Personal and Ubiquitous Computing (2024) 28:779–799
https://doi.org/10.1007/s00779-024-01819-3

ORIGINAL PAPER

Audience perceptions of Foley footsteps and 3D realism designed to convey walker characteristics

Stuart Cunningham[1] · **Iain McGregor**[2]

Received: 18 March 2023 / Accepted: 28 May 2024 / Published online: 11 June 2024

Abstract

Foley artistry is an essential part of the audio post-production process for film, television, games, and animation. By extension, it is as crucial in emergent media such as virtual, mixed, and augmented reality. Footsteps are a core activity that a Foley artist must undertake and convey information about the characters and environment presented on-screen. This study sought to identify if characteristics of age, gender, weight, health, and confidence could be conveyed, using sounds created by a professional Foley artist, in three different 3D humanoid models, following a single walk cycle. An experiment was conducted with human participants (n=100) and found that Foley manipulations could convey all the intended characteristics with varying degrees of contextual success. It was shown that the abstract 3D models were capable of communicating characteristics of age, gender, and weight. A discussion of the literature and inspection of related audio features with the Foley clips suggest signal parameters of frequency, envelope, and novelty may be a subset of markers of those perceived characteristics. The findings are relevant to researchers and practitioners in linear and interactive media and demonstrate mechanisms by which Foley can contribute useful information and concepts about on-screen characters.

Keywords Sound design · Foley · Audio features · User perceptions · Animation · Walk cycles

1 Introduction

This experiment is intended to explore the influence that Foley can have on viewers' perception of animated, neutral avatars. Foley artists go to great lengths to perform actions that correspond with visual cues such as matching age, gender, and mass, by utilising manual props to perform actions in sync with pictures that are recorded and added to a soundtrack.

Characterisation is a key aspect of a Foley artist's work as the intention is for the sounds to be *felt* rather than *heard*. It should be as if all the sounds were captured by accident during production and are impossible to separate from what is visually on the screen, adding verisimilitude to such an extent that viewers start to see elements that were not represented visually. Listening in general provides the ability to understand actions and objects well beyond the visible, and Foley facilitates this process within animation to provide a more complete understanding of a character and their motivations.

Gait refers to the nature of the way in which a person moves, most typically with respect to locomotive activities, such as walking or running. The analysis of gait is a distinct field of research in its own right [1]. Work in this area spans the health and rehabilitation domains [2, 3] as well as examining how the gait of an individual might change depending on factors such as age, weight, and gender [4–7]. Perceptions of observers may also be influenced by their gait, such as their ability to be identified [8]. From a media and creative industries perspective, the study of gait has been used to investigate its potential to communicate information about characters or avatars to audiences. For example, Hicheur and colleagues [9] examined how real and artificially generated visual gait cycles could convey four of five selected human emotions to viewers. Earlier work considered producing artificial animated gaits for avatars in virtual environments discussed

Iain McGregor contributed equally to this work.

✉ Stuart Cunningham
s.cunningham@chester.ac.uk

Iain McGregor
I.McGregor@napier.ac.uk

[1] School of Computer and Engineering Sciences, University of Chester, Parkgate Road, Chester CH1 4BJ, UK

[2] School of Computing, Edinburgh Napier University, Colinton Road, Edinburgh EH10 5DT, UK

how the manipulation of motion control parameters could communicate characteristics such as gender [10] and emotion [11]. Later research has considered how gait parameters might influence observer perceptions of an animated character's personality [12]. In many ways, the study of gait, both captured and synthesised, and observer perceptions of real and virtual walkers have mirrored one another in their range of application and utility.

Other researchers have examined how the *walking style* of virtual avatars may convey biological and personality features, such as gender and confidence [13]. In this study, experiments were conducted to establish whether concepts such as a character's age, gender, weight, health, and confidence can be communicated through *Foley*.

Much has been written about the role of Foley and sound design in communicating character information to audiences. However, to the best of our knowledge, no other studies have been undertaken to formally evaluate if, and how, Foley footstep sounds can communicate attributes of age, gender, weight, health, and confidence whilst controlling for the visual appearance of a computer-generated, animated walker that appears in a combined stimulus.

The outcomes for this work are likely to have utility not just to researchers in the sound design, audio analysis, and media domains, but to practitioners in a variety of linear and interactive media scenarios. In essence, this work aims to validate and understand the techniques that professional Foley designers employ, underlining the importance of sound in communicating anthropometric and non-anthropometric characteristics of on-screen avatars. It is hoped that this work will enhance and extend what is known in the literature about audio characteristics and their link to audience perceptions.

This article is an extended version of our previous work [14] and is concerned with exploring, and attempting to answer, the following research questions:

1. To what extent can the characteristics of age, gender, weight, health, and confidence be successfully communicated to an audience through Foley designed footsteps in an animated, neutral avatar, with a variety of visual designs?
2. What audio features contribute to participant perceptions of Foley footsteps, particularly with respect to the characteristics of the walker they are intended to represent?

2 Background and related work

2.1 Foley artistry

Foley artists make use of a variety of tools and techniques with the goal of being able to create believable, synchronised sound that can help reinforce the narrative and reality established within a production [15–17]. Three core sound elements of footsteps, prop handling, and cloth are typically performed [18–20].

Foley, being diegetic in nature, can be considered to play a part in allowing an audience to comprehend the actions and characters in a scene and therefore contribute to a narrative, although the source or object used to create the sound may be different to that depicted in a process of *"sound-acting"* [21]. Foley work is a nuanced, performative activity, and expressiveness is considered an important part of a Foley artist's experience [22].

Wright [19] argues strongly for Foley being important in the process of *"...creating sound with 'character"'* and that there is a strong human element to the work of Foley artists. The interaction of on-screen characters with props in a scene is often tailored to convey information about the character, such as their size or weight, or more abstract concepts, such as their emotional state, in what some have termed *method Foley*.

Use of Foley can go beyond purely accompanying the movements and props seen in a scene to provide the audience with expression of a range of characteristics or properties that are possessed by actors and objects. This might include elements such as the *"...the crispness of stone, the sogginess of mud, [or] the plushness of a carpet..."* and may extend to characteristics more conceptual in nature [23].

Warren, Kim, and Husney [24] discuss how a listener's auditory perception can establish dynamic physical properties of an environmental event. The results of their experiments show that different modalities offer different event information and that the auditory information given by the sound of a ball bouncing does in fact allow an accurate perception of the ball's elasticity.

[25].

Effective use of Foley can contribute to increased feelings of presence and immersion within virtual worlds, such as Virtual Reality (VR) and, by extension, Augmented Reality (AR) environments. Anderson and Casey [26] argued that effective sound design can help to compensate for limitations in visual fidelity in these situations, whilst reinforcing a sense of character for items and objects in a scene. Particularly important in these three-dimensional situations are the ability of the sound designer, or computational playback engine, to perform manipulations of these sounds to support the spatial location and/or movement of sources.

2.2 Perception of Foley

The perception of Foley within audio-visual media hinges upon the principle of *synchresis*, which is the linkage, or causation, that an audience would attribute to coincident audio and visual stimuli [27]. In other words, if a visual action is observed at the same time as a sound is heard, the two stimuli

will be attributed to the same object(s) and their interaction, assuming that the sound produced is within some boundary of being plausible or believable.

Foley is considered primarily physical and expressive in nature [28]. Such successful Foley artistry, drawing upon the principle of synchresis, is illustrated in an analysis of the film *Fight Club*, where the sounds representing the characters punching one another "*...is used to establish an authentic connection between bodies...*" [29].

Ennis, McDonnell, and O'Sullivan [30] examined the sensitivity of humans to audio mismatches and visual de-synchronisation in virtual animated conversations. The results show that the visual de-synchronisation has a more profound impact that any audio mismatch. The relationship between the audio and visual stimuli in any production is paramount to believability, and when either of these factors have to be comprised (due to budget or time restraints, for instance), there are workarounds to exploit the human cross-modal interaction for the benefit of the production. Mastoropoulou and colleagues [31] studied the effect of sound on the perceived smoothness of animations. The results showed that the addition of sound effects made it harder for all participants to distinguish which clip had the slower frame rate.

On discussing the effect of Foley within visual media, Lewis [32] explains how sound cannot be experienced in a mono-sensorial way and that when we hear a source, we try to establish a mental image of it rather than trying to describe the characteristics of the source sound. Lewis goes on to discuss the concept of a *crystalline image* [33] that is created from the original sound source and the sound heard in context at that specific point in time.

To investigate auditory perception of everyday Foley, as well as the effect of visual context upon these perceptions, a study was held to determine whether listeners could accurately identify sounds within a visual scene. Listeners accept a sound and its accuracy/appropriateness as that portrayed by the video [34].

One study explored the possibility of listeners determining physical attributes from objects they cannot see from their sound alone. Participants were asked to guess the length of several different wooden rods that were individually dropped from the same height. The results showed that listeners could scale objects appropriately without any standard or *a priori* information for comparison [35]. When applied in the context of Foley, this further supports the notion that sound can be used to depict physical attributes, independent from a visual stimulus or potentially in the absence of a visual counterpart. This may be due to the perception of an auditory event being intrinsically linked with memory traces of previous stimuli [36].

In a work concerned with *inverse-Foley*, a technique whereby a sound is taken as an input and used to synthesise a plausible animation sequence so that it is optimally synchronised, key features of sound are the timing and amplitude of the contact events, which can be used to imply physical properties and behaviours of an object as it interacts with other surfaces. As such, this work provides evidence, from another perspective, that the characteristics of an on-screen object or actor, especially in an animation setting, can be directly related to the sound that it can produce [37].

Whilst not explicitly concerned with Foley, research has been carried out to investigate the effect of posture upon the recorded sound of subjects walking. Significant relationships were first discovered between anthropometric properties of walkers and a set of bio-mechanical properties, where height, weight, gender, and shoe size all played notable roles. The ability of human listeners to correctly identify characteristics from the sound recording was variable [38]. However, it is suggested that the posture of a person walking is indicative of a range of characteristics, such as those examined in this research, but perhaps most intuitively those of age and degree of health.

Grassi [39] discusses the possibility of accurately estimating the size of an object through the sound of its impact. Three experiments were held in which various balls were dropped from a height onto different diameter plates. Grassi's study indicated that the perceived sound of objects can be manipulated, especially through impact sounds and potentially through footsteps.

2.3 Footsteps and Foley

Footstep sounds are considered one of the most important elements of Foley artistry and in the production of audio-visual media [22]. As such, prior to the term Foley *artist* being commonplace, Foley *walker* was used [15, 19]. Foley artists regularly apply their craft to convey multiple character traits, specifically concerning the use of footstep sounds, according to Beauchamp, when discussing the use of sound in animation: *"When a Foley artist walks a character, we are often able to determine the size, age, gender, and emotional state of that character without ever seeing the image. This dimension of character development is at the core of why we record Foley"* [18]. In a work examining the acoustic features of audio recordings, it was found that the gender of a person walking could be determined by participants, for example [40]. Others in the field share this view and consider that footstep sounds can be used to enhance emotions, reinforce behaviour, and support the narrative of visual media [28].

Donaldson [23] provides multiple examples of how sound can be used to convey a range of characteristics, often with examples to footstep sounds and their presence in a variety of films where they are used to convey concepts such as harshness, confidence, isolation, vulnerability, and strength. These are utilised to provide a sonic dramatisation and may

utilise sound post-production processing techniques, such as the application of reverberation and their level in the overall mix.

Work has previously been conducted to examine audience perceptions of footstep sounds in audio-visual situations. For example, interviews with a professional sound-maker exemplified that Foley practice takes into consideration the features of the characters to be presented, such as indicating age when recording footstep sounds, in addition to other contextual markers, like culture and geography related to the story being portrayed [41].

A study was performed that asked participants to rate video clips accompanied by various audio, which included real sounds (recorded on-location but not at the point of video recording); Foley; and low-quality synthesised sounds. Ratings were received in terms of the participants' confidence in gauging, perceived realism, and perceived expressiveness. It was found that realism and expressiveness were generally much higher overall for Foley and real sounds, although there was some variation within participants between the different surfaces that had been presented. Interestingly, this study also showed that participants found it difficult to identify the action causing the sound and the surfaces or materials involved when the same stimuli were presented without any visual accompaniment, arguably further demonstrating the importance of synchresis [42].

Earlier research compared audience perceptions of Foley with "real" sounds, recorded from the source objects and interactions in each case. The analysis took place in audio only and audio-visual conditions. In the study, two of the sound actions selected were footsteps. One was the sound of walking up stairs and the second that of walking upon grass. Interestingly, the real version of walking on grass was significantly preferred by participants in the audio only test, and the walking up stairs real sound was preferred in the audio-visual test. However, all other sounds that were statistically significant in these tests were Foley versions, although these represented the minority of comparisons made. In terms of the footstep sounds, this may indicate that realism is vital [43].

The importance of footstep sounds as part of Foley work may be further underlined by their focus in research studies that sought to facilitate the process by using synthesis [41, 44]. In the field of VR, work has also been done to examine how movements by users in the virtual environment might be used to synthesise plausible footstep sounds in real-time. A particularly relevant feature was that the authors were keen to synthesise the sound of a range of different surfaces that might be walked upon [44].

Other work in the field of physical modelling synthesis has been concerned with the production of footstep sounds, with the intention to convey anthropomorphic features of the walker. These characteristics were the walker's gender, foot length, and body size [45]. Importantly, it was recognised that there was a lack of a framework to produce footstep sounds and recognition that a variety of walker characteristics, such as gender, were important in accurate representations. In the synthesis models, body size, which was deemed as being an indicator of gender, was conveyed by adjusting the amplitude and spectrum of the sound. This work included the production of a tool that would allow Foley artists to easily produce appropriate and engaging footstep sounds. An evaluation with human participants was deployed to assess the plausibility of the resulting sounds. Of relevance is the intention that this evaluation would ask participants to identify the anthropomorphic features that the sound design tool was intended to convey. Results indicated that three types of body size (small, medium, and large) were successfully communicated and that in all but the big to medium case, participants could differentiate between the three sizes, with the variation in size being shown to modulate gender perception.

In a subjective evaluation, taking place in a controlled environment, it was found that manipulations to recorded Foley footstep sounds would influence the perception of individuals in terms of plausibility. In the study, a total of nine participants watched clips from a bespoke film and reported upon their perceived plausibility of Foley footsteps, that were presented in a variety of conditions, which included manipulations of volume, panning, equalisation, and reverberation. The clips with these manipulations were compared to a reference clip, produced by the researchers, and deemed to be of an industry standard in terms of mix by a professional sound designer and artist. The outcomes from this work lend support to the hypothesis that sound manipulations, and by extension variations in the sounds themselves, will be able to convey information about the nature of an actor or object in a scene [46].

3 Method

3.1 Participants

One hundred participants took part in the study, all of whom were associated with Edinburgh Napier University's Merchiston campus: 28% were academics, 34% administration or support staff, and 38% students. Invitations were sent via email, as well as made in person, and no incentive to take part was offered. The median age of those who took part was 39, with the youngest being 19 and the oldest 69; the gender split was 52% male and 48% female. All participants considered themselves to have normal hearing for their age, as well as normal or corrected-to-normal eyesight. Only a single participant had never experienced any form of media with animated characters; 95% had watched movies with animated figures

at some point, and only 26% had experienced VR with animated figures.

3.2 Materials

Forty-eight animated walk cycles were presented on a 13-inch laptop screen (Apple MacBook Pro), connected to a pair of nearfield audio monitors (Genelec 8030a) calibrated to 63.8 dBA (RMS), with a peak of 85.4 dBA. The ambient sound pressure level of the Edinburgh Napier University Auralisation Suite (listening room) was 35.8 dBA. Loudspeakers were chosen over headphones in order to convey body movement more accurately through the improved visceral perception of low frequencies which can be impaired through ear canal only delivery [47]. All the animations were QuickTime videos contained within a PowerPoint presentation.

The 13-inch laptop screen was chosen due to its convenience. As users were controlling the PowerPoint session themselves, they were sitting with their eyes approximately 75 cm from the screen. All the audio signals were mono but played through a stereo pair of speakers to create a phantom centre; the small screen sized ensured that the optimal 60 degrees could be reproduced, relying on spatial magnetization to minimise the effect of a visual mismatch.

Participants had access to paper copies of a participant information sheet, an informed consent form, and a questionnaire, all of which had previously passed successfully through the Edinburgh Napier University ethical approval process. As such, the activity was carried out in accordance with Edinburgh Napier University's guidelines and regulations.

3.3 Design

The walk cycle animation was created from a motion capture sample supplied by Autodesk (the file walksit.ma, available from www.autodesk.com/maya-creativemarket-samples). The original motion capture data represents a humanoid walking and sitting. The walk cycle was created from an extract of the mocap walk animation. Some small adjustments to the overall gait, timing, and details of the walk (feet motion) were made. The heel-to-heel interval of the character was approximately 500 ms. The walk cycle cadence, therefore, was approximately 120 steps/minute, broadly in line with comfortable walking speeds found in empirical studies of gait [48], although higher than what may be considered a moderate intensity cadence of approximately 100 steps/min [49].

The walk cycle was applied to three different characters (see Fig. 1). The first (1) was a seamless skinned humanoid based on the skinned character supplied with the mocap data. The model was modified to look as androgynous as possible. The second (2) was a model emulating the look of a wooden mannequin. The last model (3) was an abstract biped made of simple geometric shapes (mostly elongated square pyramids representing a simplified skeleton). The three animations were rendered from the same camera angle. Clips were 720p and lasted 15 s.

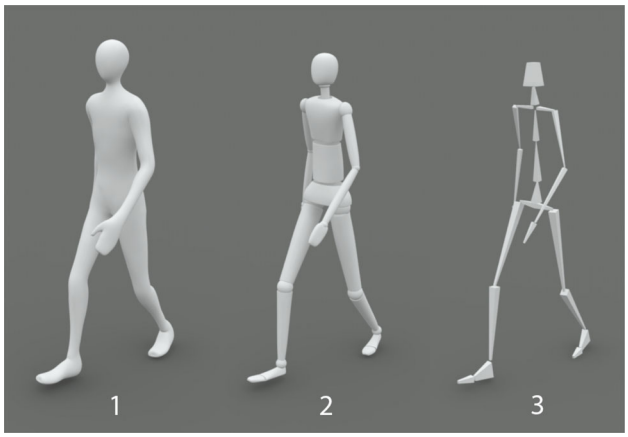

Fig. 1 Animated characters: skin (1), mannequin (2), and abstract (3)

The files were then passed to a Foley artist who specialises in AAA video games. Synchronised audio recordings were made that corresponded to *age* (young, medium, old); *gender* (male, neutral, female); *weight* (light, neutral, heavy); *health* (injured, neutral, healthy); and *confidence* (confident, neutral, and nervous). Variations in footwear, timing, posture, and cloth movements were utilised to create auditory variations (see Table 1). Each of the three clips had the same Foley performances applied to them for comparison purposes, silent versions were also prepared for inclusion. Silent versions were utilised so that any visual bias effect could be understood and accounted for. This is in recognition that that neutrality of the motion capture walk cycle neutrality could not be guaranteed. A five-point semantic differential scale was chosen for each parameter (age, gender, weight, health, and confidence) to capture responses, and participants used all five scales for each clip to obscure which parameter was being explored in any individual video.

The first set of questions addressed population, such as age, gender, occupation, hearing abilities, and eyesight correction. This was followed by asking participants about their familiarity with animated characters with regards to web videos, television, movies, games, VR, and AR, and prompted a response using a unipolar calendar scale: never; occasionally (less than once a month); once a month or more; once a week or more; daily. The sequence of video clips was fully randomised for each participant to minimise any order effect. A five-point bipolar semantic differential scale was used for all the questions about the variables controlled in the animations: *age* (young - old); *gender* (male - female);

Table 1 Foley props and techniques for chosen variables

Variable	Conditions	Props	Technique
Age	Young	Trainers, light material	Light-footed steps, random cloth Foley, scuffy feet
	Medium	Trainers, medium material	Medium-footed steps, heel-strike, quick/sharp cloth Foley
	Old	loafers, medium material	Medium-footed steps, slightly scuffy/lots of weight
Gender	Male	Trainers, medium material	Strong heel strike with medium cloth Foley
	Neutral	Trainers, medium material	Neutral steps with medium cloth Foley
	Female	Trainers, medium material	Ball of foot steps (slightly louder) with lighter/quieter cloth Foley
Weight	Light	Soft trainers, light material	Soft-footed steps, ball of feet
	Neutral	Boots, light material	Medium-footed steps
	Heavy	Boots, heavy material, rucksack, weights	Heavy-footed steps with weights, heel-strike
Health	Injured	Trainers, medium material	Limping-style footsteps with weight shifting, "flappy" cloth Foley
	Neutral	Trainers, medium material	Medium-footed steps with neutral cloth Foley
	Healthy	Trainers, medium material	Medium-footed steps, heel-strike, quick/sharp cloth Foley
Confidence	Confident	Trainers, medium material	Medium-footed steps, heel-strike, quick/sharp cloth Foley
	Neutral	Trainers, medium material	Medium-footed steps, medium cloth Foley
	Nervous	Trainers, medium material	Light-footed steps, flat-footed, "longer" cloth Foley with slight jitter

weight (light - heavy); *health* (injured - healthy); and *confidence* (confident - nervous). The order of the polar adjectives was set so that there could be no connotation of positive or negative aspect weighting due to a term being located on the left or right of the scale. There was also an option for N/A (not applicable) for each parameter. The inclusion of N/A ensured that participants could indicate when they believed a particular parameter to be unrelated to an animation, rather than make an unjustified assumption that an omitted value denoted when a parameter could not be perceived. This distinguishes it from the neutral response option on the Likert scale, which indicates the mid-point of the construct being queried. The questionnaire finished with an invitation to provide additional comments. The only compulsory question was that of age, in order to conform with the Edinburgh Napier University ethical requirements that participants be between the ages of 18 and 70.

3.4 Procedure

Participants sat at a small desk facing a laptop and a pair of loudspeakers in the Auralisation Suite at Edinburgh Napier University. They first read a printed participant information sheet, after which they read and signed an informed consent form. One of the researchers ran the experiment and sat quietly out of sight in the same room in case participants had any questions. After answering the population information questions, participants played a random sequence of clips. Each of the 15 s videos could be played only once, and participants were free to provide responses either during replay or immediately after, before progressing to the next clip. The session, which typically lasted between 30 and 40 min, finished with an invitation to provide additional written comments, and an optional debrief, where the experimental design was explained more fully and any questions answered. The responses were transcribed from the paper forms into an Excel spreadsheet to generate summary statistics, and statistical tests were performed in SPSS. All the additional comments were coded using NVivo.

4 Results

The results have been reported according to the five characteristics being explored: *age*, *gender*, *weight*, *health*, and *confidence*. Coded comments have been included in the appropriate subsections. To make informed comparisons between animations, it was necessary to exclude incomplete and N/A responses within the data collected for each characteristic measured. This accounts for the difference in sample size as each characteristic is analysed.

In the examination of each characteristic, the analysis involved two independent variables: Foley *condition*, which sought to compare a silent clip with three relevant Foley versions, and *model* as shown in Fig. 1 (abstract, mannequin, and skin). There was one dependent variable: participants' perception of each characteristic in question. Statistical testing used a two-way ANOVA with repeated measures and post-hoc analysis with the Bonferroni adjustment. Unless stated otherwise, sphericity was assumed not to have been violated by applying Mauchly's Test of Sphericity.

Sound was considered necessary to rate the requested parameters accurately; otherwise, it was *"pure guesswork"* (P01) or *"a lot harder to interpret"* (P10) and *"categorise"*

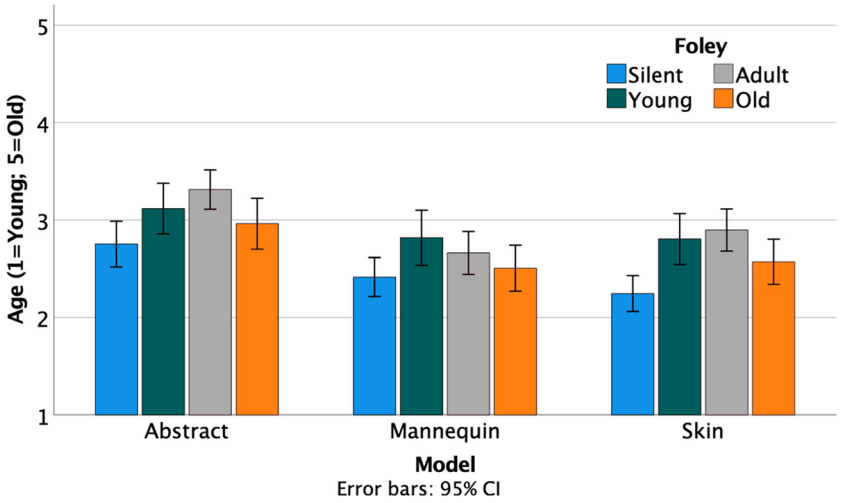

Fig. 2 Mean participant responses to age ($n = 77$) for Foley condition and model. Error bars indicate 95% confidence interval

(P52). P87 considered that their responses were mostly *"based on the sound and very little from what the animation showed"*, which was also the case for several other participants (P25, P45, P47, P55, P74, P78, P80, and P83).

4.1 Age

Figure 2 illustrates the participants' responses ($n = 77$) relating to the variable of age. The silent animations were rated, on average, as slightly young ($M = 2.47, SD = 0.93$), along with the abstract model ($M = 2.75, SD = 1.03$), the mannequin model ($M = 2.42, SD = 0.89$), and the skin model ($M = 2.25, SD = 0.81$). Upon the addition of Foley, the participants perceived the adult condition as being oldest on average ($M = 2.96, SD = 0.97$), followed by the young condition ($M = 2.91, SD = 1.18$), and then the old condition ($M = 2.68, SD = 1.08$). The condition rated as oldest overall was the adult abstract model ($M = 3.31, SD = 0.89$), whilst the youngest was the silent skin model ($M = 2.25, SD = 0.81$).

The number of participants who responded N/A for each of the animation model and Foley conditions is presented in Table 2. These figures indicate that participants found it less appropriate to rate the silent clip variations, when compared to those featuring sound. The abstract and mannequin models featured comparable levels of N/A responses.

Table 2 Participant N/A responses to age

Condition	Abstract	Mannequin	Skin	Σ
Silent	13	11	9	**33**
Young	7	5	2	**14**
Adult	5	4	3	**12**
Old	4	4	1	**9**
Σ	**29**	**24**	**15**	**68**

The effects of the Foley condition and model upon perceptions of age were examined. There was no interaction between condition and model, but significant differences were identified between conditions ($F(3, 228) = 7.698, p < .001$) and model types ($F(2, 152) = 20.473, p < .001$). Post-hoc analysis showed that for condition, significant differences existed between the silent and Foley clips of young ($p = 0.004$) and adult ($p < .001$), which were both perceived as being older. In the case of models, the mannequin ($p < .001$) and skin ($p < .001$) models were perceived as younger than that of the abstract type.

Participants' age had a small influence over the responses. The under 55 s all gave average age estimates of under three, whilst the 55 s and over were over three. In the under 55 s group, average age estimates grew slightly from the 18 to 24 group ($M = 2.31, SD = 0.41$), through the 25 to 34 s ($M = 2.66, SD = 0.41$), to the 35 to 44 group ($M = 2.76, SD = 0.27$), and then dipped slightly in the 45 to 54 s ($M = 2.74, SD = 0.36$).

Participant 30 reported that *"Young and Old are the most difficult"* with P92 stating that *"Age was a struggle"*. The speed of the walk cycle (P54, P60) and the *"upright walking position"* (P64) were two of the reasons provided for the animations not being considered old by some of the participants.

4.2 Gender

Figure 3 shows results for the animations without sound were on average rated by participants ($n = 73$) as slightly male ($M = 2.41, SD = 1.18$) with the mannequin close to neutral gender ($M = 2.97, SD = 1.28$) and the skin model towards male ($M = 1.96, SD = 0.98$). The addition of Foley moved the overall averages towards the gender-neutral position (3) for all animations with sound: male Foley animations ($M = 2.52, SD = 1.32$); neutral Foley ($M = 2.81, SD = 1.28$); and female Foley ($M = 2.88, SD = 1.34$).

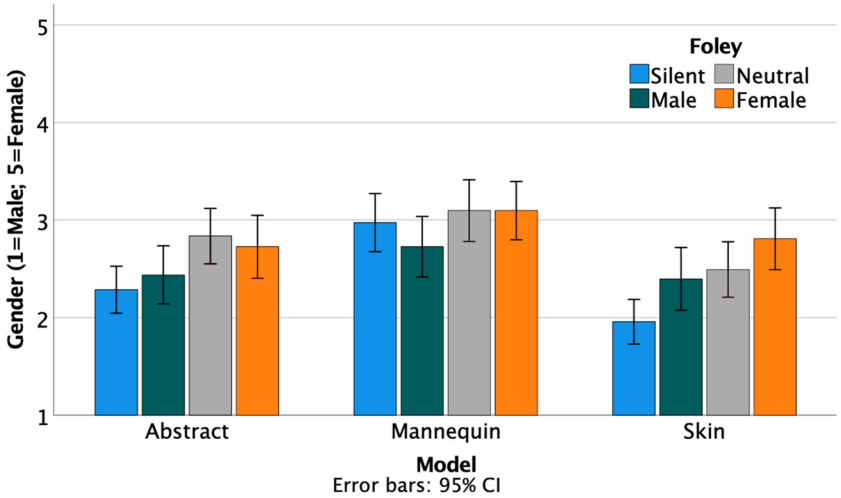

Fig. 3 Mean participant responses to gender ($n = 73$) for Foley condition and model. Error bars indicate 95% confidence interval

The number of participants who responded N/A for each of the animation model and Foley conditions is presented in Table 3. These figures indicate that participants found it less appropriate to rate the silent clip variations and similar, albeit lower, levels of N/A with respect to versions featuring sound. The abstract modes received the highest level of N/A responses, followed by mannequin and skin.

The effects of the Foley condition and model upon perceptions of gender were examined. There was no interaction between condition and model, but significant differences were identified between conditions ($F(3, 216) = 7.053$, $p < .001$) and model types ($F(2, 144) = 13.103$, $p < .001$). Post-hoc analysis showed that for condition, significant differences existed between the silent and Foley clips of neutral ($p = .004$) and female ($p = .002$) gender. The abstract model was perceived as more male than the mannequin ($p = .001$), and the skin model was perceived as more male than the mannequin ($p < .001$).

Participants' own reported gender did not have any influence over their responses for each of the animations between females ($M = 2.67, SD = 1.35$) and males ($M = 2.64, SD = 1.24$). Participants described how difficult gender was to determine (P01, P02, P18). It was thought that the animations looked masculine (P05, P25, P29, P30, P34, P39, P61), with *"not many female characters"* (P06). Perceived weight was reported as a way of gauging gender (P30, P38, P42, P64, P95), with heavier being male and lighter female. The choice of footwear also influenced responses (P54, P57).

Table 3 Participant N/A responses to gender

Condition	Abstract	Mannequin	Skin	Σ
Silent	17	8	8	**33**
Male	5	6	2	**13**
Neutral	10	3	3	**16**
Female	10	4	2	**16**
Σ	**42**	**21**	**15**	**78**

4.3 Weight

Analysis of the participants' perceptions of weight ($n = 83$) is illustrated in Fig. 4. All silent animations were considered by participants to be slightly light ($M = 2.29, SD = 0.93$) with the mannequin the lightest ($M = 2.21, SD = 0.85$), followed by the abstract ($M = 2.20, SD = 1.01$) and skin ($M = 2.54, SD = 0.89$) models. The addition of Foley raised overall mean ratings for the light ($M = 2.46, SD = 0.92$), neutral ($M = 3.81, SD = 1.03$), and heavy ($M = 3.82, SD = 1.06$) conditions. The skin model with neutral weight Foley was rated heaviest ($M = 3.98, SD = 0.88$).

The number of participants who responded N/A for each of the animation model and Foley conditions is presented in Table 4. These figures indicate that participants found it less appropriate to rate the silent clip when compared to the sounding versions, with no participants submitting N/A responses in the neutral and heavy Foley versions. The three models received comparable numbers of N/A responses.

The effects of Foley condition and model upon perceptions of weight were examined. There was no interaction between the Foley condition and model, but significant differences existed between conditions ($F(2.401, 195.885) = 148.086$, $p < .001$), with Greenhouse-Geisser correction applied as the assumption of sphericity had been violated ($\chi^{2(5)} = 29.425$, $p < .001$). Significant differences were found between model types ($F(2, 164) = 5.422$, $p = .005$). Post-hoc analysis showed that for condition, differences existed between the silent and Foley clips of neutral ($p < .001$) and heavy ($p < .001$). Comparable differences were also present between the light and neutral ($p < .001$) and heavy weights ($p < .001$), with neutral and heavy consistently perceived as heavier. The abstract model was perceived

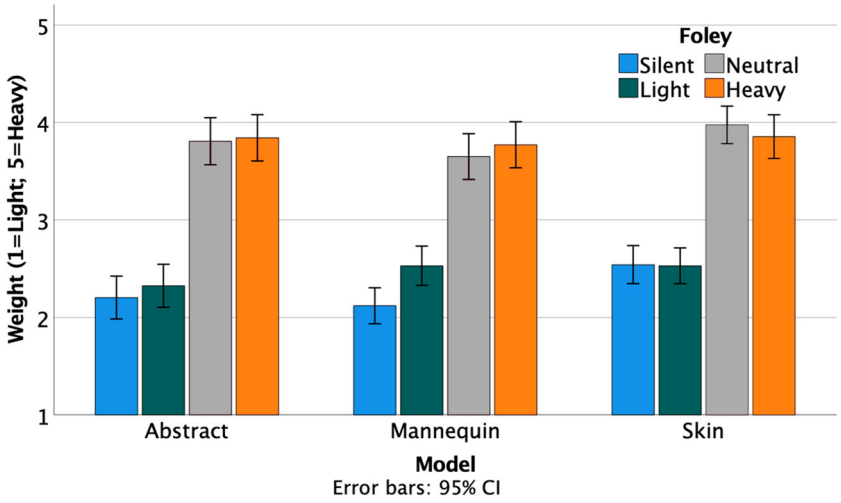

Fig. 4 Mean participant responses to weight ($n = 83$) for Foley condition and model. Error bars indicate 95% confidence interval

as lighter than the skin model ($p = .04$), as was the mannequin ($p = .006$).

Participants reported that *"Light and Heavy was the easiest"* (P30), with P31 stating that sound was required to make an estimation possible. The *"clothing rub"* suggested heavy to P46, which was made more explicit by P33 *"to indicate heaviness, as limbs rub against each other"*. The speed of the animation translated to lightness by P60 who also compared the animations to their own weight.

4.4 Health

The results for participants' perception of health ($n = 80$) are presented in Fig. 5. The animations without audio were rated on average by participants as slightly healthy ($M = 3.73, SD = 1.22$) with the mannequin considered healthiest ($M = 3.79, SD = 1.18$), followed by the skin ($M = 3.74, SD = 1.25$), and the abstract model ($M = 3.65, SD = 1.23$). The injured Foley condition was much lower with an overall rating tending towards injured ($M = 1.66, SD = 0.94$), whilst the healthy Foley condition was rated similar overall ($M = 3.75, SD = 1.07$) to the animations without sound.

The number of participants who responded N/A for each of the animation model and Foley conditions is presented in Table 5. These figures indicate that participants found it less appropriate to rate the silent clip when compared to the sounding versions. The three models received comparable numbers of N/A responses, with the mannequin model obtaining the lowest count of the N/A option.

Table 4 Participant N/A responses to weight

Condition	Abstract	Mannequin	Skin	Σ
Silent	9	9	10	**28**
Light	1	0	1	**2**
Neutral	0	0	0	**0**
Heavy	0	0	0	**0**
Σ	**10**	**9**	**11**	**30**

The effects of the Foley condition and model upon perceptions of health were examined. There was no interaction between condition and model or between the three different character models. However, significant differences were identified between conditions ($F(3, 237) = 180.660, p < .001$). Post-hoc analysis showed that significant differences existed between the injured clip and the silent ($p < .001$), neutral ($p < .001$), and healthy ($p < .001$) clips, with participants considering all but the injured clip were on the healthy side of the five-point scale. Differences were also noted between the silent and neutral clips ($p = .005$) and between the neutral and healthy clips ($p < .001$), suggesting that the neutral condition was able to be clearly differentiated from the other conditions presented, in the direction intended by the Foley design.

A couple of participants found it difficult to consider health on a scale (P13, P28). The irregularity of the sound conveyed injury (P18, P30), but hesitancy was considered applicable to both injury and nervousness (P60, P75). P04 found that the abstract figure could only be thought of as *"non-human"* and as such *"injury/ailment"* could not be applied to it, only a form of *"robotic awkwardness"*.

4.5 Confidence

The participants' perception of confidence ($n = 78$) in the animation clips is presented in Fig. 6. All the silent animations were slightly confident ($M = 2.35, SD = 1.00$) with the skin model being most confident ($M = 2.19, SD = 0.93$) followed by the mannequin ($M = 2.35, SD = 1.02$) and abstract model ($M = 2.50, SD = 1.04$). Overall, the confident animations were rated slightly more confi-

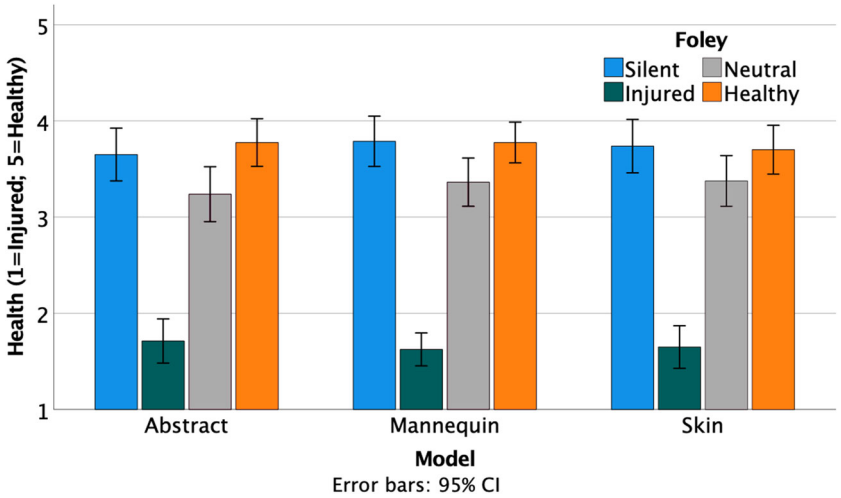

Fig. 5 Mean participant responses to health ($n = 80$) for Foley condition and model. Error bars indicate 95% confidence interval

dent ($M = 2.20, SD = 0.97$) than any of the other Foley conditions, although the neutral ($M = 2.75, SD = 1.04$) and nervous ($M = 2.78, SD = 1.04$) mean ratings were highly comparable. The nervous Foley condition with the mannequin model received a score that was furthest from confident, although it was nearest to the midpoint of the scale ($M = 2.87, SD = 1.02$), rather than the nervous end (5).

The number of participants who responded N/A for each of the animation model and Foley conditions is presented in Table 6. These figures indicate that participants found it less appropriate to rate the silent clip when compared to the sounding versions. The three models received similar numbers of N/A responses, with the abstract obtaining the largest count, followed by the mannequin and skin models, respectively.

The effects of the Foley condition (silent, confident, neutral, and nervous) and model upon perceptions of confidence were examined. There was no interaction between condition and model or between the three character models. However, significant differences were identified between Foley conditions ($F(3, 231) = 17.827, p < .001$). Post-hoc analysis showed significant differences between the silent clips and the neutral ($p = .002$) and nervous ($p < .001$) clips with the silent clip being perceived as more confident. This pattern was replicated between the confident Foley design clips and the neutral ($p < .001$) and nervous ($p < .001$) clips.

Table 5 Participant N/A responses to health

Condition	Abstract	Mannequin	Skin	Σ
Silent	11	8	11	**30**
Injured	1	0	0	**1**
Neutral	2	1	1	**4**
Healthy	1	1	2	**4**
Σ	**15**	**10**	**14**	**39**

Confidence was considered a difficult aspect to gauge (P30, P50). Once participant recognised a *"natural bias"* in themselves that they associated a *"heavy and confident style of walking with males"*. *"Loudness of step was associated with confidence"* (P43). P46 considered that *"without sound everyone seemed confident"*, P43 thought that *"heads down... might have indicated lack of confidence"* but *"all [of the figures were] facing forwards"*, and P64 stated that *"none of them felt nervous, largely based on the stride"*. Conversely, P18 reported that the *"floor creaking made the character seem more nervous"*.

5 Acoustic features and perceived characteristics

For both Foley artists and sound designers, being able to identify the acoustic features or other qualities of footstep sounds that convey characteristics of *age*, *gender*, *weight*, *health*, and *confidence* would be advantageous. Such awareness would present benefits to the sound design process, namely streamlining the procedure and ensuring that creative intentions could be communicated to audiences with the greatest likelihood of success. This section summarises what is already known in this respect, to what extent this existing knowledge has manifested itself in the sounds used within the study undertaken earlier, and to identify where additional insight would be beneficial.

5.1 Related work

5.1.1 Anthropometrics

From an analysis of the literature, characteristics of *weight* and *gender* have been consistently found to be related to spectral features of footstep sounds. This is irrespective of

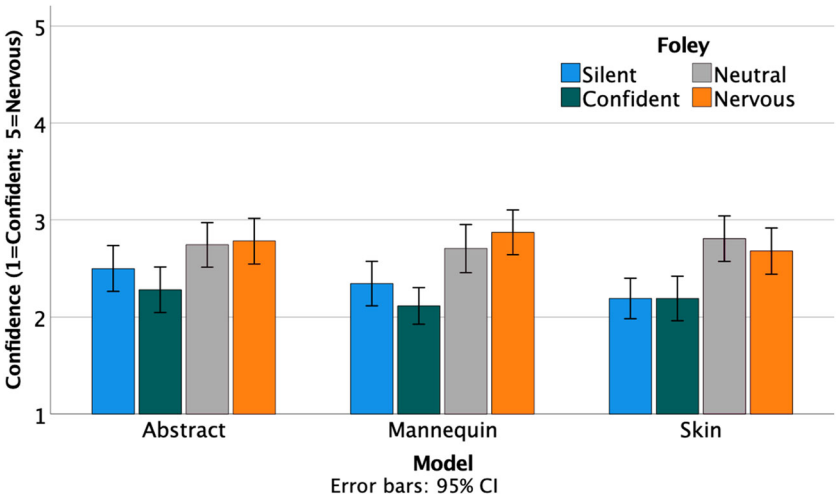

Fig. 6 Mean participant responses to confidence ($n = 78$) for Foley condition and model. Error bars indicate 95% confidence interval

whether these actions were recorded and subsequently analysed or designed using audio synthesis.

In the case of *weight*, the manipulation of frequency of footstep sounds has been shown to impact listener perceptions of an avatar's weight, although the consistency of this may be influenced by the type of surface that the footstep interacts with [50]. In a study where walkers' footstep sounds were played back to them via headphones in real-time, the manipulation of frequency spectra was shown to change perceptions of body weight [51]. Specifically, the enhancement, through amplification, of high frequency components (1–4 kHz) and attenuation of low frequencies (83–250 Hz). These manipulation of the sonic cues were related to perceptions of being lighter when compared to the unaltered sound, along with a low frequency version with inverse frequency manipulations to that of the high frequency one.

Beyond footsteps and considering sound design techniques more generally, it has also been shown that spectral frequency is a communicator of aspects such as *weight*. Low frequency sounds are perceived as heavier than high frequency or silence [52].

Later work reinforced these principles relating to perception of *weight*, with experiments conducted where participants exercised with a gym step and climbed flights of stairs [53]. That work extended the earlier findings to provide indications that, in terms of *gender* perception, participants felt more feminine in the presence of enhanced high frequency components and attenuated low frequency relayed footstep sounds. Those who took part in the study also felt more masculine in the inversely manipulated footstep sounds (attenuated high frequencies and amplified low frequencies). The authors additionally highlighted that a participant's own personal characteristics may influence their perceptions. A comparable set of outcomes was discovered in more recent work [54], which replicated many of the principles from [51, 53]. The authors also highlighted that there may be differences between male and female participant perceptions, with females identifying more pronounced effects of footstep manipulation upon the perceived characteristics of gender and weight.

A notable difference is that these studies were concerned with a person's perception of their own, personal characteristics due to hearing associated footstep sounds. Whereas the research conducted for this study seeks to identify the perception of such characteristics outside of oneself and, in this case, in a virtual avatar.

In terms of a walker's *gender*, high frequency spectral components are commonly found to be indicative of females, whilst males tend to have greater low frequency components in their footstep sounds [40, 55, 56].

In studying the spectral energy distributions of male and female footstep sounds, Li, Logan, and Pastore [40] found that *gender* was communicated by spectral peak and high frequency spectral components, affirming that frequency domain information is most important. Similar observations are also found in other research, where differences in the frequency content of the sole and heel strikes of the foot showed spectral differences between male and female walkers on a vibration sensor [56].

In examining a consistent set of synthesised footstep sounds that were designed to portray a *genderless* walker, Turchet and Serafin [57] found that amplitude variations

Table 6 Participant N/A responses to confidence

Condition	Abstract	Mannequin	Skin	Σ
Silent	8	10	8	**26**
Confident	1	0	1	**2**
Neutral	4	3	1	**8**
Nervous	4	1	1	**6**
Σ	**17**	**14**	**11**	**42**

and surface (for example, sand, grass, metal, snow) made no significant difference to gender perceptions by listeners employing both loudspeakers and headphones. This outcome supports the view that gender characteristics are unlikely to be communicated by the use of volume, at least in isolation, with spectral features being more likely to be useful in this respect.

In another study of footstep sound synthesis, relatively high frequency components within step impacts and the sound of particles associated with surfaces were linked with identification of females (and conversely for males). These spectral components were in addition to temporal aspects, specifically variations in walking pace, with faster steps connected with females [58]. However, the pace of the walker is deliberately controlled in this research and so not currently applicable.

From a principal components analysis of footstep sounds [38], spectral slope was identified as a factor that may allow the characteristics of weight to be communicated and be attributable to the changes encountered at the phases of *weight* transfer, between heel and sole strike, in walkers. This component also contributed to the footstep sounds of walkers of different gender in upright positions. Peak frequency and spectral centroid contributed to principal components, providing characteristics of weight and gender in stooped walking positions.

5.1.2 Non-anthropometrics

Whilst our own exploration of the literature has not been systematic, there appears to be very little in the way of published studies into listener perceptions of *age* from footstep sounds. Nevertheless, a relationship between age [58, 59], gait [51, 60, 61], and associated footsteps is suggested and is considered to be intuitive. There is the suggestion that footstep events, specifically heel and toe strikes, are significant factors in perception of age [62], indicating that changes in loudness and ADSR envelope may be relevant.

In considering techniques and factors in synthesised footsteps intended to express emotions, *age* is identified as a dependent variable (along with gender, weight, and others) that can be manipulated by controlling synthesis parameters for the footstep sound, such as impact force, distance, ground texture, materials, and so forth [59]. However, no direct or indirect mapping of these variables is suggested that would provide an insight into what aspects of the footsteps contribute to the perception of these characteristics.

The definition of *health* and being *healthy* is very broad. In relation to footsteps, this is most intuitively able to manifest itself through the representation of a walker's gait. It has been discovered that emotional arousal and positive feelings were also increased with high frequency sound versions of footsteps, possibly indicating that this would contribute to impressions of healthiness [51].

Spectral analysis is an important contributor to the identification of different footstep characteristics, use of similar approaches can also be found in more general footstep and *gait*-related research studies. For instance, person identification using spectrotemporal modulation features can be achieved via a wavelet transform of sound spectrum, which forms the input to a Support Vector Machine classifier [63]. Person identification has been facilitated through analysing gait with Mel-Frequency Cepstral Coefficients (MFCCs) and Hidden Markov Models [64]. In other research, employing MFCCs, represented as images, was utilised as input to Convolutional Neural Networks [61]. Person identification using psychoacoustic features of loudness, sharpness, fluctuation strength, and roughness, in addition to spectral envelope and cepstral features, using k-means clustering are possible [60, 65]. Spatio-temporal features are able to help identify clinical and biometric gait parameters [66].

Confidence, possibly by its nature as a difficult concept to define, is much less discussed in the literature relating to footsteps and sound design in general. Results have been gained in studies relating to speech, where fundamental frequency, loudness, and speech rate were discriminating factors. Lower mean frequency and faster speaking rates have been associated with greater confidence [67]. A related example in the field of speech analysis was conducted to examine features, and their predicative ability when combined with boosted decision tree machine learning techniques, to identify confidence as well as doubt in English spoken word with source signals drawn from three different groups of accents [68]. The range of fundamental frequency, mean amplitude, duration of utterance, and mean harmonic-noise ratio were found to influence perceptions of confidence.

Similarly, concepts of *confidence* in speech were investigated that made use of fundamental frequency analysis of speech, in addition to temporal elements that measured the speaking rate along with presence of filled pauses (um and ers) [69]. The latter of these characteristics may be considered akin to being able to clearly identify the signal from background noise or distractors. The findings are further supported in speech research where variations in loudness can communicate the level of certainty of a speaker [70] and confidence [71].

Published literature suggests that frequency components may contribute to conveying characteristics of *confidence*, although this has yet to be tested or discovered in the case of footstep sounds. The presence of these aspects, particularly identifying gaps between the signal component of the audio as well as its rate, has parallels with footstep sounds. Specifically, this may lead to a postulation that the cadence or walking rate observed by audiences might be indicative of the walker's confidence along with the nature of any auditory

filler content in gaps between footstep strikes. Whilst the walking rate was fixed in the sample clips that participants in this research rated, the sound design did intentionally add aspects of jitter and longer duration cloth artefacts to communicate the nervous condition (see Table 1). As such, there may be aspects of confidence that relate to rhythmic qualities and the presence of noise between intentional footstep sounds. However, there is also evidence from a study where participants rated confidence in their own utterances, that speech rate may not be a universally reliable indicator. Being potentially dependent upon the task and situation [71], which suggests that these concepts require further, more detailed investigations.

Overall, there is a mixed pool of existing research into the characteristics of *age*, *gender*, *weight*, *health*, and *confidence* that this research focuses upon. A significant body of work exists relating to both gender and weight, especially with respect to footstep sounds, of which the auditory features relating to frequency and spectrum are most salient. However, the non-anthropometric characteristics of age, health, and confidence are much less explored, although some parallels can be explored in other audio and signal processing domains, notably in the case of confidence and its manifestation in speech.

5.2 Analysis of Foley footstep sounds

To explore and validate these findings further, the sounds used in this study were subjected to an acoustic analysis. Namely the extraction of a set of audio features (frequency analysis on the Bark scale, amplitude envelope, and novelty) which could be examined as factors contributing to the average of the perceptual responses obtained from participants in the study.

Since the number of unique Foley sequences produced for this study was small (15), the approach of performing an inspection of specific audio features, the selection of which was directed by the literature above, with respect to the five characteristics being portrayed, was taken. To facilitate this procedure, each of the 15 Foley sound recordings (silent versions were not considered) was peak normalised to -3 dB and processed in MATLAB using version 1.8.1 of the *MIRtoolbox* [72]. Specifically, use was made of frequency, envelope, and novelty analysis, which were applied to Foley characteristics as indicated in the literature. Age was an exception of age, where a more exploratory approach was adopted due to a lack of consensus in prior research.

As highlighted in Sect. 5.1, frequency characteristics have been a key factor to almost all analysis and synthesis of footstep sounds with respect to a range of characteristics of the respective walker. The one exception being the characteristic of *age*, where there is less existing information available.

As such, each of the five characteristics was investigated to allow for an inspection and comparison of the frequency profile across the three Foley sequence variations produced. To accomplish this, a plot of each Foley sound's spectrum was realised using a Discrete Fourier Transform (DFT). To make this process more representative of a human listener's perception, the frequency axis of the DFT was transposed to the Bark scale, with each element of the axis representing one of the 24 critical bands identified for human hearing [73, 74].

To examine aspects of the sounds relating to loudness and to potentially reveal differences that may occur due to gait, jitter, and heel-toe strikes, envelope analysis was undertaken to help visualize each event in the Foley sequences and their relative amplitude. Turchet [45] found the analysis of amplitude envelopes useful to illustrate different footsteps created with a variety of footwear and on a range of ground surfaces. The envelope analysis figures are produced in the *MIRtoolbox* [72] by filtering the audio signal using a Hilbert transform, extracting the absolute envelope shape, and then smoothed using lowpass filtering to remove high-frequency components. Finally, the signal is down-sampled to remove redundant information from the smoothing step. Each 15 s sequence was analysed overall, to help identify any similarities or differences that may exist across the sequence.

As an additional tool, particularly in trying to quantify and identify features of footstep strikes and how different each strike may be from one another, a novelty curve for clips is also included, where pertinent. This draws upon a deeper analysis of the data produced from an audio similarity matrix [75, 76]. It is hoped that this too may illustrate aspects of differing gait and jitter, especially in the less explored characteristics of *age* and the non-anthropometrics of *health* and *confidence*. When computing the novelty information, use was made of the multi-granular approach [77].

5.2.1 Age

The Foley sounds intended to convey *age* differences are shown in Fig. 7. Recall from Sect. 4.1 that significant differences were only discovered between the Foley versions and the silence. The general profile of frequency distribution for the Foley versions is similar, with the bulk of content being in the lower critical bands (1 to 5) in slightly differing amounts. The sound of old has a notable increase in higher critical bands (10 to 24) compared to young and neutral, which may partly explain the mean rating differences participants reported, though these were not statistically significant. This finding would suggest that frequency information alone is unlikely to be used in the discrimination of a walker's age.

In reviewing envelope information from Fig. 8, there is a similar pattern of events occurring between the adult and old

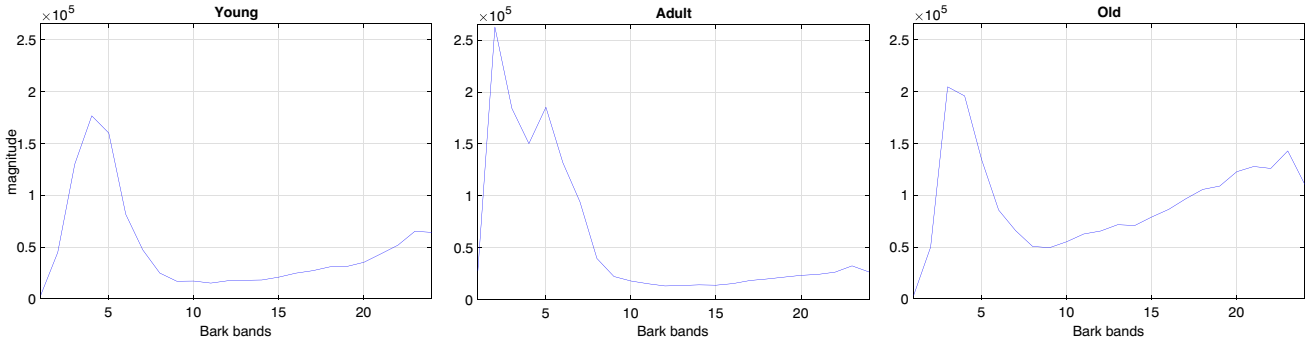

Fig. 7 Age - frequency analysis (Bark scale)

pairing. Although the cleaner strikes of the adult condition are evidenced by more pronounced spikes, followed by a drop to near zero amplitude. The young clip follows a similar pattern to old, albeit with peaks reaching an amplitude that is, on average, lower than the adult and old sounds.

The novelty curves for each of the three sounds, in Fig. 9, are generally unremarkable in terms of their differences. However, the neutral condition shows slightly less presence of novel events, potentially indicating a greater amount of consistency in strikes across the duration of the clip.

5.2.2 Gender

A similar profile is encountered in the review of the *gender* characteristic. In Sect. 4.2 the male Foley version differed more from neutral and female, which participants tended to rate more closely, though not with any statistical significance, other than compared with the silent versions. An inspection of the frequency profile for each sound is displayed in Fig. 10, illustrating that each version mainly contained information in between the same critical bands (1 to 8). With the male version having a greater magnitude in this region, especially in critical band two, which may indicate subtle differences being conveyed, but not enough in isolation to allow the sounds to be sufficiently distinguished.

5.2.3 Weight

Notwithstanding the silent versions, for the *weight* characteristic, the differences in frequency distributions again mirror the ratings received by participants in Sect. 4.3. It was noted that the light and neutral pairing along with the light and heavy pairing were significantly different. It can be seen from Fig. 11 that the neutral and heavy sounds have broadly similar profiles, with most of the signal magnitude in the lower critical bands (1 to 7). The light Foley sound also follows such a profile but has additional spikes in higher critical bands (12, 15, 17, and 19). This finding is in consonance with the literature, where frequency was highlighted as the most dominant indicator of weight.

5.2.4 Health

For *health*, the results from Sect. 4.3 indicated that, aside from silence, there were perceived differences between all pairings of the injured, neutral, and healthy sounds. Upon analysing the frequency graphs in Fig. 12, the stark differences between the pairs of injured and neutral, as well as injured and healthy, from Sect. 4.3 are not as immediately apparent. There is certainly an increased magnitude in the neutral sound clip, though all three tend to contain strongest elements in lower points (1 to 8) on the Bark scale. Whilst the

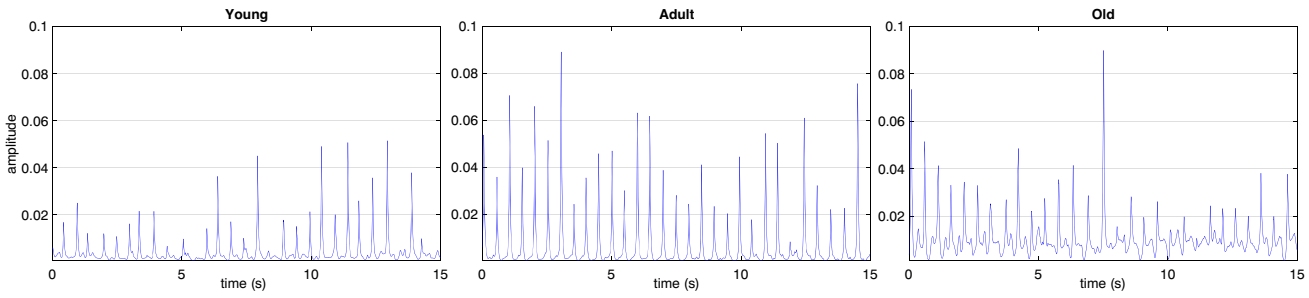

Fig. 8 Age - envelope analysis

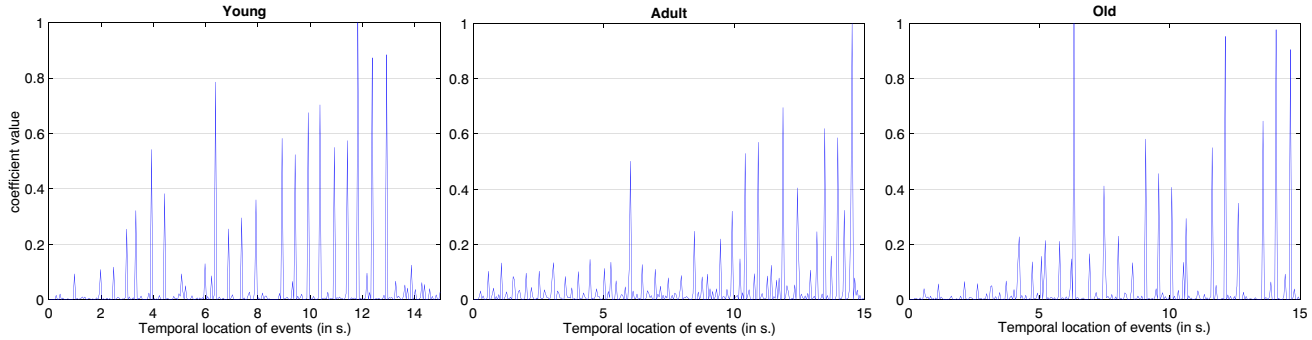

Fig. 9 Age - novelty

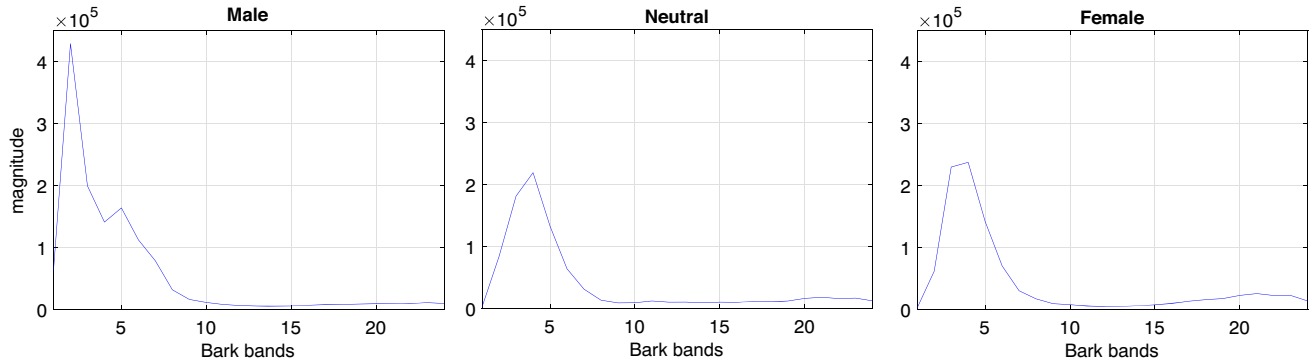

Fig. 10 Gender - frequency analysis (Bark scale)

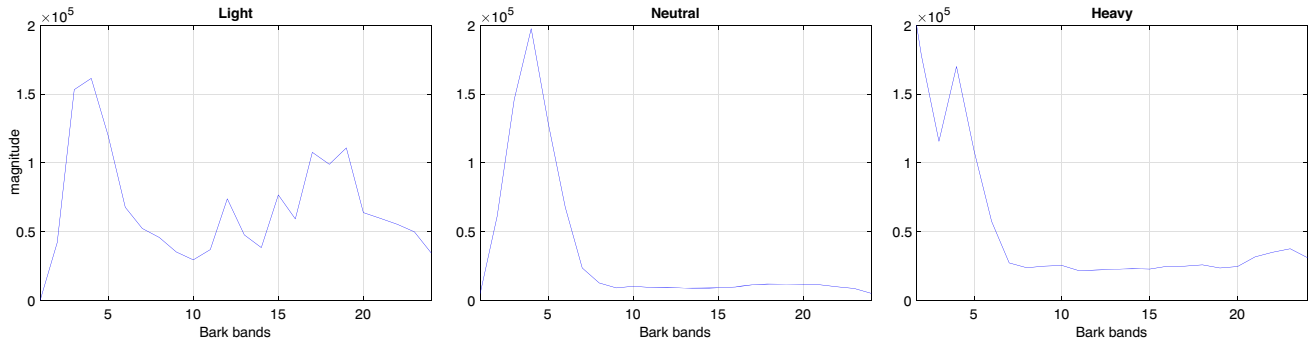

Fig. 11 Weight - frequency analysis (Bark scale)

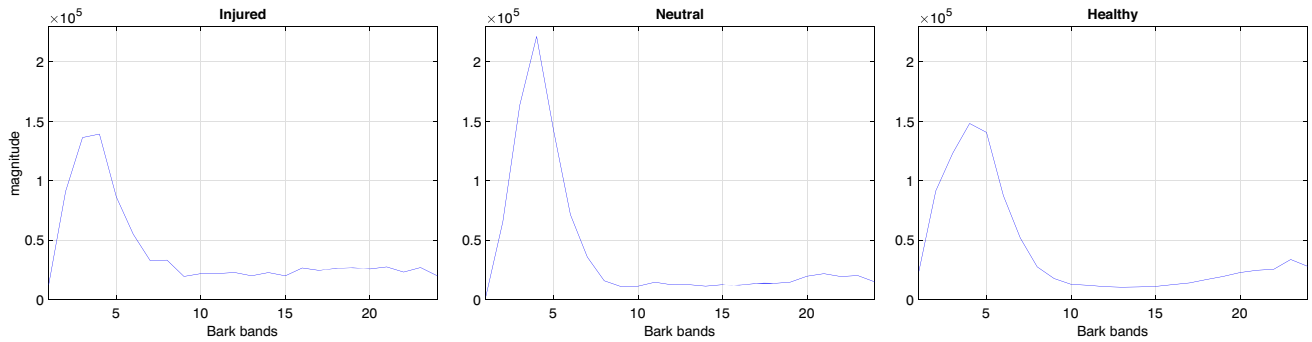

Fig. 12 Health - frequency analysis (Bark scale)

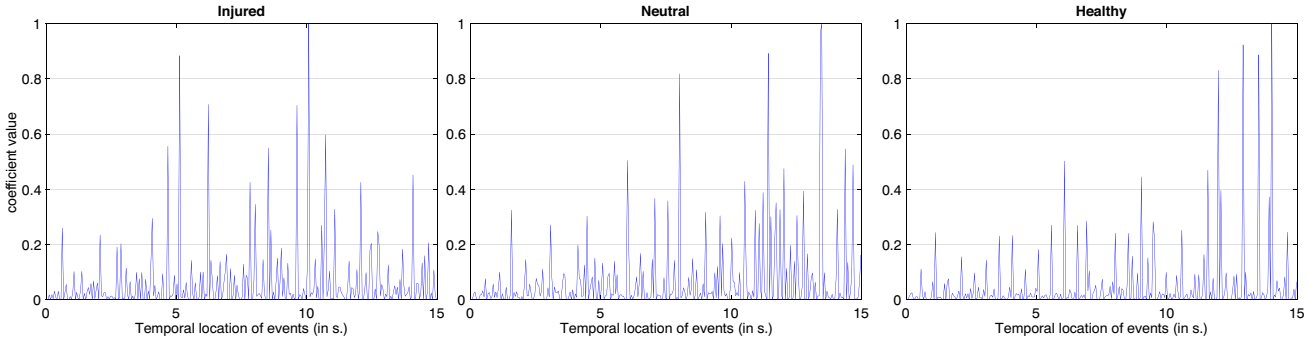

Fig. 13 Health - novelty

magnitude difference may party account for the participants' perceptions, it would seem the frequency information alone is not enough to clearly distinguish between these sounds.

The novelty analysis of the three Foley sequences relating to health, shown in Fig. 13, demonstrates predominantly similar patterns. There is a slightly lower occurrence of larger novelty events present in the healthy condition, indicating potentially the presence of more consistency over its duration and fewer events with large novelty spikes. This may go some way to account for differences between the healthy clip and the other two, but it is not proposed that this feature alone may account for the differences in participants' perceptions.

As such, the health characteristic is one that may warrant more detailed study and investigation in future in an effort to understand how it is perceived. The literature has suggested that aspects of gait are contributory and finding additional measures of this may be advantageous in the future.

5.2.5 Confidence

Finally, in the case of *confidence*, when disregarding the silent versions, participant data in Sect. 4.5 found the nervous and neutral clips to be very similar, whilst significant differences existed between the pairings of confident and neutral, along with confident and nervous. Examining the frequency graphs of Fig. 14 would seem to explain some of these findings although, as per the literature, frequency information alone may not be sufficient. The distributions between the pair of confident and neutral are stark, with the neutral clip having greater magnitude in almost every point on the Bark scale, in addition to a rise in the later critical bands (18 to 23). There is a less pronounced difference between confident and nervous clips, although the magnitude across all bands is more consistent in nervous than confident, in addition to confident having a more pronounced spike in lower critical bands (1 to 7).

The envelope analysis, shown in Fig. 15, of the confidence sound clips is particularly supportive of the difference between the confident and neutral pair of sounds, as was also discovered in their frequency content. In complement to the frequency analysis, a greater amplitude of events is clearly visible along with a greater noise floor between strikes. However, the differences reported by participants between the confident and nervous pair are still yet less obvious from their envelopes. Whilst being similar in amplitude, although less pronounced and consistent in the case of the nervous clip, which likely relates to the 'lighter' intention of the Foley artist. In this sense, the amplitude envelope may also be a useful feature in discriminating between these different levels of confidence.

A review of the novelty curves in Fig. 16 further illustrates the perceived differences between the confident and neutral sound clips participants experienced, showing much higher

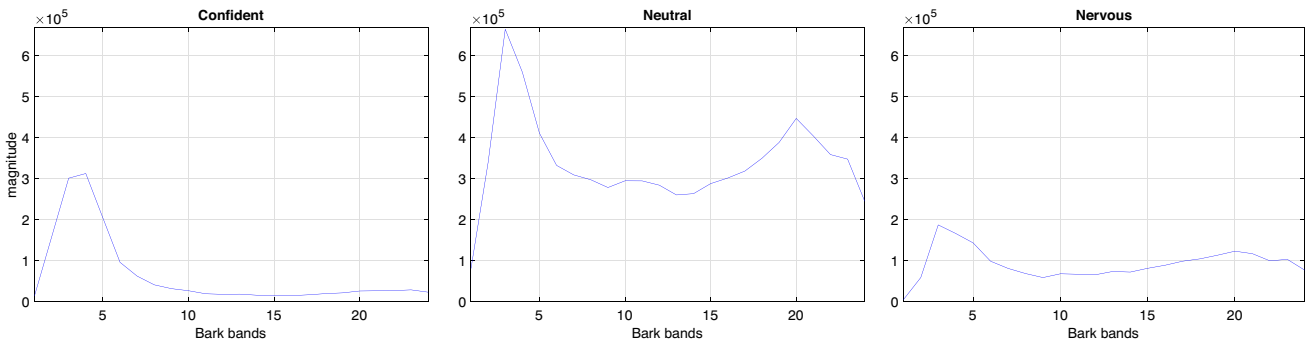

Fig. 14 Confidence - frequency analysis (Bark scale)

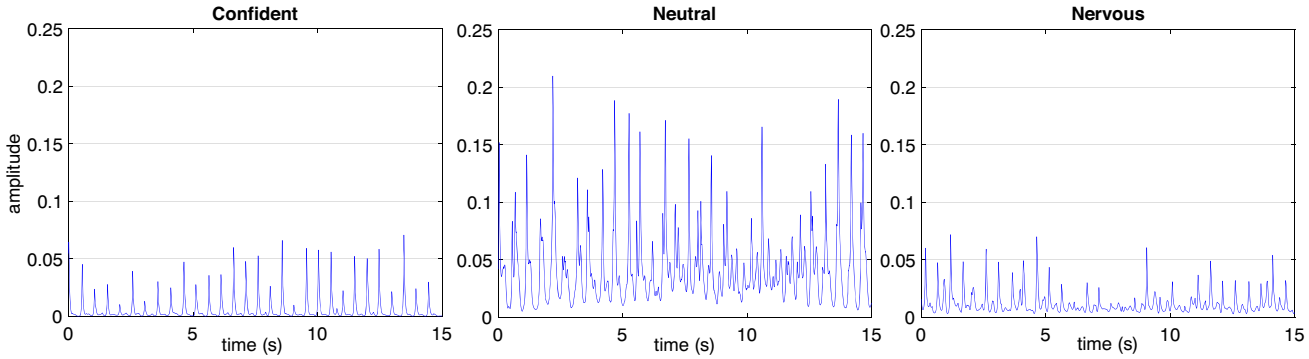

Fig. 15 Confidence - envelope analysis

spikes of novelty in the neutral clip. The novelty curves also appear to provide a mechanism by which the confident and nervous clips can be differentiated. This is something which was lacking in the frequency and envelope analysis. Between these two pairs, we are now able to identify greater novelty and more significant spikes, as was demonstrated between the confident and neutral pairing. It is hypothesised that these differences may be indicative of greater consistency and absence of jitter in the gait of the confident walker. Thereby making it likely that the novelty feature would be valuable as a marker of differences in walking behaviour.

6 Discussion

Regarding the first research question of this article (*To what extent can the characteristics of age, gender, weight, health, and confidence be successfully communicated to an audience through Foley-designed footsteps in an animated, neutral avatar, with a variety of visual designs?*), in general terms, the Foley created for this study could be considered successful in conveying all the characteristics, but arguably most clearly in the case of weight, health, and confidence. As is perhaps best illustrated in the confidence characteristic, with the nervous condition making the animations seem more neutral rather than nervous, Foley is often better at communicating one of the extremes of the scale than the other. Whilst there are statistically significant differences present between the different Foley conditions, it would, of course, be desirable to note larger spreads across the five-point scale used by participants to report upon these characteristics. In this respect, the characteristics of weight and health may be considered most successfully communicated using the Foley designs.

Upon inspection of the N/A responses from participants of the study, there was a clear trend of the largest numbers occurring in the silent condition across all the different characteristics being examined. This is likely due to the absence of sound and *synchresis* in those stimuli, suggesting that sound is indeed a key component in allowing audiences to perceive, and form opinions upon, characteristics of an avatar. The number of N/A responses followed less of a pattern over all Foley characteristics when it came to comparing between the three visuals that were used for the avatar, although for most characteristics, the abstract model tended to achieve the highest count. This situation is less clear but may be due to difficulty participants had recognising these anthropometric, "human-like" qualities in models that appeared less human.

The skill of the Foley artist and choice of props will almost certainly have had an impact upon the perceived parameters, as will the visual elements, having shown frequent

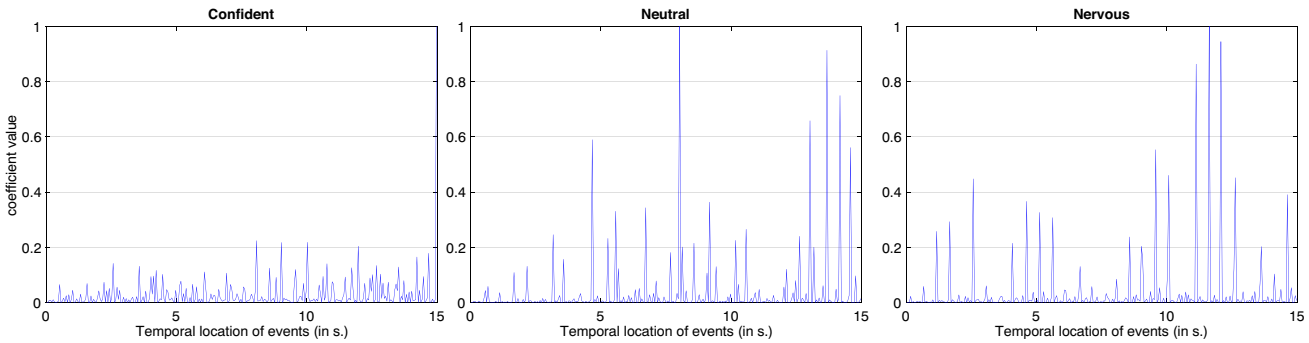

Fig. 16 Confidence - novelty

differences between the three model types in characteristics of age, gender, and weight. Since the walk cycle used was the same in all the video clips, participants reported perceiving variations where the only visible change was the character models employed. The current work is limited by having one Foley artist produce the sounds for the study, and this may have affected the generalisability of the results. This is a factor that could be addressed in subsequent studies, although it was hoped that the training and professional standing of the Foley artist mitigated this. Nevertheless, the artistic and creative nature of Foley could be controlled in future work, and the research presented here provides a framework for other researchers.

The effect of the three animated models (abstract, mannequin, and skin) was variable, with significant effects being found in the case of age, weight, and gender, but these did not interact with the Foley conditions. As such, the anthropometric features of a walker seem to be partly conveyed by the nature of the model, whilst the non-anthropometric characteristics of health and confidence do not appear to be. Although the focus of our work was primarily upon the role that footstep sounds played in this process, future work should continue to investigate how the visual appearance of a neutral avatar might influence these and other judgements about a character.

The selection of the same motion capture data to be used across all the Foley conditions and animated avatar models was a deliberate choice for the purposes of this research, to control for any effect that this might have upon participant perceptions. As identified during the design of the study, the motion capture data cannot be guaranteed to be completely neutral, at least with respect to the five characteristics that this study investigated, since the authors of this research do not know the characteristics of the walker used to create the motion capture data. As such, whilst this is useful as a *control* condition, it cannot be considered strictly *neutral*.

The participant perceptions of the silent models provide the closest insight into the neutrality of the data across the three different avatar models. The silent clips were also the ones that obtained the largest number of N/A responses. As such, future work in this area would do well to create as neutral as possible a walk cycle, potentially by synthesising this, rather than using motion capture data, which will inherently be limited by the nature of the walker whose data was captured. Similarly, it would be interesting to study the effect that motion data from walkers who are representative of the spectrum within the five characteristics (for example, in the case of age, a young, adult, and old walker) have in combination with the Foley conditions.

In terms of the second research question of this article (*What audio features contribute to participant perceptions of Foley footsteps, particularly with respect to the characteristics of the walker they are intended to represent?*), for the characteristics studied, it became apparent in the literature that certain types, or classes, of audio parameters are markers that human listeners might use to perceive different levels or states within each. There was a greater body of literature related to anthropometric qualities of weight and gender. For example, when assessing synthesised footstep sounds, Turchet's participants could successfully identify the body size intended to be conveyed with a range of footwear and surfaces, along with modulating perception of gender, with changes primarily being facilitated by manipulations to the amplitude and frequency [45]. The non-anthropometric characteristics of health, confidence, and age are somewhat less developed and conclusive, especially in the case of age. When it comes to health and confidence, the broader concept of gait becomes increasingly relevant. This arguably has application in the case of age too, since these characteristics are likely to be interrelated. It was found that audio features pertaining to frequency could be related to participant perceptions of walker characteristics. In the case of the non-anthropometric measures, amplitude envelope and novelty might also be viable. This aspect requires much further study, ideally with a large dataset of footsteps, where the walker characteristics are known. Such a dataset would provide more conclusive insight into the ability of these features, and likely others, to predict walker characteristics. Such knowledge would be beneficial not only in informing sound designers of how to produce Foley footsteps, but also in person and walker identification systems for a wide variety of applications.

This work could be considered relevant not only to animators within linear and interactive media, but also by any designer who wishes to integrate additional information in a natural manner to any form of moving object. Sound seems to be perceived by the participants in this study as a natural extension of the abstracted figures, despite the artificial rendering. This synchronicity is something that Foley artists have been perfecting for decades, and which arguably sound effects artists have evolved over Millenia, first for religious ceremonies and latterly for theatre.

The next stage is to explore these techniques within Augmented and Mixed reality so that the line between real and augmented artefacts can be blurred. Sometimes, it is essential that end users understand that an element is only an augmentation, whereas at other times it is desired that any augmentation is indistinguishable from the real-world objects it is overlaying. Sound can be used to make the object seem more artificial or natural according to the needs at any point. Simple parameters like weight can be altered so that the expected amount is conveyed, or that the object is too light or heavy. The health of an artefact might also be suggested as invisible corrective parameters are adjusted in real-time. Foley may also be applied in these situations for practical reasons, such as blurring between when a real person is presented or an artificial avatar, which could be used to mask

bandwidth difficulties, for example. There is also the potential to apply this to any video capture or camera-driven avatar to either make it more extreme or neutral in order to optimise anonymity within the metaverse. Future studies to validate the efficacy of sound design techniques, whether recorded or synthesised, might also make greater use of qualitative data capture and analysis, to better understand the perceptions of participants.

Acknowledgements Many thanks to Richard Hetherington, Gregory Leplatre, and Rob Brown who assisted with the research.

Data Availability The authors did not seek permission to share the dataset from participants, so it is not possible to make it available.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Whittle MW (2014) Gait analysis: an introduction. Butterworth-Heinemann, Oxford, UK
2. Baker R (2006) Gait analysis methods in rehabilitation. J Neuroeng Rehabilitation 3:1–10
3. Kirtley C (2006) Clinical gait analysis: theory and practice. Elsevier Health Sciences, Philadelphia, USA
4. Nigg B, Fisher V, Ronsky J (1994) Gait characteristics as a function of age and gender. Gait & Posture 2(4):213–220
5. Ostrosky KM, VanSwearingen JM, Burdett RG, Gee Z (1994) A comparison of gait characteristics in young and old subjects. Phys Therapy 74(7):637–644
6. Samson MM, Crowe A, De Vreede P, Dessens JA, Duursma SA, Verhaar HJ (2001) Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and body weight. Aging Clinical Experimental Res 13:16–21
7. Bonilla Yanez M, Kettlety SA, Finley JM, Schweighofer N, Leech KA (2023) Gait speed and individual characteristics are related to specific gait metrics in neurotypical adults. Scientific Reports 13(1):8069
8. Stevenage SV, Nixon MS, Vince K (1999) Visual analysis of gait as a cue to identity. Appl Cognitive Psychol: The Official J Soc Appl Res Memory Cognition 13(6):513–526
9. Hicheur H, Kadone H, Grezes J, Berthoz A (2013) Perception of emotional gaits using avatar animation of real and artificially synthesized gaits. In: 2013 Humaine association conference on affective computing and intelligent interaction, IEEE pp 460–466
10. Chung S-k, Hahn JK (1999) Animation of human walking in virtual environments. In: Proceedings computer animation 1999, IEEE, pp 4–15
11. Rose C, Cohen MF, Bodenheimer B (1998) Verbs and adverbs: multidimensional motion interpolation. IEEE Comput Graphics Appl 18(5):32–40
12. Badathala SP, Adamo N, Villani NJ, Dib HN (2018) The effect of gait parameters on the perception of animated agent's personality. In: Augmented reality, virtual reality, and computer graphics: 5th international conference, AVR 2018, Otranto, Italy, June 24–27, 2018, Proceedings, Part I 5, Springer pp 464–479
13. Thaler A, Bieg A, Mahmood N, Black MJ, Mohler BJ, Troje NF (2020) Attractiveness and confidence in walking style of male and female virtual characters. In: 2020 IEEE Conference on virtual reality and 3d user interfaces abstracts and workshops (VRW), pp 678–679. https://doi.org/10.1109/VRW50115.2020.00190
14. Cunningham S, McGregor IP (2022) Manipulating Foley footsteps and character realism to influence audience perceptions of a 3D animated walk cycle. In: Proceedings of the 17th international audio mostly conference, pp 113–120
15. Ament VT (2014) The foley grail: the art of performing sound for film, games, and animation. Routledge, New York, USA
16. Yewdall DL (2012) The practical art of motion picture sound. Routledge, New York, USA
17. Holman T (2012) Sound for film and television. Routledge, Burlington, MA, USA
18. Beauchamp R (2013) Designing sound for animation, 2nd edn. Routledge, New York, USA. https://doi.org/10.4324/9780240825007
19. Wright B (2014) Footsteps with character: the art and craft of Foley. Screen 55(2):204–220
20. Winters P (2017) Sound design for low and no budget films. Routledge, New York, USA
21. Pinheiro S (2016) Acousmatic Foley: staging sound-fiction. Organised Sound 21(3):242–248
22. Aly L, Penha, R, Bernardes G (2017) Digit: a digital Foley system to generate footstep sounds. In: International Symposium on Computer Music Multidisciplinary Research, Springer, pp 429–441
23. Donaldson LF (2017) You have to feel a sound for it to be effective: Sonic surfaces in film and television. In: The Routledge companion to screen music and sound. Routledge, New York, USA, pp 85–95
24. Warren WH Jr, Kim EE (1987) Husney, R The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. Perception 16(3):309–336
25. Hug D (2014) Kemper, M From Foley to function: a pedagogical approach to sound design for novel interactions. J Sonic Stud 6(1):1–23
26. Anderson DB, Casey MA (1997) The sound dimension. IEEE Spectrum 34(3):46–50
27. Chion M (2019) Audio-vision: sound on screen. In: Audio-Vision: Sound on Screen. Columbia University Press, New York, USA
28. Donaldson LF (2014) The work of an invisible body: the contribution of Foley artists to on-screen effort. Alphaville: J Film Screen Med (7):1–15
29. Hagood M (2014) Unpacking a punch: transduction and the sound of combat Foley in fight club. Cinema Jo 53(4):98–120
30. Ennis C, McDonnell R (2010) O'Sullivan, C Seeing is believing: body motion dominates in multisensory conversations. ACM Trans Graphics (TOG) 29(4):1–9
31. Mastoropoulou G, Debattista K, Chalmers A, Troscianko T (2005) The influence of sound effects on the perceived smoothness of rendered animations. In: Proceedings of the 2nd symposium on applied perception in graphics and visualization, pp 9–15

32. Lewis M (2015) Ventriloquial acts: critical reflections on the art of Foley. New Soundtrack 5(2):103–120
33. Deleuze, G (2020) Cinema II: the time-image. In: Philosophers on Film from Bergson to Badiou. Columbia University Press, New York, USA. Chap. 9, pp 177–199
34. Bonebright TL (2012) Were those coconuts or horse hoofs? Visual context effects on identification and veracity of everyday sounds. Georgia Institute of Technology
35. Carello C, Anderson KL, Kunkler-Peck AJ (1998) Perception of object length by sound. Psychol Sci 9(3):211–214
36. Demany L, Semal C (2008) In: Yost WA, Popper AN, Fay RR (eds.) The role of memory in auditory perception, Springer, Boston, MA, pp 77–113
37. Langlois TR, James DL (2014) Inverse-Foley animation: synchronizing rigid-body motions to sound. ACM Trans Graphics (TOG) 33(4):1–11
38. Pastore RE, Flint JD, Gaston JR, Solomon MJ (2008) Auditory event perception: the source—perception loop for posture in human gait. Perception & Psychophys 70(1):13–29
39. Grassi M (2005) Do we hear size or sound? Balls dropped on plates. Perception & Psychophys 67(2):274–284
40. Li X, Logan RJ, Pastore RE (1991) Perception of acoustic source characteristics: walking sounds. J Acoustical Soc America 90(6):3036–3049
41. Pauletto S, Selfridge R, Holzapfel A, Frisk H (2021) From Foley professional practice to sonic interaction design: initial research conducted within the radio sound studio project. In: Nordic sound and music computing conference
42. De Götzen A, Sikström E, Grani F, Serafin S (2013) Real, Foley or synthetic? An evaluation of everyday walking sounds. Proceedings of SMC
43. Trento S, De Götzen A (2011) Foley sounds vs real sounds. In: Sound and music computing conference (SMC2011)
44. Nordahl R, Turchet L (2011) Serafin, S Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications. IEEE Trans Visualization Comput Graphics 17(9):1234–1244
45. Turchet L (2016) Footstep sounds synthesis: design, implementation, and evaluation of foot–floor interactions, surface materials, shoe types, and walkers' features. Appl Acoustics 107:46–68
46. Hughes B, Wakefield J (2015) An investigation into plausibility in the mixing of Foley sounds in film and television. In: Audio Engineering society convention 138. Audio Engineering Society
47. Zelechowska A, Gonzalez-Sanchez VE, Laeng B, Jensenius AR (2020) Headphones or speakers? An exploratory study of their effects on spontaneous body movement to rhythmic music. Front Psychol 11. https://doi.org/10.3389/fpsyg.2020.00698
48. Soulard J, Vaillant J, Balaguier R (2021) Vuillerme, N Spatio-temporal gait parameters obtained from foot-worn inertial sensors are reliable in healthy adults in single-and dual-task conditions. Scientific Reports 11(1):1–15
49. Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna Jr JM, Kang M, Rowe DA (2018) How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. British J Sports Med 52(12):776–788. https://bjsm.bmj.com/content/52/12/776.full.pdf. https://doi.org/10.1136/bjsports-2017-097628
50. Sikström E, De Götzen A, Serafin S (2015) Self-characterstics and sound in immersive virtual reality—estimating avatar weight from footstep sounds. In: 2015 IEEE virtual reality (VR), IEEE, pp 283–284
51. Tajadura-Jiménez A, Basia M, Deroy O, Fairhurst M, Marquardt N, Bianchi-Berthouze N (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, pp 2943–2952
52. Takashima M (2018) Perceived weight is affected by auditory pitch not loudness. Perception 47(12):1196–1199
53. Tajadura-Jiménez A, Newbold J, Zhang L, Rick P, Bianchi-Berthouze N (2019) As light as you aspire to be: changing body perception with sound to support physical activity. In: Proceedings of the 2019 CHI conference on human factors in computing systems, pp 1–14
54. Clausen S, Tajadura-Jiménez A, Janssen CP (2021) Bianchi-Berthouze, N Action sounds informing own body perception influence gender identity and social cognition. Front Hum Neurosci 15:688170
55. Giordano B, Bresin R (2006) Walking and playing: what's the origin of emotional expressiveness in music. In: Proc. Int. Conf. Music Perception and Cognition (2006)
56. Sudo K, Yamato J, Tomono A, Ishii K-i (2002) Gender recognition method based on silhouette, footstep, and foot pressure measurements for counting customers. Electron Commun Japan (Part II: Electronics) 85(8):54–64
57. Turchet L (2013) Serafin, S Investigating the amplitude of interactive footstep sounds and soundscape reproduction. Appl Acoustics 74(4):566–574
58. Visell Y, Fontana F, Giordano BL, Nordahl R, Serafin S (2009) Bresin, R Sound design and perception in walking interactions. Int J Human-Comput Stud 67(11):947–959
59. DeWitt A, Bresin R (2007) Sound design for affective interaction. In: Affective computing and intelligent interaction: second international conference, ACII 2007 Lisbon, Portugal, September 12-14, 2007 Proceedings 2, Springer, pp 523–533
60. Itai A, Yasukawa H (2006) Footstep recognition with psyco-acoustics parameter. In: APCCAS 2006-2006 IEEE Asia pacific conference on circuits and systems, IEEE, pp 992–995
61. Algermissen S (2021) Hörnlein, M Person identification by footstep sound using convolutional neural networks. Appl Mech 2(2):257–273
62. Cook PR (2002) Modeling bill's gait: analysis and parametric synthesis of walking sounds. In: Audio Engineering society conference: 22nd international conference: virtual, synthetic, and entertainment audio. audio engineering society
63. DeLoney C (2008) Person identification and gender recognition from footstep sound using modulation analysis. Technical Report
64. Geiger JT, Kneißl M, Schuller BW, Rigoll G (2014) Acoustic gait-based person identification using hidden Markov models. In: Proceedings of the 2014 workshop on mapping personality traits challenge and workshop, pp 25–30
65. Shoji Y, Takasuka T, Yasukawa H (2004) Personal identification using footstep detection. In: Proceedings of 2004 international symposium on intelligent signal processing and communication systems, 2004. ISPACS 2004., IEEE, pp 43–47
66. Altaf MUB, Butko T, Juang B-H (2015) Acoustic gaits: gait analysis with footstep sounds. IEEE Trans Biomed Eng 62(8):2001–2011
67. Jiang X, Pell MD (2014) Encoding and decoding confidence information in speech. In: Proceedings of the 7th International Conference in Speech Prosody (social and Linguistic Speech Prosody), vol. 5762579
68. Jiang X, Pell MD (2018) Predicting confidence and doubt in accented speakers: human perception and machine learning experiments. In: Proceedings of speech prosody, pp 269–273
69. Kirkland A, Lameris H, Székely E, Gustafson J (2022) Where's the uh, hesitation? The interplay between filled pause location, speech rate and fundamental frequency in perception of confidence. In: Proceedings of interspeech, pp 18–22
70. Goupil L, Ponsot E, Richardson D, Reyes G, Aucouturier J-J (2021) Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature. Nature Commun 12(1):861

71. Goupil L, Aucouturier J-J (2021) Distinct signatures of subjective confidence and objective accuracy in speech prosody. Cognition 212:104661
72. Lartillot O, Toiviainen P, Eerola T (2008) A Matlab toolbox for music information retrieval. In: Data Analysis, Machine Learning and Applications: Proceedings of the 31st Annual Conference of the Gesellschaft Für Klassifikation eV, Albert-Ludwigs-Universität Freiburg, March 7–9, 2007, Springer, pp 261–268
73. Zwicker E, Flottorp G, Stevens SS (1957) Critical band width in loudness summation. J Acoustical Soc America 29(5):548–557
74. Zwicker E (1961) Subdivision of the audible frequency range into critical bands (frequenzgruppen). J Acoustical Soc America 33(2):248–248
75. Foote J (1999) Visualizing music and audio using self-similarity. In: Proceedings of the seventh ACM international conference on multimedia (Part 1), pp 77–80
76. Foote J, Cooper M (2002) Nam. ISMIR, U Audio retrieval by rhythmic similarity. In: ISMIR
77. Lartillot O, Cereghetti D, Eliard K, Grandjean D (2013) A simple, high-yield method for assessing structural novelity. In: The 3rd international conference on music & emotion, Jyväskylä, Finland, June 11-15, 2013. University of Jyväskylä, Department of Music