Taking a Walk on the Wild Side: Effects of Walking in Synchrony with Pitch-Altered Footstep Sounds on Body Perception in Outside the Lab Contexts SUPPLEMENTARY MATERIAL

1 Supplementary Methods

1.1 Recording Hardware Configuration

Recordings were conducted in an anechoic environment at Carlos III University of Madrid. To record the footstep sounds we used two sets of wired binaural microphones (SOUNDMAN OKM-II), with a frequency response of 20 Hz–20 kHz, a sensitivity of approximately 7 mV/Pa, and an impedance of 2.2 kΩ. Designed for in-ear placement, they support a maximum sound pressure level (SPL) of 120 dB and operate on plug-in power (1.5 V-10 V) via a 3.5 mm TRS stereo mini-jack. We also used a handheld portable recorder (ZOOM h4n), in four-channel configuration. One set of microphones was positioned at the tip of the shoes and the other one in the walker's ears. The files were recorded with a 16-bit resolution and 44100 Hz sampling rate. A mobile application with a flashing-screen metronome functionality was used to provide a synchronized visual tempo cue (80, 100, 120 bpm) during the recordings.

1.2 Materials and Recording Setup

Two different ground materials were used: wood (MDF) planks, positioned on the floor of the anechoic environment to form a runway measuring 480 cm in length and 80 cm in width, and marble tiles (40 x 40 x 4 cm), which were placed on top of the wooden runway. Further, two types of shoes were used: leather dress shoes (EU sizes 42 and 46) and sandals with hard rubber soles (EU size 42). To simulate increased body weight and create varied weight recording conditions, participants wore a fitness weight set that added 5 kg of extra mass.

The choice to include sandals with hard-rubber soles as footwear and wood (MDF) as the walking surface was made to replicate the materials setup used in [3]. However, in line with the main aim of the experimental study – namely, to explore and test the "Footsteps Illusion" in a more realistic, everyday scenario – we selected an outdoor area on the university campus for conducting the experiment. This area was moderately large (approximately 70 x 90 m), mostly flat, typically not too crowded, and featured a variety of surface materials, including stone and concrete tiles. For this reason, we decided to incorporate marble tiles into the recording session to better reflect this diversity. As for the footwear, we also included leather dress shoes in the recordings, to include other types of shoes commonly worn in daily life.

Three people, weighing approximately 50, 60 and 70 kg, participated in the sound recordings. Each of them performed eight recording sessions: four while wearing the 5 kg weights (sandals on marble, sandals on wood, shoes on marble, and shoes on wood), and the same four without the additional weight. This approach allowed us to cover the following weight ranges: 50-55 kg, 55-60 kg, 60-65 kg, 65-70 kg, 70-75 kg and 75-80 kg.

To prevent introducing any additional sound sources into the recordings, participants synchronized their footsteps using a visual metronome on a smartphone, which emitted a screen flash at each beat. The recording consisted of five 1-minute phases: walking in place at 80 bpm, walking normally at 80 bpm, walking in place at 100 bpm, walking

Author's Contact Information:, 4m4r000@gmail.com.

normally at 100 bpm, and walking in place at 120 bpm. We only recorded walking in place at 120 bpm due to the limited path length, as walking at this tempo along the short distance would have resulted in unnaturally short steps.

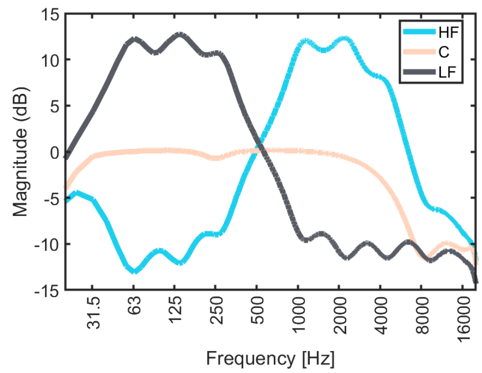

Fig. S1. Frequency response of the High Frequency (HF), Low Frequency (LF) and Control (C) filters.

1.3 Soundtrack Selection and Processing

After the recording sessions, we went through all the audio files and selected the candidate ones for the study. Specifically, we focused on the walking pace and decided to exclude 80 bpm, 120 bpm and walking-in-place recordings. This decision was made considering that, in the study, we wanted participants to walk at a "natural" speed, which we defined as a pace that people typically use in everyday walking scenarios. We felt that 80 bpm and 120 bpm might not represent this natural pace, and walking-in-place recordings were excluded to avoid any artificial constraints. Although studies involving similar pre-recorded sounds like [5] have used a 120 bpm walking pace, they only involved the walking-in-place activity, which is different from the walking scenario we aimed to replicate. Once we had preliminarily selected the recordings, we decided to obtain, for each of the six weight-groups, a 5-minute footsteps audio track, that we could later filter to produce the three experimental sound conditions. The strategy we adopted was to extract several "footsteps pairs" and then combine them in random order to obtain 5-minute tracks. For each of the selected recordings, we analyzed the footsteps peaks to extract the time between two adjacent footsteps (T) (e.g., left-right). We then manually cut the footsteps pairs, leaving $\frac{T}{2}$ before the first peak and after the second peak. To automate the process of generating the audio tracks, we developed a MATLAB script that randomly selected and concatenated the footstep pairs. To ensure smooth transitions between the segments, we applied a 10 ms fade-out/fade-in effect at each junction. This approach allowed us to create seamless and randomized 5-minute audio tracks. Furthermore, since the study setup employed passive noise-canceling headphones, we decided to incorporate background noise into the footsteps tracks to make the listening experience more realistic and less frustrating. To do so, we recorded the background noise at the study location and added it to the soundtrack. One of the most important parts for the design of the soundtracks for our experiment was to create a realistic auditory experience in the tracks. To this end, one of the key aspects is the plausibility of the background noise for the specific location of the test. For this reason, instead of using publicly available background noise recordings from online databases, we decided to record the background noise at the specific location where the experiment was to be conducted, being cautious to avoid the inclusion of salient noise events (e.g., a clear sound of a motorbike or a bus passing by). The integration of the background noise to the footstep sounds in the track heard by the participants is not trivial, as it should maintain a signal (i.e., footsteps) to noise (i.e., background noise) ratio, which allows for a realistic contrast. Thus, the algorithm for the creation of the soundtracks includes an SNR parameter as input, to control this behavior. However, we did not pursue a specific SNR value, instead we used the piloting session to select the one that rendered the most realistic sensation, by evaluating several files created with different SNRs (i.e., 0, 5, 10 and 15 dB). After the piloting session, the selected file had an SNR = 0 dB, as the manuscript explains. To replicate this work, with the same objective of generating a realistic audio experience for the designated setup, we encourage researchers to focus on following the same procedure we presented to obtain the audio files, rather than on reaching a specific SNR or spectrogram of the background noise. Only for completeness, two spectrograms (footstep sounds alone and background noise), for a 30-second fragment, have been included in Figure S2. As it can be seen, both signals have similar energy according to a SNR of 0 dB. However, the footsteps signal presents a clear pseudo-periodic pattern, consistent with a walking pace.

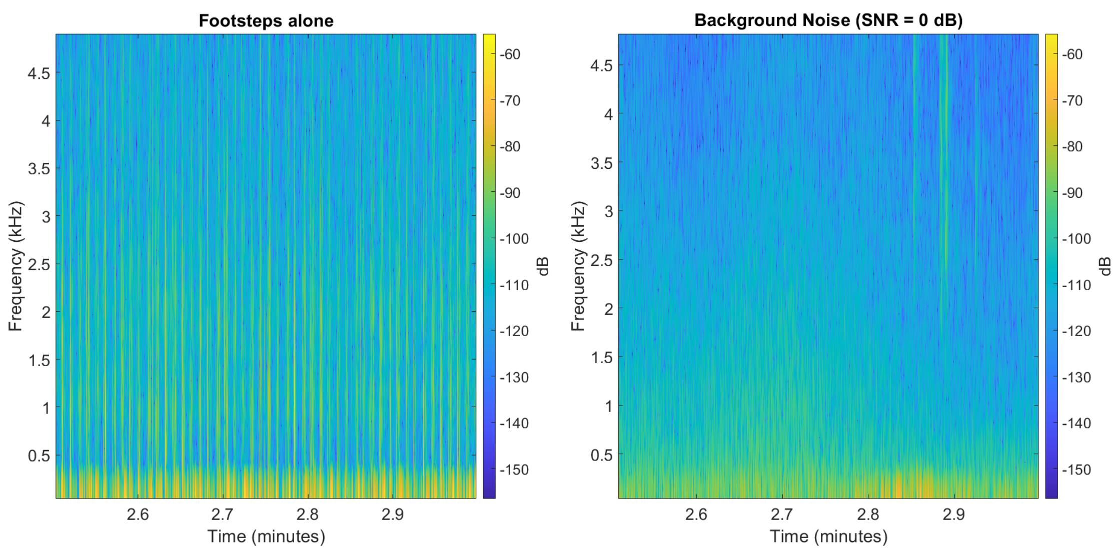

Fig. S2. Spectrograms of a 30-second fragment of the 70-75 kg footstep soundtrack, and of background noise.

Since the recordings were composted of four channels (two for the microphones placed in the ears and two for those placed on the shoes), the footstep pairs extraction involved all four channels, allowing us to generate audio tracks for both microphones locations. We ultimately decided to perform the piloting using only the foot-located microphones for the following reasons. Firstly, in the footsteps recordings with the microphones in the ears, the footstep sound level was lower than with the foot-located microphones, and this, after amplifying the signal, generated a perceivable click at the footstep pairs joints. Secondly, considering the the influence of height differences between the individuals who produced the recorded sounds and the participants in the study. Finally, to remain consistent with previous studies [3, 6] that use the feet as microphone locations.

The same filters developed for the SoniWeight Shoes device [3], as described in [4], were used to equalize the pre-recorded footstep sounds, obtaining: an HF version and a LF version, simulating the acoustic characteristics of a lighter and a heavier body, respectively. In HF, higher frequency bands (1-4 kHz) were amplified by 12 dB, while lower frequencies (83-250 Hz) were attenuated by 12 dB. In contrast, LF inverted this pattern. The filters followed a cascaded bi-quadratic filtering approach, to ensure high efficiency and low latencies, replicating the ones used in [3], used for real-time equalization of instantaneous gathered footstep sounds. While for this experiment real-time equalization was not required (as the filtering of the soundtracks is conducted prior to the experiment), we decided to apply the same filtering approach to ensure equal filtering conditions. Figure S1 shows the frequency response of the two filters, along with the Control condition (referred to as C). Considering that equalizing the soundtracks with the HF and LF filters would result in different frequency ranges either amplified or attenuated, and taking also into consideration that the ear is more sensitive to higher frequencies than to low frequencies, the same soundtrack in the two conditions (HF and LF) would result, in addition to the spectral shaping sought to be studied, in a difference in absolute loudness, which could be a bias factor, if not corrected. Keeping this in mind, after filtering the soundtracks with the HF and the LF conditions, the A-weighted RMS (to take into account the frequency sensitivity of the ear) value of the three conditions (HF, LF and C) was calculated and broadband correction factors were applied to ensure equal A-weighted RMS among the three spectral settings. Specifically, the A-weighting curve is used (and no other, i.e., the, B, C or D) considering the moderate sound pressure level of standard footstep sounds. Once this process was finished, the three spectral versions of the same soundtracks sounded different in terms of their spectral shaping but had the same absolute loudness.

1.4 Synchronization Analysis

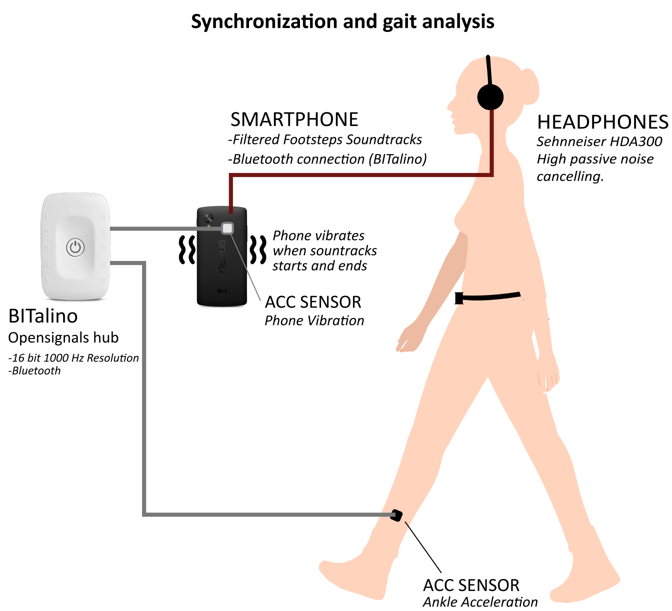

Fig. S3. Synchronization analysis and experiment setup diagram

The smartphone was programmed to emit a vibration pattern of 500 ms, oscillating at a frequency of roughly 166 Hz. For the analysis, the onset of the vibration within the 1000 Hz accelerometer signal was identified through visual inspection of the characteristic vibration pattern. Given the sampling rate and the clear periodic structure of the signal (one peak every 6 ms ca., see Figure S4), the onset could be reliably extracted with an estimated temporal accuracy of approximately 1 ms. At this point, having both the phone- and the participant's leg- accelerometer data synchronized in the same file, we could use the vibration onset time index as the beginning the soundtrack reproduction and analyze the synchronization between the soundtrack and the participant's walking activity.

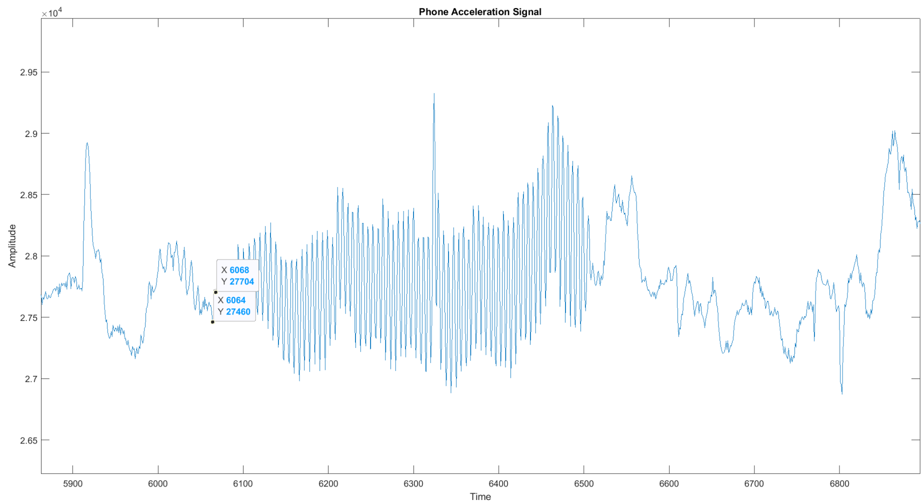

Fig. S4. Phone acceleration Signal: vibration pattern extraction.

1.5 Questionnaire on Body feelings and Emotional Experience

Participants were asked to answer the question "How did you feel during the experience?" from 1 to 9, selecting accordingly a figure from each of the 3 different scales of the Self-Assessment Manikin[1]–a non-verbal, graphic representation of the three fundamental emotional dimensions. The "Valence" or "Pleasure" scale ranges from 1 (Unhappy, Negative), to 9 (Happy, Positive); the "Arousal" scale ranges from 1 (Unaroused, Calm), to 9 (Aroused, Excited); and the Dominance scale ranges from 1 (Submissive, Awed), to 9 (Dominant, Important). Participants were then asked to select, on 1–7 bipolar scales, the number that best represented where their experience fell between each pair of opposite descriptors: Slow/Quick; Light/Heavy; Weak/Strong; Crouched, Stooped/Elongated, Extended; Very Feminine/Very Masculine. Finally, participants were asked to indicate the extent to which they agreed with the following statements, using a scale from 1 (strongly disagree) to 7 (strongly agree): "During the experience it seemed like the sounds I heard were produced by my own footsteps/ body"; "During the experience it seemed the feeling of my body was less vivid than normal"; "During the experience the feelings about my body were surprising and unexpected"; "During the experience it seemed like I could really tell where my feet were".

1.6 Interview Questionnaire

- **General observations:** How was the experience of walking freely with the sound in general? Did you notice any differences with (Experiment 1)? If so, what were they? Why do you think they were different?

- **Sound 1**: Describe how you felt with the first sound. How did you perceive the sound of your footsteps when walking on different surfaces? How did you perceive your body when walking on different surfaces? Did you feel anything different from your usual way of walking, in particular? For example, these could be sensations, emotions, thoughts, etc. (where, what kind...) In this first experience, you may have experienced some changes in your body, or you may not have. Did you feel any? (If so, what and why?) You may have felt that the sound impacted your movements, although you may not have noticed any impact. Do you think your movement was impacted by those sensations/changes? How? (Observations about your path, such as "I see you spent more time in this area than others. Tell me about this.") Were there things in the environment that influenced your experience - e.g. people, weather.
- **Sound 2:** (same questions as Sound 1)
- **Comparison of the two experiences:** There are similarities and differences between the two patio maps. Why do you think this is? Why do you think there are similarities if you heard different sounds? Do you think the different sounds had a similar or different impact on your experience? Why? In what way?
- **Reflections on the experiment design:** Including the first part of the experiment where we told you where to walk, what do you think about the experiment itself, for example, what we asked you to do? What did you like, and why? If you could change something about the experiment, what would it be? What was it like to hear the footsteps in the headphones? Did you feel like they were your footsteps? Was it easy/difficult to synchronize with them? Was it different to coordinate with the footsteps when you had a predetermined path and your own path?

2 Supplementary Results

2.1 Experiment 1

2.1.1 Questionnaire on Body Feelings and Emotional Experience. In Table S1 we report the Confidence Interval Values for the Questionnaire on Body Feelings and Emotional Experience results.

2.1.2 Body Visualization. During the analysis, we examined the individual data distribution and observed that one participant's responses diverged substantially from the rest (see Figure 4) Although we did not have a strong a priori reason to exclude this participant, we note that re-running the ANOVA without this case—treated as a potential outlier—yielded a significant main effect of sound condition excluding the outlier: F=5.131, p=0.01, $\eta_p^2 = 0.20$, medium to large effect size). (F=3.73, p=0.032, $\eta_p^2 = 0.15$, medium effect size [2]). Post hoc t-tests in this restricted sample further showed lower represented body weight in HF than in LF (t=2.58, p=0.04).

2.1.3 Gait Biomechanics. ANOVA for the stance durations didn't reveal a significant differences across sound conditions (F=0.02, p=0.979, $\eta_p^2 = 0.01$), Means (± SD) were HF: 1.283 (± 0.06), C: 1.299 (± 0.05) and LF: 1.297 (± 0.06).

Table S1. Median and Confidence Interval (CI) values for the Questionnaire on Body Feelings and Emotional Experience Items scores.

	Median			Lower CI			Higher CI		
Measure	C	HF	LF	C	HF	LF	C	HF	LF
Valence	6	7	7	6	6	6	7	8	8
Arousal	4	3	3	2	2	2	5	4	5
Dominance	5	6	5	4	5	4	6	8	7
Quickness	4	4	4	3	3	2	5	5	5
Weight	4	2	4	3	2	3	4	4	5
Strength	4	5	5	4	3	4	5	5	5
Straightness	5	5	4	4	4	3	6	6	6
Masculinity	4	4	5	4	4	4	5	5	5
Agency	5	6	5	3	5	3	6	6	6
Vividness	4	3	4	3	3	3	4	5	5
Surprise	3	4	2	2	2	2	4	5	4
Proprioception	5	5	5	3	2	3	5	6	5

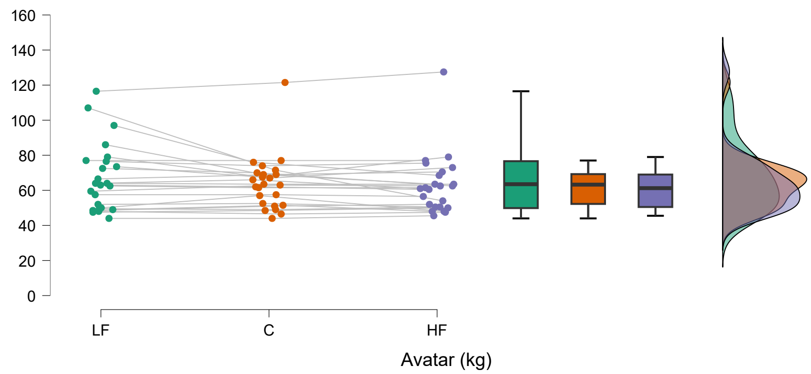

Fig. S5. Experiment 1 Results for Body Visualized Weight

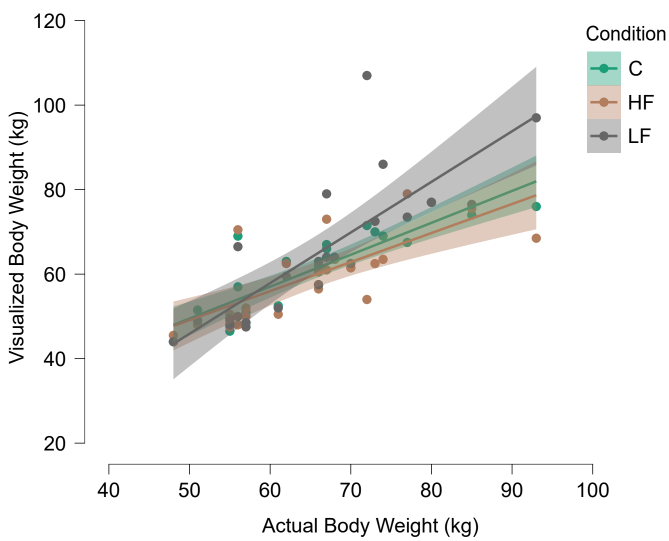

Fig. S6. Experiment 1 Results for Body Visualized Weight Correlation with Participants' Actual Body Weight

References

[1] Margaret Bradley and Peter Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49–59. doi:10.1016/0005-7916(94)90063-9

[2] Jacob Cohen. 1988. *Statistical Power Analysis for the behavioral sciences.* L. Erlbaum Associates, Hillsdale, N.J.

[3] Amar D'Adamo, Marte Roel, Laia Turmo-Vidal, Dehshibi Mohammad M., Daniel De La Prida, Joaquín R. Diaz-Durán, Luis Antonio Azpicueta-Ruiz, Aleksander Väljamäe, and Ana Tajadura-Jiménez. 2024. SoniWeight Shoes: Investigating Effects and Personalization of a Wearable Sound Device for Altering Body Perception and Behavior. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24).* Association for Computing Machinery, New York, NY, USA. doi:10.1145/3613904.3642651

[4] Daniel de la Prida Caballero, Joaquín Roberto Díaz Durán, Luis Antonio Azpicueta Ruiz, and Ana Tajadura Jiménez. 2022. As light as your footsteps: design and evaluation of a portable device for changing body perception through a sound illusion. (11 2022). https://e-archivo.uc3m.es/handle/10016/37191

[5] Merle T. Fairhurst, Ana Tajadura-Jiménez, Peter E. Keller, and Ophelia Deroy. 2023. You, me, and us: Maintaining self-other distinction enhances coordination, agency, and affect. *iScience* 26, 12 (2023), 108253. doi:10.1016/j.isci.2023.108253

[6] Ana Tajadura-Jiménez, Maria Basia, Ophelia Deroy, Merle Fairhurst, Nicolai Marquardt, and Nadia Bianchi-Berthouze. 2015. 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 (CHI '15).* ACM, New York, NY, USA, 2943–2952. doi:10.1145/2702123.2702374