Manipulating Foley Footsteps and Character Realism to Influence Audience Perceptions of a 3D Animated Walk Cycle

Stuart Cunningham
Manchester Metropolitan University
Manchester, UK
s.cunningham@mmu.ac.uk

Iain McGregor
Edinburgh Napier University
Edinburgh, UK
I.McGregor@napier.ac.uk

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 models were capable of communicating characteristics of age, gender, and weight. 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.

CCS CONCEPTS

• **Applied computing** → **Sound and music computing**; • **Human-centered computing** → User studies.

KEYWORDS

Sound design, Foley, user perceptions, animation, walk cycles

ACM Reference Format:
Stuart Cunningham and Iain McGregor. 2022. Manipulating Foley Footsteps and Character Realism to Influence Audience Perceptions of a 3D Animated Walk Cycle. In *Proceedings of Audio Mostly 2022 (AM '22).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

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.

AM '22, September 06–09, 2022, St. Pölten, Austria

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

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. Within animation, this is obviously not possible, but the intention is still the same in that every performed sound should belong to an on-screen character and their actions, 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.

Other researchers have examined how the *walking style* of virtual avatars may convey biological and personality features, such as gender and confidence [28]. 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*.

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 [2, 16, 34]. Three core sound elements of footsteps, prop handling and cloth are typically performed [4, 32, 33].

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*" [26].

Foley will often contribute to the drama of the scene and respond to the demands of the story. Foley work is a nuanced, performative activity and expressiveness is considered an important part of a Foley artist's experience [1].

Wright [33] 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*.

The sound produced by different surfaces and their interactions is a vital part of establishing the feel of a scene, location, or object in a scene, even if those surfaces and objects might never appear on-screen and are assumed to be present in the situation depicted. 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 [12].

Warren, Kim and Husney [31] 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.

Foley, along with other techniques, has also shown the potential to be used as a pedagogical tool in training sound designers, especially in the early stages where ideas are being explored and developed in prototype form [17].

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 [3] 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 [7]. 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 [11]. 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...*" [15].

Ennis, McDonnell and O'Sullivan [13] 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 [22] studied the effect of sound on the perceived smoothness of animations. Their experiment used an independent samples design; the dependent variable was perceived motion smoothness of the two animated sequences in each test pair. The independent variable was the auditory background of the clips (sound effects or silence). Test groups were separated by participants' familiarity with computer graphics. 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 [20] 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 the source rather than trying to describe the characteristics of the source sound. Lewis goes on to discuss the concept of a *crystalline image* [9] of a sound source 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. The scenes were played with either: the actual sound made by the object in the scene; an acoustically similar sound made by the object in the scene; or an acoustically dissimilar sound to the object in the scene. The results showed that a visual accompaniment to an auditory event has an effect on the listener's perception of the sound. Listeners accept a sound and its accuracy/appropriateness as that portrayed by the video [5].

One study explored the possibility of listeners determining physical attributes from objects they cannot see from 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 [6]. 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 [10].

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 [19].

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 relationship between these anthropomorphic and bio-mechanical properties, and the resulting postures and footsteps sounds were found to be complex and the ability of human listeners to correctly identify characteristics from the sound recording was variable [24]. 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 [14] 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. Further detail is given relating to these two objects: the non-sounding object (NSO) and the sounding object (SO). Physical properties of the NSO were reasonably estimated without any information given to participants about the sound

source event. Changes in the SO's properties affected the perceived properties of the NSO. Results showed that power (the rate of work over time) was the dominant factor and may be the link between actual and subjective size. Grassi's study showed 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 [1]. As such, prior to the term Foley *artist* being commonplace, Foley *walker* was used [2, 33]. 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"* [4]. 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 [21]. 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 [11].

Donaldson [12] 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, strength, and more. These are utilised to provide a sonic dramatization and may utilise sound 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 and 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 [25].

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 ratings, perceived realism, and perceived expressiveness. It was found that realism and expressiveness were generally much higher overall for Foley and real sounds when compared to those that had been synthesised, 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 [8].

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. Where a visual scene accompanied the audio, the clips were extracted from a set of four commercially released movies. 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 [29].

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 [23, 25]. 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 [23].

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. These significant manipulations related to changes in level, panning and equalisation. 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. Whilst the number of participants in this study was relatively small, their demographic was of students and academics in the music technology and multimedia field, and so can arguably be considered expert listeners, at least to a degree. 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 [18].

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 the 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 more accurately convey body movement through the improved visceral perception of low frequencies which can be impaired through ear canal only delivery [35]. All the animations were QuickTime videos contained within a PowerPoint presentation. 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.

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 inline with comfortable walking speeds found in empirical studies studies of gait [27], although higher than what may be considered a moderate intensity cadence of approximately 100 steps/min [30].

The walk cycle was applied to three different characters (see Figure 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) 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 seconds.

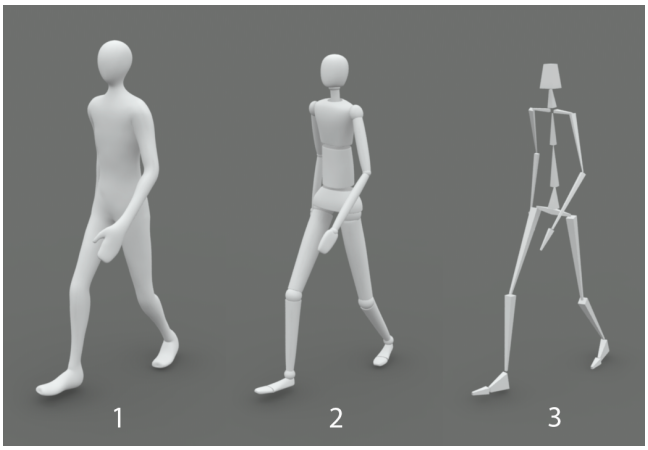

Figure 1: Animated characters: Skin (1), Mannequin (2) and Abstract (3)

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. 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); *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. 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 second 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 to 40 minutes, 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.

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 w/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 w/slight jitter

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 differences in sample size differences 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 Figure 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"* (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 ($M = 2.47, SD = 0.93$), 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 ($M = 2.25, SD = 0.81$).

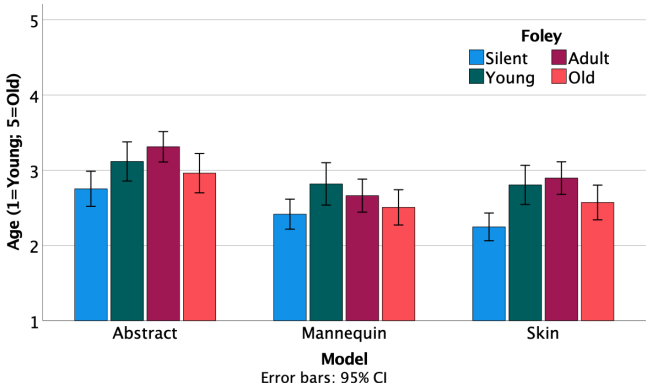

Figure 2: Participant Responses to Age (n=77) for Foley Condition and Model

The effects of the Foley condition (silent, young, adult, and old) 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 55s all gave average age estimates of under three, whilst the 55s and over were over three. In the under 55s group, average age estimates grew slightly from the 18 to 24 group ($M = 2.31, SD = 0.41$), through the 25 to 34s ($M = 2.66, SD = 0.41$), to the 35 to 44

group ($M = 2.76, SD = 0.27$), and then dipped slightly in the 45 to 54s ($M = 2.74, SD = 0.36$).

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 ($M = 2.88, SD = 1.34$).

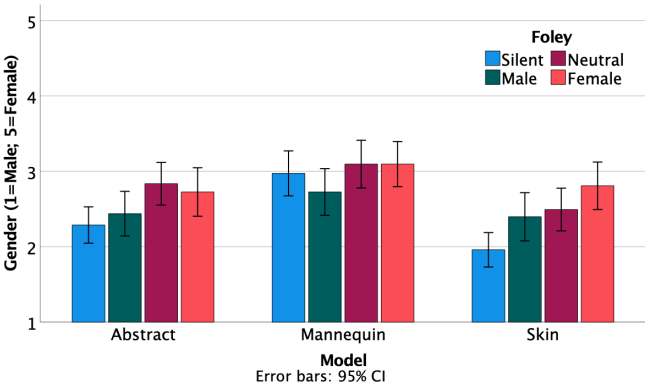

Figure 3: Participant Responses to Gender (n=73) for Foley Condition and Model

The effects of the Foley condition (silent, male, neutral, and female) 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 also 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 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).

4.3 Weight

Analysis of the participants' perceptions of weight (n=83) is illustrated in Figure 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 ($M = 3.98, SD = 0.88$).

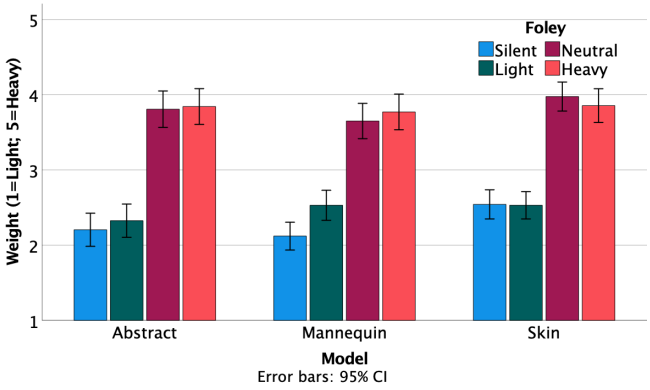

Figure 4: Participant Responses to Weight (n=83) for Foley Condition and Model

The effects of Foley condition (silent, light, neutral, and heavy) and model upon perceptions of weight were examined. There was no interaction between the Foley condition and model, but significant differences existed ($F(2.401, 195.885) = 148.086, p < .001$) between conditions, 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 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 Figure 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.

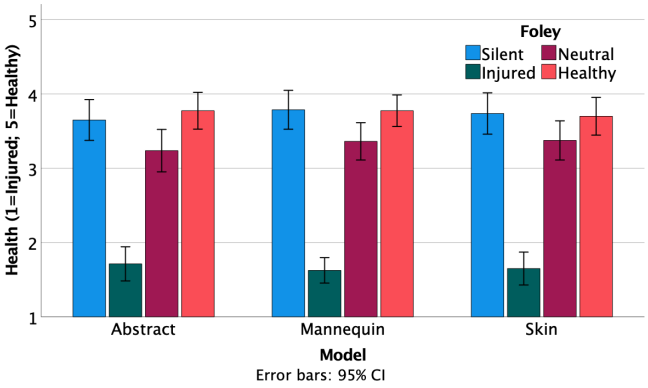

Figure 5: Participant Responses to Health (n=80) for Foley Condition and Model

The effects of the Foley condition (silent, injured, neutral, and healthy) 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 in the animation clips are presented in Figure 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 confident ($M = 2.20, SD = 0.97$) 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 effects of the Foley condition (silent, confident, neutral, and nervous) and model upon perceptions of confidence were examined.

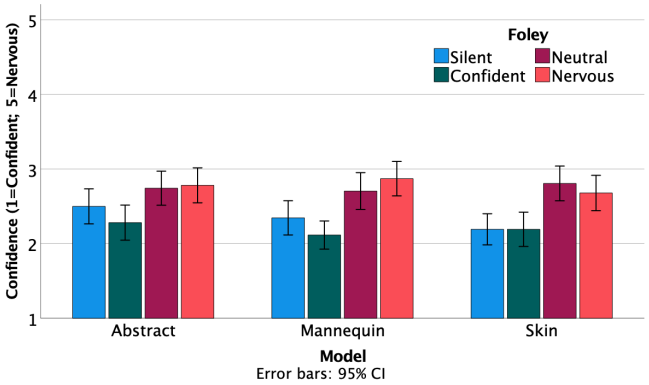

Figure 6: Participant Responses to Confidence (n=78) for Foley Condition and Model

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.

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 DISCUSSION

In general terms, the Foley created for this study could be considered successful in conveying all the characteristics (age, gender, weight, health, and confidence), 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.

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 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 the research presented here provides a framework for other researchers. A detailed acoustic analysis of the sounds created in this study and how acoustic features correlate with participant's perception of the five characteristics may also yield useful insight for practitioners.

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 might 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 any video capture or camera driven avatar to either make it more extreme or neutral in order to optimise anonymity within the metaverse.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Luis Aly, Rui Penha, and Gilberto Bernardes. 2017. Digit: A digital foley system to generate footstep sounds. In *International symposium on computer music multidisciplinary research*. Springer, 429–441.
[2] Vanessa Theme Ament. 2014. *The Foley grail: The art of performing sound for film, games, and animation*. Routledge.
[3] David B Anderson and Michael A Casey. 1997. The sound dimension. *IEEE spectrum* 34, 3 (1997), 46–50.
[4] Robin Beauchamp. 2013. *Designing Sound for Animation* (2 ed.). Routledge, New York, USA. https://doi.org/10.4324/9780240825007
[5] Terri L Bonebright. 2012. Were those coconuts or horse hoofs? visual context effects on identification and veracity of everyday sounds. Georgia Institute of Technology.
[6] Claudia Carello, Krista L Anderson, and Andrew J Kunkler-Peck. 1998. Perception of object length by sound. *Psychological science* 9, 3 (1998), 211–214.
[7] Michel Chion. 2019. Audio-vision: sound on screen. In *Audio-Vision: Sound on Screen*. Columbia University Press.
[8] Amalia De Götzen, Erik Sikström, Francesco Grani, and Stefania Serafin. 2013. Real, foley or synthetic? An evaluation of everyday walking sounds. *Proceedings of SMC* (2013).
[9] Gilles Deleuze. 2020. Cinema II: The Time-Image. In *Philosophers on Film from Bergson to Badiou*. Columbia University Press, Chapter 9, 177–199.
[10] Laurent Demany and Catherine Semal. 2008. *The Role of Memory in Auditory Perception*. Springer US, Boston, MA, 77–113. https://doi.org/10.1007/978-0-387-71305-2_4
[11] Lucy Fife Donaldson. 2014. The work of an invisible body: the contribution of foley artists to on-screen effort. *Alphaville: Journal of Film and Screen Media* 7 (2014), 1–15.
[12] Lucy Fife Donaldson. 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, 85–95.
[13] Cathy Ennis, Rachel McDonnell, and Carol O'Sullivan. 2010. Seeing is believing: body motion dominates in multisensory conversations. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 1–9.
[14] Massimo Grassi. 2005. Do we hear size or sound? Balls dropped on plates. *Perception & psychophysics* 67, 2 (2005), 274–284.
[15] Mack Hagood. 2014. Unpacking a punch: transduction and the sound of combat Foley in Fight Club. *Cinema Journal* 53, 4 (2014), 98–120.
[16] Tomlinson Holman. 2012. *Sound for film and television*. Routledge.
[17] Daniel Hug and Moritz Kemper. 2014. From foley to function: A pedagogical approach to sound design for novel interactions. *Journal of Sonic Studies* 6, 1 (2014), 1–23.
[18] Braham Hughes and Jonathan Wakefield. 2015. An investigation into plausibility in the mixing of Foley sounds in film and television. In *Audio Engineering Society Convention 138*. Audio Engineering Society.
[19] Timothy R Langlois and Doug L James. 2014. Inverse-foley animation: Synchronizing rigid-body motions to sound. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–11.
[20] Matt Lewis. 2015. Ventriloquial acts: Critical reflections on the art of Foley. *The New Soundtrack* 5, 2 (2015), 103–120.
[21] Xiaofeng Li, Robert J Logan, and Richard E Pastore. 1991. Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America* 90, 6 (1991), 3036–3049.
[22] Georgia Mastoropoulou, Kurt Debattista, Alan Chalmers, and Tom Troscianko. 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*. 9–15.
[23] Rolf Nordahl, Luca Turchet, and Stefania Serafin. 2011. Sound synthesis and evaluation of interactive footsteps and environmental sounds rendering for virtual reality applications. *IEEE transactions on visualization and computer graphics* 17, 9 (2011), 1234–1244.
[24] Richard E Pastore, Jesse D Flint, Jeremy R Gaston, and Matthew J Solomon. 2008. Auditory event perception: The source—perception loop for posture in human gait. *Perception & psychophysics* 70, 1 (2008), 13–29.
[25] Sandra Pauletto, Rod Selfridge, Andre Holzapfel, and Henrik Frisk. 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*.
[26] Sara Pinheiro. 2016. Acousmatic Foley: Staging sound-fiction. *Organised Sound* 21, 3 (2016), 242–248.
[27] J Soulard, J Vaillant, R Balaguier, and N Vuillerme. 2021. 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 (2021), 1–15.
[28] Anne Thaler, Andreas Bieg, Naureen Mahmood, Michael J. Black, Betty J. Mohler, and Nikolaus F. Troje. 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)*. 678–679. https://doi.org/10.1109/VRW50115.2020.00190
[29] Stefano Trento and Amalia De Götzen. 2011. Foley Sounds vs Real Sounds. In *Sound and Music Computing Conference (SMC2011)*.
[30] Catrine Tudor-Locke, Ho Han, Elroy J Aguiar, Tiago V Barreira, John M Schuna Jr, Minsoo Kang, and David A Rowe. 2018. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *British Journal of Sports Medicine* 52, 12 (2018), 776–788. https://doi.org/10.1136/bjsports-2017-097628 arXiv:https://bjsm.bmj.com/content/52/12/776.full.pdf
[31] William H Warren Jr, Elizabeth E Kim, and Robin Husney. 1987. The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. *Perception* 16, 3 (1987), 309–336.
[32] Patrick Winters. 2017. *Sound Design for Low and No Budget Films*. Routledge.
[33] Benjamin Wright. 2014. Footsteps with character: the art and craft of Foley. *Screen* 55, 2 (2014), 204–220.
[34] David Lewis Yewdall. 2012. *The practical art of motion picture sound*. Routledge.
[35] Agata Zelechowska, Victor E. Gonzalez-Sanchez, Bruno Laeng, and Alexander Refsum Jensenius. 2020. Headphones or Speakers? An Exploratory Study of Their Effects on Spontaneous Body Movement to Rhythmic Music. *Frontiers in Psychology* 11 (2020). https://doi.org/10.3389/fpsyg.2020.00698