2

DOI: 10.5281/zenodo.7552228

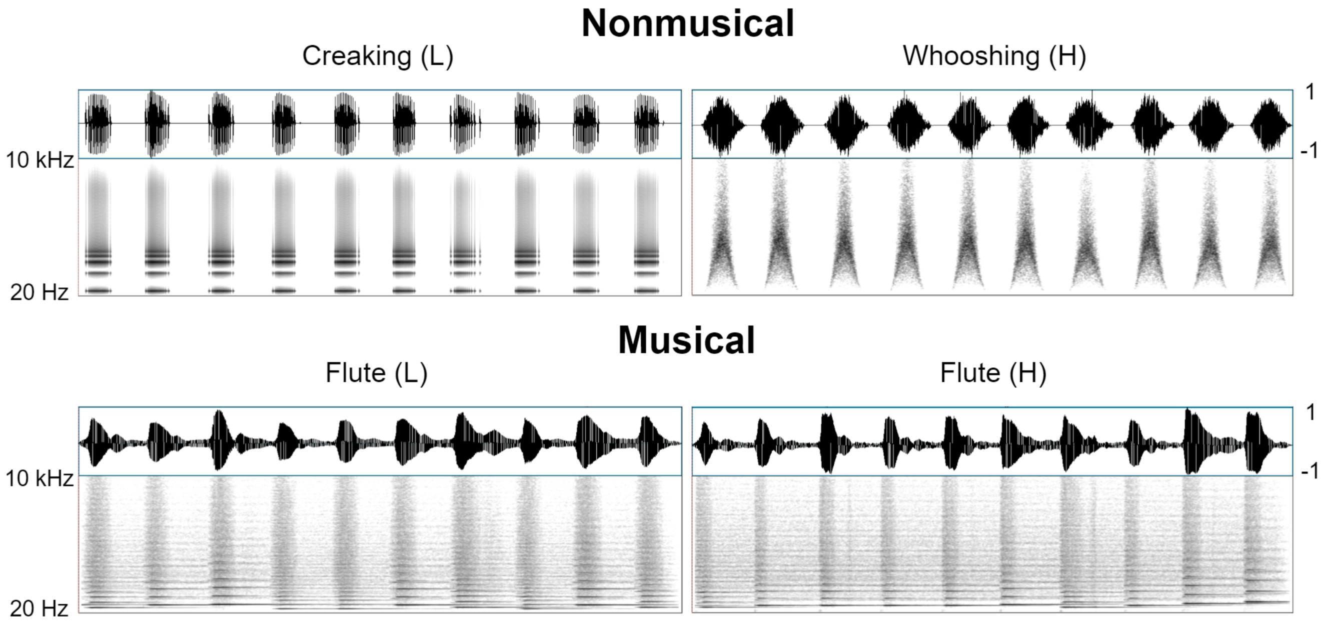

Figure 4: PRAAT-generated waveform and spectrogram plots for each mapping scheme when sonifying a 12-sec excerpt of shank IMU data recorded during treadmill walking at 1.5 kmph. It is possible to see the spectral differences between the creaking and whooshing schemes, as well as the attack time and dynamic range differences in the flute L and H waveforms.

energy levels) for both IMU recordings and we rendered them as WAV files sampled at 48 kHz/24 bit resolution.

These were imported into a REAPER session along with the video recordings and synchronized accurately by transient alignment using the tab-to-transient function in REAPER. The sonified sequences were adjusted by ear to have roughly equal loudness. To make the motion of the lower extremities easily discernible, the video was converted to grey-scale and underwent brightness and contrast adjustment followed by edge detection filtering (see video materials). From the original clips, two sets of 12 sec excerpts were randomly chosen for each walking speed (slow and fast). One set was rendered as audio-only (48 kHz/24-bit stereo WAV files), and the other one as audio+video (h.264 encoded M4V files) for each of the four mappings. Hence, the complete set of stimuli comprised eight audio clips and eight video clips (2 sonic categories $\times$ 2 energy levels $\times$ 2 walking speeds).

3.2. Participants

A convenience sample of 52 participants (33 men, 19 women) aged 44 $\pm$ 15.7 years (ranging from 25-82) were invited to participate via mailing lists and social media. The evaluation was conducted anonymously, and no sensitive information was collected.

3.3. Experimental Procedure and Outcomes

The evaluation was set up as a survey on Google Forms. Participants were initially briefed about the purpose of the research as well as the structure of the survey, after which they were instructed to use headphones and presented with a 1 kHz sine tone as a sound intensity reference to help them adjust their listening volume to a comfortable level. The survey was divided into two main parts - *Audio Only* and *Audio + Video*:

- Audio Only
  - Arousal Ratings: Participants listened to each of the eight audio clips (presented in a random order) and were asked to rate on a 9-point scale the arousal level of the sounds (1 = very calm/passive, 9 = very excited/active).
  - Valence Ratings: They then listened to the same clips in a different random order and were asked to rate the valence of the sounds (1 = very sad, 9 = very happy).
- Audio + Video
  - Congruence Ratings: Participants were then presented with the eight audio+video clips in random order and asked to rate the level to which they felt the energetic properties of the sound matched the energy they observed in the walking movement (1 = doesn't match at all, 9 = perfect match).
  - Intrusiveness Ratings: The same clips were shuffled and participants were asked to rate how mentally disturbing/bothersome they would perceive the sounds to be if they had to listen to them while walking (1 = not at all disturbing, 9 = very disturbing).

After completing the above, participants were asked to specify their age and gender, and subsequently answered six questions about their music perception and emotion-specific cognition abilities (selected from the Goldsmith Musical Sophistication Index questionnaire [25]). Completing the survey took approximately 10-15 minutes.

DOI: 10.5281/zenodo.7552228

Factor Combo	Arousal	Valence	Congruence	Intrusive-ness
M - H	***	***	NS	NS
M - L	***	***	***	NS
N - H	**	***	***	NS
N - L	***	***	NS	NS

Table 1: A summary of detected significant differences between slow and fast walking for each combination of sonic category and energy level shown for all outcomes based on the results of the Bonferroni-corrected Wilcoxon signed-rank tests. * = p $<$ 0.05, ** = p $<$ 0.01, *** = p $<$ 0.001, NS = non-significant differences.

3.4. Data Analysis

The rating data were exported from Google Forms in a comma-separated format, rearranged into matrices using MATLAB 2018b, and statistically analyzed in SPSS 27.0. In correspondence with the experimental structure and the collected ratings, we aimed to analyze the effects of (a) Walking Speed (slow, fast), (b) Sonic Category (musical, nonmusical), and (c) Energy Level (high, low) on *Arousal, Valence, Congruence, and Intrusiveness* ratings. We first checked all data for normality (Shapiro-Wilks test) and homogeneity of variance (Levene's test) for each set of factors, and found the distributions to both exhibit significant deviations from normality and significantly non-homogeneous variance between factor levels. Therefore, we chose to adopt a non-parametric repeated measures analysis for each outcome. We first checked for main effects across all eight factor level combinations using Friedman tests. If significant effects were detected, planned pairwise comparisons were carried out using Wilcoxon signed-rank tests. For each factor, inter-level comparisons were carried out only between equivalent pairs of *other* factor combinations. This allowed us to validate H1-4 by (a) studying the effects of each individual factor while simultaneously accounting for the others, and (b) reducing the total number of pairwise comparisons to 12 of a possible 28. A significance criterion $\alpha = 0.05$ was used for all statistical analyses. The reported p-values are those obtained post-Bonferroni correction.

4. RESULTS

In terms of self-reported musical sophistication, the participants had mean (std. dev) aggregated scores of 19.63 (5.80) and 20.57 (5.36) for music-related emotion and auditory perception respectively (max possible score 27).

The Friedman tests showed significant main effects for Arousal ($\chi^2(7)$ =201.14, $p < 0.001$); Valence ($\chi^2(7)$ =139.86, $p < 0.001$); Congruence ($\chi^2(7)$ =48.57, $p < 0.001$); and Intrusiveness ($\chi^2(7)$ =200.46, $p < 0.001$). The results of the planned Wilcoxon signed-rank comparisons are shown in Fig. 5 and Table 1.

4.1. Perceived Arousal

Walking Speed had a strong effect on perceived arousal, with the fast walking clips receiving significantly higher ratings for all combinations of *Sonic Category* and *Sound Energy*, specifically **M-H** (Z = -4.18, $p < 0.001$), **M-L** (Z = -4.19, $p < 0.001$), **N-H** (Z = -4.0, $p < 0.01$), and **N-L** (Z = -4.98, $p < 0.001$). Ratings also differed based on sonic category for each *Walking Speed* and *Sound Energy* combination; with slow walking, the **M** clips received significantly lower ratings than **N** for both energy levels - **H** (Z = -5.82, $p < 0.001$) and **L** (Z = -4.17, $p < 0.001$). The same trend was seen with fast walking - (Z = -5.92, $p < 0.001$) and (Z = -3.66, $p < 0.01$) respectively. *Energy Level* also impacted perceived arousal ratings, but only for the **N** clips, where **N-H** received significantly higher ratings than **N-L** for both fast walking (Z = -4.89, $p < 0.001$) and slow walking (Z = -5.30, $p < 0.001$). No differences were seen between **M-H** and **M-L** for fast (Z = -0.1, $p = 0.92$) or slow (Z = -0.63, $p = 0.53$) walking.

4.2. Perceived Valence

There was a strong effect of *Walking Speed* on perceived valence; the fast walking clips were rated as significantly happier-sounding for all combinations of *Sonic Category* and *Sound Energy*, namely **M-H** (Z = -5.49, $p < 0.001$), **M-L** (Z = -4.75, $p < 0.001$), **N-H** (Z = -3.56, $p < 0.001$), and **N-L** (Z = -4.47, $p < 0.001$). There were also differences based on *Sonic Category* at both walking speeds, with the **M** clips being perceived to sound significantly happier than their **N** counterparts. This effect was strong with fast walking for both **H** (Z = -5.67, $p < 0.001$) and **L** (Z = -3.67, $p < 0.001$) clips. A similar, albeit less pronounced significant effect was seen with slow walking for **H** (Z = -3.30, $p = 0.012$) and **L** (Z = -3.27, $p = 0.012$) clips. There was an effect of *Energy Level* (particularly for **M** clips), with **M-H** rated significantly happier-sounding than **M-L** with both fast walking (Z = -4.55, $p < 0.001$) and slow walking (Z = -3.7, $p < 0.001$). For the **N** clips, a difference was seen between **N-H** and **N-L** with fast walking (Z = -3.034, $p = 0.024$), but not with slow walking (Z = -1.44, $p = 1.00$).

4.3. Perceived Congruence

For 2 out of 4 combinations of *Sonic Category* and *Energy Level*, there was a strong effect of *Walking Speed* on participant ratings of perceived congruence between the energetic qualities of the sound and the observed walking. Specifically, **N-H** was rated as significantly more congruent with fast walking than slow walking (Z = -4.38, $p < 0.001$), and **M-L** significantly more congruent with slow walking than fast walking (Z = -4.35, $p < 0.001$). With *slow* walking, we also observed the following effects: (a) *Sonic Category* - **M-H** was rated as significantly more congruent than **N-H** (Z = -3.23, $p = 0.012$), (b) *Sound Energy* - **N-L** was rated as significantly more congruent than **N-H** (Z = -4.33, $p < 0.001$). With fast walking, congruence ratings were consistently on the high side (median above 5) for all factor combinations with no differences between any pairs of them.

4.4. Perceived Intrusiveness

For all combinations of *Walking Speed* and *Energy Level*, the **N** clips were rated as being more intrusive than the **M** clips if they were to be listened to while walking. **N-H** received significantly higher intrusiveness ratings than **M-H** for fast (Z = -5.42, $p < 0.001$) as well as slow (Z = -6.03, $p < 0.001$) walking. A similar difference was observed between **N-L** and **M-L** for fast (Z = -5.33, $p < 0.001$) and slow (Z = -5.79, $p < 0.001$) walking. No differences were seen between slow and fast walking or high- and low-energy clips within each sonic category.

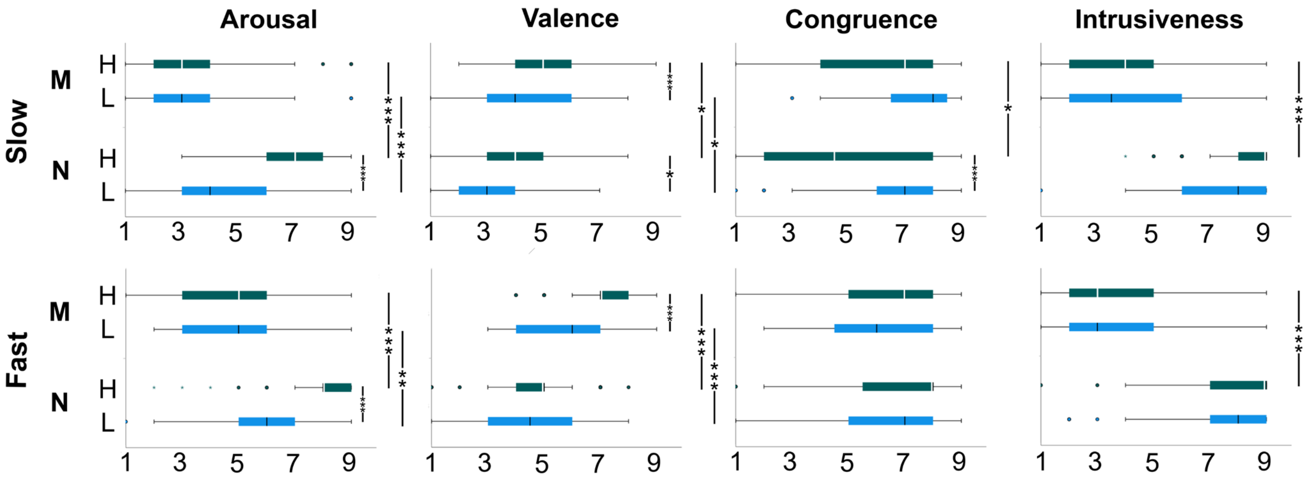

Figure 5: Boxplots visualizing the results of the survey for all factor levels and outcomes. The upper row represents slow walking, and the lower one represents fast walking. Within each plot, the ratings are clustered vertically based on sonic category (N = nonmusical, M = musical) and energy levels (H = high energy, dark blue box, L = low energy, light blue box). In each case, the boxes represent the interquartile range (IQR), and the notches within the boxes represent the median. The whiskers indicate variability outside the IQR. The small circles represent potential outliers (> 1.5 IQR but <= 3 IQR above (below) the upper (lower) quartile. The dots denote extreme values (> 3 IQR above (below) the upper (lower) quartile). Significant differences are indicated by the asterisks between levels. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

5. GENERAL DISCUSSION

In this study, we developed an interactive gait sonification prototype and devised four mapping schemes (two musical and two nonmusical, each with a high- and low-energy design). We carried out a web-based perceptual test to evaluate how the schemes differed in perceived arousal, valence, congruence with observed energetic qualities of gait, and intrusiveness.

We hypothesized that high-energy sounds would be rated higher (e.g. more active and more positive) than their low-energy counterparts in terms of both perceived arousal and valence **(H1)**, which was partially validated. With respect to arousal, **H1** held true for the nonmusical schemes (**N-H** v/s **N-L**) but not for the musical schemes (**M-H** v/s **M-L**). Conversely, **M-H** and **M-L** showed clear valence differences (**M-H** rated as happier), but this contrast was less pronounced between **N-H** and **N-L**. Hence, the **M-H** v/s **M-L** differences in envelope properties, dynamics, musical scale, and pitch range translated to differences in valence ratings (**M-H** rated as happier than **M-L**) but *not* arousal ratings, contrary to our expectations. Also, both **M-H** and **M-L** received lower arousal ratings than their nonmusical counterparts.

Based on [16, 15], we primarily attribute the arousal results to the spectral characteristics of the sounds. Looking at the upper half of Fig. 4, it is clear that **N-H** had both a higher spectral centroid and a greater concentration of high frequency energy (= brighter timbre) than **N-L**. Comparing **M-H** and **M-L** (lower half of Fig. 4), any spectral differences are far less apparent, which may have contributed to them being rated so similarly despite other signal properties being distinct.

We next hypothesized that sound clips generated from fast walking would receive higher arousal and more positive valence ratings than those from slow walking **(H2)**. **H2** was clearly validated irrespective of sonic category or energy level. As the step rate was greater for fast walking, this translated to sound events being generated at a faster rate/tempo. Our arousal and valence results line up well with past findings related to tempo and emotion in both music performance and walking [11, 12, 14, 15], where fast tempi have been associated with high arousal as well as happiness, whereas slow tempi have been linked with sadness. These findings suggest that all our schemes were successful at communicating the temporal characteristics of the recorded walking data through sound.

Our next hypothesis **(H3)** was that the energetic properties of the **H** schemes would be rated as more congruent with fast walking than slow walking, and vice versa. **H3** was partially validated for slow walking, with a significant difference between **N-H** v/s **N-L**, and a similar tendency (not significant) for the musical schemes (see Fig. 5). This result could be because **N-H** and **N-L** were farther apart than **M-H** and **M-L**, both in terms of their generation and their perceptual characteristics. For fast walking, there were no congruence rating differences regardless of sonic category or energy level. This lack of differences can have several explanations. Due to differences in inter-modality temporal resolution [26], it may have been harder for participants to form accurate mental associations between the visual and auditory stimuli when the observed limbs (and resulting sounds) moved (evolved) at a faster rate. However, it is also known that users interact differently with movement sonification systems depending on whether they are movement performers or observers [27] (possibly due to respective presence or absence of interactions between the visual, auditory, haptic, and proprioceptive channels) so the results may be different when users are movement performers.

It is curious that the differences in arousal and valence ratings between **H** and **L** schemes in many cases largely did not equate to equivalent differences in their perceived congruence with fast and slow walking. Because the walking speed itself (and result-

DOI: 10.5281/zenodo.7552228

ing sound event rate/tempo) had such a strong effect on perceived arousal and valence regardless of sound energy (see Table 1), any differences between **H** and **L** schemes had very little impact on congruence ratings. In other words, the **L** schemes sounded happy and aroused enough when applied to fast walking that they were not perceived to be incongruent with the movement. Also, the sounds were temporally synchronized with the visual stimulus in our experiment, which may have led participants to rate their properties as congruent [19].

Our final hypothesis **(H4)** was that the musical schemes would receive lower intrusiveness ratings than the nonmusical, and this was clearly borne out by the results. The flute is known to be the preferred instrument for peaceful, neutral, and sad music performance [15] and the mellow timbre of the flute model might have been preferred over the relatively bright sounding nonmusical schemes. The use of random musical scale notes during each swing phase may have added a degree of variety and unpredictability to the listening experience, which is known to be an important element of user engagement when using movement sonification [22]. The results are in line with those of [21], where synthesized musical auditory guidance was rated as being more pleasant and preferable for a longer duration of use than nonmusical guidance based on the same auditory perceptual properties. It is important to note that participants still did rate the musical sounds as *somewhat* intrusive (median 3-4 out of 9), indicating that the musical schemes have room for improvement in terms of sound quality, expressive properties, and musical content (e.g. a robust generative composition engine rather than random notes every time).

The present study also has some limitations. Constant-speed treadmill recordings of a single walker were used both for tuning of the mapping schemes and generating the experimental stimuli. This was done to ensure temporal consistency in the data, but data collected from overground walking may be different, particularly in terms of velocity variability. The sound generation algorithms used were chosen in accordance with the FAUST libraries, but for the redesign process we will consider using third party sound libraries for greater sonic versatility. The experiment was carried out online so as to reach a diverse audience, but allowed us less experimental control over participants' listening conditions and sound hardware, which may have led to greater variance in participant ratings. Last but not least, the participants listened to the sonifications of observed movements, and our findings (particularly movement-sound congruence) may not directly apply to participants listening to their own movements [27].

With respect to induced gait characteristics, the results of [11] showed that gait parameters differ based on both the arousal (aggressive v/s tender) and valence (happy v/s sad) of the walker's emotional intent. As our eventual goal is to assess induced differences in gait kinematics between **H** and **L** schemes, it is necessary that these schemes differ in *both* perceived arousal and valence irrespective of sonic category. In general, both the musical and nonmusical schemes should be equidistant within their categories as well as in their morphology. We estimate that a redesign is necessary, primarily focusing on a) greater morphological uniformity between the musical and nonmusical categories, and b) introducing more pronounced spectral differences between **M-H** and **M-L**, especially brightness. The redesign will be carried out such that **H** and **L** sounds uniformly maintain their intended arousal and valence properties regardless of walking speed. For instance, the noise-based whooshing scheme **N-H** received considerably high perceived arousal ratings even for slow walking, which probably led to it receiving significantly lower congruence ratings with slow walking than the creaking scheme **N-L** (see Fig. 5). After the redesign, we will perform a real-life walking experiment where the mapping schemes will be applied interactively, and differences in their induced gait characteristics will be assessed. Lastly, it is known that sound intensity is also an important factor in emotion perception and recognition in both music and walking [12]. The audio clips we used were normalized for roughly equal loudness, but for subsequent experiments it may be wiser to introduce loudness differences between **H** and **L** schemes to exaggerate the differences in their perceived arousal and valence.

6. CONCLUSION

In this study we found the perceived arousal, valence, intrusiveness and movement-sound congruence to vary depending on the energy level and musical nature of the sonification scheme, as well as walking speed. The potential of music as sonification was reinforced by virtue of our musical schemes being rated as significantly less intrusive than the nonmusical schemes. In order to extend the gait-altering potential uncovered by past footstep sonification research to swing phase sonification schemes, our designs must be reworked so as to achieve more distinct arousal and valence characteristics with more uniform perceptual differences between categories prior to real-life testing with walkers. Overall, we believe that the present work is a firm step in the direction of developing swing phase sonification schemes that can induce meaningful motor performance change in gait rehabilitation and exercise settings.

7. ACKNOWLEDGMENTS

We would like to thank the participants who took the time to complete the survey. This work is partially funded by by NordForsk's Nordic University Hub, Nordic Sound and Music Computing Network NordicSMC, project number 86892.

Author contributions: PRK was main responsible for technical development, experiment design, data analysis, and writing. PRK and RB devised the original research proposal. SD, EGS, and RB provided suggestions related to past literature, experimental design, sound design, data analysis procedures, and contributed to the writing. All authors approved of the final manuscript.

8. REFERENCES

[1] Ana Tajadura-Jiménez, Nadia Bianchi-Berthouze, Enrico Furfaro, and Frederic Bevilacqua, "Sonification of surface tapping changes behavior, surface perception, and emotion," *IEEE MultiMedia*, vol. 22, no. 1, pp. 48–57, 2015.

[2] Stephanie Studenski, Subashan Perera, Kushang Patel, Caterina Rosano, Kimberly Faulkner, Marco Inzitari, Jennifer Brach, Julie Chandler, Peggy Cawthon, Elizabeth Barrett Connor, et al., "Gait speed and survival in older adults," *Jama*, vol. 305, no. 1, pp. 50–58, 2011.

[3] Martin Alfuth and Dieter Rosenbaum, "Effects of changes in plantar sensory feedback on human gait characteristics: a systematic review," *Footwear Science*, vol. 4, no. 1, pp. 1–22, 2012.

[4] Yon Visell, Federico Fontana, Bruno L Giordano, Rolf Nordahl, Stefania Serafin, and Roberto Bresin, "Sound design

and perception in walking interactions," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 947–959, 2009.

[5] Tania Lam and Keir G Pearson, "The role of proprioceptive feedback in the regulation and adaptation of locomotor activity," *Sensorimotor Control of Movement and Posture*, pp. 343–355, 2002.

[6] Matthew WM Rodger, William R Young, and Cathy M Craig, "Synthesis of walking sounds for alleviating gait disturbances in parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 543–548, 2013.

[7] Julia Reh, Tong-Hun Hwang, Gerd Schmitz, and Alfred O Effenberg, "Dual mode gait sonification for rehabilitation after unilateral hip arthroplasty," *Brain Sciences*, vol. 9, no. 3, pp. 66, 2019.

[8] Luca Turchet, Stefania Serafin, and Paola Cesari, "Walking pace affected by interactive sounds simulating stepping on different terrains," *ACM Transactions on Applied Perception (TAP)*, vol. 10, no. 4, pp. 1–14, 2013.

[9] Ana Tajadura-Jiménez, Maria Basia, Ophelia Deroy, Merle Fairhurst, Nicolai Marquardt, and Nadia Bianchi-Berthouze, "As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 2943–2952.

[10] Fritz Menzer, Anna Brooks, Pär Halje, Christof Faller, Martin Vetterli, and Olaf Blanke, "Feeling in control of your footsteps: conscious gait monitoring and the auditory consequences of footsteps," *Cognitive neuroscience*, vol. 1, no. 3, pp. 184–192, 2010.

[11] Luca Turchet and Roberto Bresin, "Effects of interactive sonification on emotionally expressive walking styles," *IEEE Transactions on affective computing*, vol. 6, no. 2, pp. 152–164, 2015.

[12] Bruno L Giordano, Hauke Egermann, and Roberto Bresin, "The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity," *PLoS One*, vol. 9, no. 12, pp. e115587, 2014.

[13] P. N. Juslin, "Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 6, pp. 1797–1813, 2000.

[14] Patrik N Juslin and Petri Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?," *Psychological bulletin*, vol. 129, no. 5, pp. 770, 2003.

[15] Roberto Bresin and Anders Friberg, "Emotion rendering in music: range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.

[16] Faranak Abri, Luis Felipe Gutiérrez, Prerit Datta, David RW Sears, Akbar Siami Namin, and Keith S Jones, "A comparative analysis of modeling and predicting perceived and induced emotions in sonification," *Electronics*, vol. 10, no. 20, pp. 2519, 2021.

[17] Jeska Buhmann, Frank Desmet, Bart Moens, Edith Van Dyck, and Marc Leman, "Spontaneous velocity effect of musical expression on self-paced walking," *PloS one*, vol. 11, no. 5, pp. e0154414, 2016.

[18] Zachary Thomas Wallmark, *Appraising Timbre: Embodiment and Affect at the Threshold of Music and Noise*, Ph.D. thesis, UCLA, 2014.

[19] Pieter-Jan Maes, Marc Leman, Caroline Palmer, and Marcelo Wanderley, "Action-based Effects on Music Perception," *Frontiers in Psychology*, vol. 4, pp. 1008, 2014.

[20] Ashutosh Kharb, Vipin Saini, YK Jain, and Surender Dhiman, "A review of gait cycle and its parameters," *IJCEM International Journal of Computational Engineering & Management*, vol. 13, pp. 78–83, 2011.

[21] Prithvi Ravi Kantan, "Comparing sonification strategies applied to musical and non-musical signals for auditory guidance purposes," in *Proceedings of the 19th Sound and Music Computing Conference (SMC 2022)*, 2022.

[22] Pieter-Jan Maes, Jeska Buhmann, and Marc Leman, "3MO: A Model for Music-Based Biofeedback," *Frontiers in Neuroscience*, vol. 1, 12 2016.

[23] Prithvi Ravi Kantan, Erika G Spaich, and Sofia Dahl, "A Metaphor-Based Technical Framework for Musical Sonification in Movement Rehabilitation," in *The 26th International Conference on Auditory Display (ICAD 2021)*, 2021.

[24] Emma Frid, Jonas Moll, Roberto Bresin, and Eva-Lotta Sallnäs Pysander, "Haptic feedback combined with movement sonification using a friction sound improves task performance in a virtual throwing task," *Journal on Multimodal User Interfaces*, vol. 13, no. 4, pp. 279–290, 2019.

[25] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart, "The Musicality of Non-musicians: An Index for Assessing Musical Sophistication in the General Population," *PloS One*, vol. 9, no. 2, 2014.

[26] John G. Neuhoff, "Is Sonification Doomed to Fail?," in *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, Newcastle upon Tyne, June 2019, pp. 327–330, Department of Computer and Information Sciences, Northumbria University.

[27] Roberto Bresin, Maurizio Mancini, Ludvig Elblaus, and Emma Frid, "Sonification of the Self vs. Sonification of the Other: Differences in the Sonification of Performed vs. Observed Simple Hand Movements," *International Journal of Human-Computer Studies*, vol. 144, pp. 102500, 2020.

 DOI: 10.5281/zenodo.7552228