The Augmented Floor - Assessing Auditory Augmentation

Katharina Groß-Vogt, Iason Svoronos Kanavas, Marian Weger
vogt@iem.at
Institute of Electronic Music and Acoustics (IEM), University of Music and Performing Arts Graz
Graz, Austria

Clemens Amon
Joanneum Research - DIGITAL
Graz, Austria
Clemens.Amon@joanneum.at

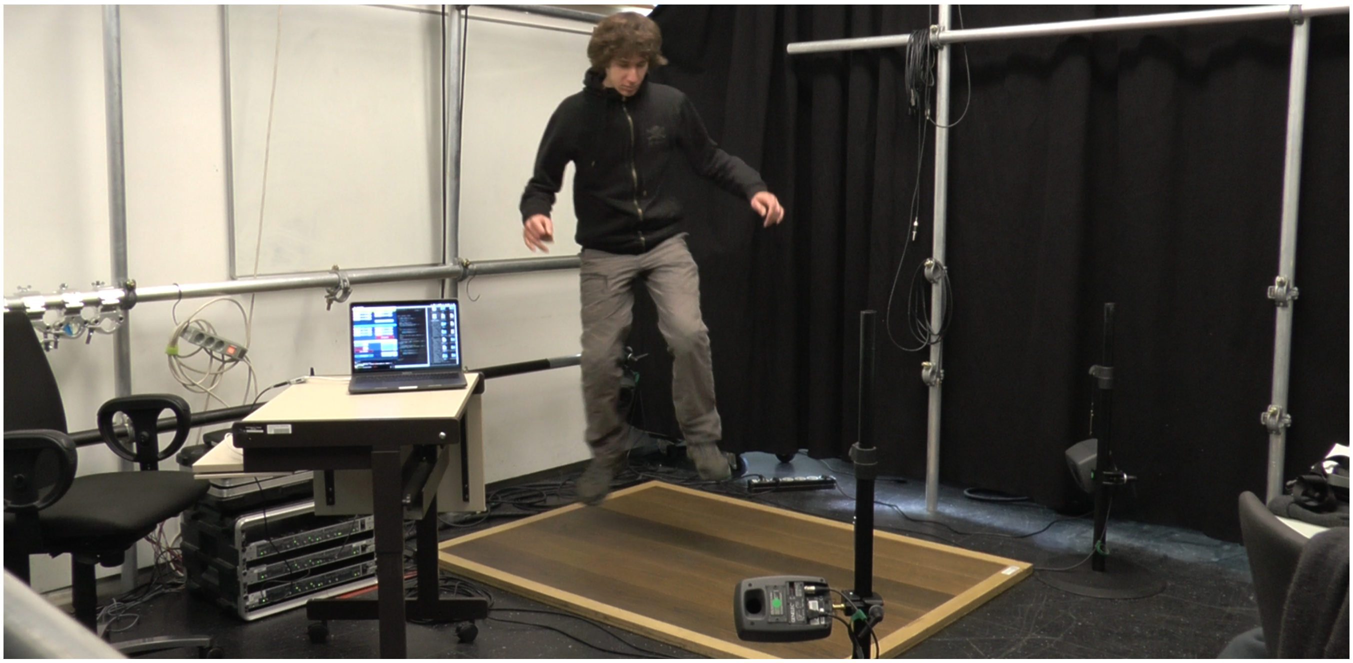


Figure 1: One of the authors jumping on the Augmented Floor in the Sonic Interaction Design Lab of the IEM.

ABSTRACT

The Augmented Floor is a wooden floor with in-built PyzoFlex® printed electronics. It measures any change of pressure on its surface. For users walking on this floor, we created an auditory augmentation of their footsteps. Different sound designs were implemented, both realistic and synthetic ones. We developed a method based on envelope reusing and compared it to the triggering of step-sounds. The envelope method shows a higher correlation between the input signal and the playback sound; it was rated better. Besides, realistic sounds were preferred over synthetic ones.

AM '22, September 6–9, 2022, St. Pölten, Austria

ACM ISBN 978-1-4503-9701-8/22/09…$15.00
https://doi.org/10.1145/3561212.3561219

KEYWORDS

sonification, auditory augmentation, onset detection, interactive sonification

ACM Reference Format:
Katharina Groß-Vogt, Iason Svoronos Kanavas, Marian Weger and Clemens Amon. 2022. The Augmented Floor - Assessing Auditory Augmentation. In *AudioMostly 2022 (AM '22), September 6–9, 2022, St. Pölten, Austria.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3561212.3561219

1 INTRODUCTION

The concept of auditory augmentation has been elaborated by Bovermann et al. [3] with their exemplary prototype of the WetterReim. In this work, the typing sound on a keyboard was recorded with a Piezo microphone, and played back in real-time with added resonances which depend on external data - in that case, weather data. The typing sound groups in the perception with the added resonances to sound differently depending on the weather outside. Bovermann et al. defined auditory augmentation as *"a paradigm to vary the objects' sonic characteristics such that their original sonic response appears as augmented by an artificial sound"*. Later work of part of the authors broadened this definition to *"the augmentation of* *a physical object and/or its sound by sound [...]"* [7]. In this definition, auditory augmentation makes use of either object-specific sounds caused by interaction, or adds environmental sounds that blend into the sound scene. An exemplary installation of this work used artificial reverberation in a kitchen to convey its current use of electricity; in the case of extra-ordinary usage of electric energy, the kitchen sounded like a church [8].

Various related concepts to the one of auditory augmentation have been developed in different fields. To mention a few, Pousman and Stasko [14] review prototypes of *ambient information design* that provide information without distracting the users in a pleasant way, via different modalities; the authors define the term and create a taxonomy for such systems. In the concept of *ambient sonification systems*, actually quite similar prototypes have been developed to the ones cited above (for instance, the opening of a wardrobe door making sounds that refer to the actual weather outside) [6]; mentioning one more related concept, *embedded sonification* explores mobile sonifications, i.e. where the sonification is embedded in *things* [2]. In the latter paper, prototypes of hand-size have been developed for the scenario of a phone being dropping into the lake and how to circumvent such a risk.

The ultimate goal of auditory augmentation is to use the current, ambient sounds and to provide additional, external, information - in the way this happens in *augmented reality*. This can be achieved without disturbing the real surrounding too much, so the inherent assumption is that the added sounds will blend into the real world and do not hinder the original interaction of the user. In previous works, the additional value of auditory augmentation has not been studied systematically. This paper describes an effort to start with such research.

Our test environment was based on a wooden floor that is equipped with printed electronics, a technology called PyzoFlex® [15], by courtesy of Joanneum Research Digital. It is based on ferroelectric polymers that are printed on a thin foil and glued between the wooden layers of the floor (in this case). These sensors detect any change of pressure as excerted when, e.g., walking on the floor.

Related Projects

We first conducted a literature review on projects concerning either walking interactions or interactive floors, both in combination with sound. Applied research was performed about using interactive floors to support specific goals. One example is building an automated dance recognition tool for training the correct step sequence in standard dances [13]. A very different goal is the automatic tracking of people with a smart floor [1], or, more specifically, surveying elderly people who have a high risk of falling on the floor [5]. The latter was also the initial use case scenario for which the PyzoFlex® floor has been developed. In more fundamental research, one insightful project showed that altering walking sounds of people made them perceive themselves as thinner or even happier when, for instance, their walking sounded more dynamically [17]. Another study focused on the rhythm of walking [11]. Different auditory and haptic feedback and interaction modes were studied with respect to their effect on the walking tempo. All of the above systems differ in their technical realization from our PyzoFlex® floor, some of them used smart floors equipped with different sensors, others equipment on the shoes.

We glean from these works that there appears to be a closed loop in perception between walking and step-sounds, just as can be expected from research on embodiment (as a starting point to the literature on embodiment and HCI we refer to the special issue of [12]).

Hypothesis and Procedure

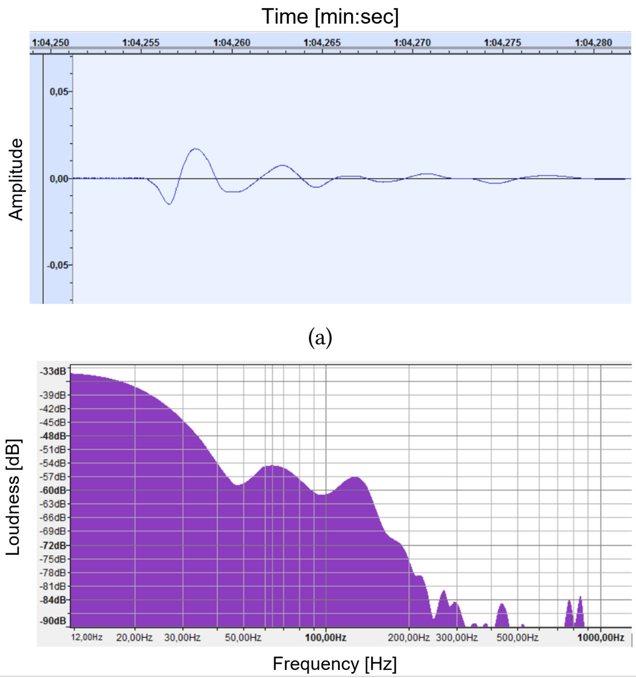
In the Augmented Floor project, we aimed at evaluating possible advantages of auditory augmentation, with the following hypothesis: the more precise a raw input signal of an interaction is followed by the sound of the auditory augmentation, the better the perception of the interaction is assessed. We implemented two different algorithms with varying degrees of preserving information of the original input signal in the playback sound. The first we called *"envelope reusing"* or *"env"*; it is based on using the envelope of the real steps in real-time as envelope for the playback sound, which is based on a sound sample with a rather steady spectrum. The second algorithm is a much more classic approach. We pre-render step-sounds in various surroundings and trigger them following an onset detection of the walking person - we refer to it as *"step-sound triggering"* or *"trig"*. Both algorithms realize auditory augmentation, but we argue that they do so at different levels of auditory correspondence between the ambient sound (the steps on the wooden floor) and an added sound (for instance, walking on gravel or in a spaceship). All details of the setup and software implementation will be discussed in Sec. 2.

In an experiment (see Sec. 3) we tested if subjects prefer the same step-sounds in the *"env"* condition over the *"trig"* condition. Secondly, we wanted to examine different sounds: on the one hand, realistic step-sounds in different conditions (such as walking on a gritted street, on gravel, snow, or wading in water), on the other hand synthetic sounds (such as walking on a "gong" or sci-fi sounds). Would people prefer synthetic sounds when missing the correct clues of other modalities (wading in water is not realistic when standing on a wooden floor)? Or do people prefer realistic step-sounds, because they are less salient and fit better to their experience? We both collected quantitative data (rating of preference on a slider interface, indirect measure of "time-on-exhibit" - in this case the time used to rate a certain preset), and qualitative data in a post-hoc questionnaire.

A demo video can be found at https://phaidra.kug.ac.at/o:127081.

2 PROTOTYPE REALIZATION

We first considered synthesizing step-sounds. Extensive research has been reported in this field (for an overview, we refer to [18] and the seminal PhISM algorithm of Cook [4]). Tests with our given technical system revealed that the recorded signal is difficult to work with if used directly as step-sound. On the one hand, the envelope of the system follows any pressure impact very quickly and precisely. This can be observed in our calibration setup. We let a sand ball fall from a defined height on 18 positions of the floor, where one of the sensors is located. The system response can be seen in Fig. 2(a) as tiny but accurate signal. On the other hand, the raw signal sounds dull because it lacks high frequencies.



(a)

(b)

Figure 2: Examining the raw signal from the Augmented Floor: (a) shows the impact of a calibration event (a sand ball with diameter of 7 cm falling from 1 m height roughly at the position of one sensor). The system response is weak (0,02 from a normed amplitude of 1), but accurate, and as short as 25 ms. (b) shows the spectrum of the aggregated signal of one step on the floor (all 18 channels mixed, one of the authors stepping on the floor). The signal is very silent and has a huge noise content. Besides, what is of interest here, is the fact that spectral content is missing in the central hearing range and completely above ca. 1000 Hz.

This is illustrated in Fig. 2(b), where the spectrum of one step of one of the authors is shown. Middle and higher frequencies are missing. So, augmenting the signal itself by, for instance, adding of resonances, would not lead to a high-quality step-sound. Due to the system’s properties, we decided to use only the signal’s envelope for the auditory augmentation. This allows for a very quick system response and low latency and is discussed in Sec. 2.2 and 2.3.

2.1 Hardware setup

The PyzoFlex® floor is roughly 2.2 square meters large, and is surrounded by a ring of four loudspeakers with a diameter of ca. 2 meters. The space of the floor is thus rather limited for real walking, but turned out to be sufficient for our purpose. The floor consists of 6 wooden floor tiles each of them equipped with 3 single PyzoFlex® sensors [20] printed on 175 μm thin PET substrate. Each sensor has a size of 8 times 4 centimeters. The sensors are directly integrated into the wooden floor tiles for maximum mechanical coupling. This ensures a high sensitivity to structure-borne sound induced by physical interaction on its surface. Around the floor, four Genelec 8020 loudspeakers were placed on each corner, tilted downwards in order to use their first reflections to enhance the feeling that the sound comes from the floor itself as much as possible.

We used TouchOSC[1] to collect data in the experiment over a smartphone and send them to the computer via WLan using the OSC protocol.

The main software for the auditory augmentation was written in SuperCollider. In the following subsections, the two different algorithms are explained in detail.

2.2 (I) Envelope reusing (*env*)

As preparation, we recorded or generated sound samples with a rather steady spectrum; for instance, in the case of gravel as walking surface, we have recorded a file of dragging an object over gravel; in the case of wading in water this was the sound of rushing water. A sample player continuously looped the sound sample with - initially - zero amplitude. In the further course, the amplitude of the signal was controlled by a step’s envelope and thus created a high-quality step sound for various surroundings.

The DSP for the envelope reusing is shown in Fig. 3. First, we wanted to filter out internal low frequency noise produced by the floor’s electronics. Therefore, each signal from the 18 channels was filtered with a highpass filter at a variable cutoff frequency for the different playback sounds (between 220 Hz and 440 Hz). Then, the signals were half-wave rectified and the input signal envelopes were calculated by utilizing an envelope follower. Following that, the 18 signals were added and amplified. The now mono signal was smoothed by a filter. Using a mix of both the smoothed and non-smoothed signal further on provided better listening experience, establishing a compromise between quick reactions of high transients but still smoothing out every detail of the highly sensitive floor. Then, a steep-slope expander was configured for each playback sound separately. The threshold of the expander was set at the highest noise level in the signal; anything above was leveled up in order to improve the SNR. Furthermore, the real envelop of the input signal is very short, while real footsteps are longer due to reverberation issues. Therefore, we used another envelope follower with an instant attack time but a configurable smoothed release. Whenever the signal level was above the threshold of the expander, the amplitude of the playback synthesizer was modulated by the input signal’s smoothed envelope resulting in an audible sound. As next step, the output signal was filtered by a lowpass filter followed by a compressor so we could level up very soft footsteps that were inaudible. In the last stage, we utilized the summed enveloped and we linearly mapped it to the playback rate of the sound loop, as well as to frequency values in order to modulate the low pass filter’s cutoff frequency values. This was to create a wide variety of different sound textures controlled by the input signal as an additional interaction feature. Furthermore, it mimics the physical reality of surfaces reacting differently to soft and heavy impacts, when different resonances come into play.

The *env* algorithm has several advantages. The input signal’s waveform is derived almost instantaneously, latency only arising from input/output buffering, AD/DA conversion, and very little delays caused by basic DSP such as filtering. Another advantage of

[1] TouchOSC by hexler, https://hexler.net/touchosc.

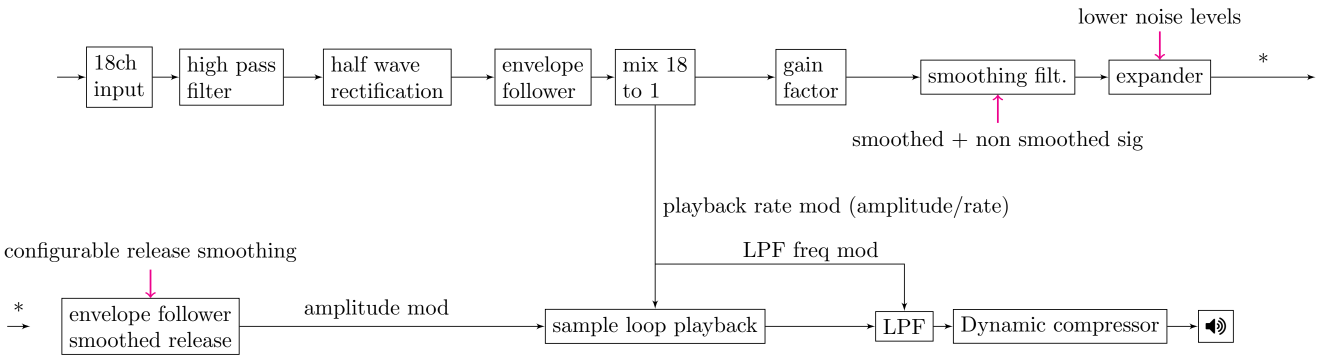


Figure 3: Block diagram of the DSP for envelope augmentation.

this rather simplistic method is that it provides reliable results with little possibility for error. In particular, using a simple threshold level, by means of a noise gate or an expander with a steep slope, we can easily prevent hardware noise to be blended significantly with the actual input signal. Overall, the augmentation is very fast and directly follows any modulations of the step signal.

2.3 (II) Step-sound triggering (*trig*)

A more classical approach in sonic interaction design than the *env* algorithm is the triggering of sound samples whenever a step has been detected; which we refer to as *trig* algorithm. As preparation, we created a variety of step-sound samples for each walking surface, that corresponded to the ones used with the *env* algorithm.

Effective, real-time onset detection for interactive applications is challenging due to the need for low latency. Kemp [10] showed that in case of an auditory stimulus, a latency of 8-10 ms was perceivable, and more recently, according to Jack et al. [9], a latency of sound with respect to a haptic stimulus is imperceptible if it stays below 10 ms. Therefore, as a rule of thumb, the overall time of the detection of the step and playback to the listener's position should remain below 10 ms. The main additional delay of our system as compared to the above described *env* algorithm stems from the real-time onset detection.

A large variety of onset detection functions are based on DFT processing. In our case, the minimum delay we could achieve by utilising onset detection based on DFT was 5.3 ms with a sampling rate of 48 kHz and a window size of 256. On top of that, we add additional delay times due to other DSP processes such as noise filtering and hardware buffering. The whole algorithm did not meet the needs of the desired latency threshold. Another issue with DFT-based algorithms is related to computing resources. Practice shows that intensive DFT onset detection algorithms can be detrimental to the overall performance, with additional delays added because of buffer underrun issues. This is typically a complex problem since it is related to low level processes, so the question lies whether developing effective algorithms as alternatives to DFT-based onset detection could be the solution. For the above reasons, we used an alternative to DFT-based onset detection based on amplitude thresholding as has been recently developed by Weger et al. [19]. Please refer to this paper for details on the algorithm.

In addition to the onset detection, the following steps of DSP have been implemented. First, we calculated the mean value of all 18 channels in order to prevent false onset detection triggers as well as extremely high amplitude levels. The latter were observed whenever users stepped directly on top of the individual sensors. Furthermore, to achieve a better signal to noise ratio (SNR), we applied a noise gate with a threshold level corresponding to the noise level. Finally, we tried to enhance the correspondence between walking interaction and sound even more. For this goal, we both mapped amplitude values of the input signal to a cutoff frequency of a low pass filter, as well as to the playback rate values (between 0.7 and 1.3 where 1.0 was normal playback speed). Finally, we mapped amplitude of the input signal to the playback sound.

With these adjustments and a roundabout time of below 10 ms, the *trig* algorithm performed better than we had initially suspected and, just as the *env* algorithm, created a highly corresponding experience of auditory augmentation.

2.4 Sound design

We chose and created different sound samples, out of which eight were to be used in the subsequent experiment, see Tab. 1. The criteria to choose them were (a) to provide a very good sound quality, and (b) to be consistent with the physicality of a step. The latter requirement was especially important for the synthetic sounds, as a way to fit them to the acoustic response of a step. A real step sound is a rather short sound event with one loud attack in the beginning (or possibly a second, much smaller one, reflecting the landing of the heel followed by the ball of the foot) and varying reverberation depending on the surface. As further criteria worth examining we defined (c) to provide varieties of both realistic and synthetic step-sounds, and (d) varying in terms of their spectral range – in order to explore different frequency regimes.

The type of chosen step-sound was restricted by the DSP methods described above. Especially walking on a hard surface with a very dry acoustics is not possible to realize with neither the *env* method – because it was not possible to find a good "homogeneous" sound to work with–, nor with the *trig* method –because the latency of the system would be even more problematic. Therefore, we ended up with mostly aggregated surfaces, such as gravel or icy snow, or with sounds containing hard attacks in reverberant

sound ID	metaphor	type
1	"gritted street"	realistic
2	"giant steps"	super-realistic
3	"phaser"	synthetic
4	"out of engine"	synthetic
5	"icy snow"	realistic
6	"wading in water"	realistic
7	"gong"	synthetic
8	"gravel"	realistic

Table 1: Sounds used in the experiment.

surfaces (e.g., a "gong" sound with, for a step-sound, especially long release time).

For both DSP methods, we created two versions of these samples that should lead to a very similar listening experience when walking on the floor. This should ensure that participants rated the inherent features of the responsiveness of the system rather than the sound quality or their personal preference of the sound.

3 EXPERIMENT

3.1 Experiment design

The experiment took place at the Institute of Electronic Music and Acoustics (IEM) in Graz, Austria, in February 2022. 14 participants took part in the experiment. They were recruited out of the staff, students, and friends of the authors; the group was rather heterogeneous in the sense that roughly half of them had no prior knowledge with interactive auditory systems.

The experiment included a short written introduction and the signing of the data privacy statement; then, in a learning phase, all sounds of the experiment were played in random order for only five seconds each, in total taking 80 seconds of getting to know the whole range of sounds. This phase could be repeated upon request.[2] The actual experiment then played all 16 sounds in randomized order for each participant, and for every sound people could take their time for rating, while walking on the floor. Thus, each participant heard all eight sounds in both variants, *env* and *trig*. After the experiment, a questionnaire had to be filled out.

In the test, participants were asked to rate their overall impression of the step-sounds – this rating should include both sound quality issues as well as the responsiveness of the system – from very poor to excellent. As an interface, a smartphone was used with TouchOSC that contained a simple slider, see Fig. 4. For each sound, a rating had to be given before the next sound could be started.

The post-hoc questionnaire comprised the following questions:

- on the criterion that the participant had elaborated for him-/herself according to which s/he rated the sounds (asked as an open question);
- on the perception and description of differences of the responsiveness of the system;
- on the sounds that could be remembered post-hoc;
- if and which differences in one's own movements were observed;

[2]This part of the experiment is also shown at the end of the demo video, see the link within the Procedure section.

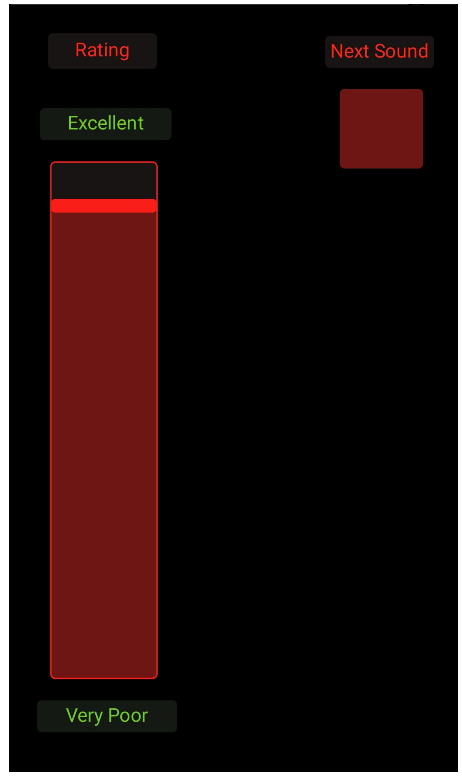


Figure 4: Screenshot of the smartphone's experiment interface.

- on the affective reaction of the participant following the SAM manikin rating for valence, arousal, and dominance of the interactive system [16];
- basic questions on the person and comments.

3.2 Results

Quantitative results have been collected and different conditions were compared to each other. Visual inspection already indicated that the rating's data are not normally distributed, as many high ratings were given, see Fig. 5. This is supported by a Shapiro-Wilk Test (stat=0.948, p=0.000). Therefore, we used the Mann-Whitney U test as a nonparametric statistical test for determining whether two different sub-categories of the data differed significantly with a significance level of $\alpha = 0.05$. All statistical results are given in Tab. 2.

Fig. 6 shows the boxplots of various conditions. In most but for two sounds (ID 2 and 4), the *env* method was rated higher than the *trig* one, though differences are not big – on average over all sounds, the difference is 0.15. For three sounds (ID 3, 5, and 7) these differences are significant. When pooling over all *env* sounds as compared to all *trig* ones, the difference is significantly higher, with a median value of 0.75 vs. 0.6. Furthermore, participants significantly evaluated the realistic sounds better than the synthetic ones, with respective median values of 0.7 vs. 0.63.

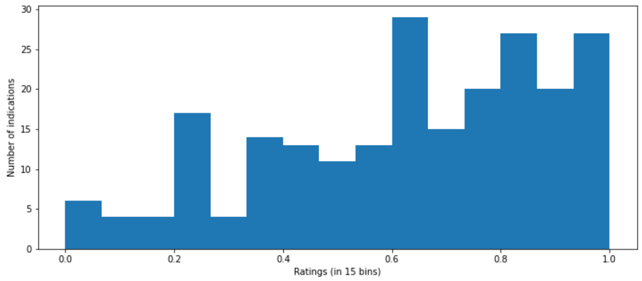


Figure 5: Histogram over all rating's data, sorted in 15 bins, from 0.0 (*very poor*) ot 1.0 (*excellent*).

data A	med.(Q1, Q3)	data B	med.(Q1, Q3)	statistics p-value
all-trig	0.6(0.35, 0.75)	**all-env**	0.75(0.55, 0.9)	**stat=4601 p=0.001**
1-trig	0.58(0.5, 0.72)	1-env	0.78(0.59, 0.85)	stat=66 p=0.147
2-trig	0.83(0.44, 0.9)	2-env	0.6(0.31, 0.85)	stat=121 p=0.284
3-trig	0.58(0.3, 0.65)	**3-env**	0.75(0.66, 0.85)	**stat=48.5 p=0.023**
4-trig	0.58(0.43, 0.73)	4-env	0.55(0.35, 0.73)	stat=103 p=0.818
5-trig	0.53(0.32, 0.7)	**5-env**	0.85(0.71, 0.94)	**stat=44.5 p=0.015**
6-trig	0.75(0.5, 0.83)	6-env	0.85(0.6, 1.0)	stat=76 p=0.318
7-trig	0.35(0.11, 0.54)	**7-env**	0.73(0.61, 0.8)	**stat=35 p=0.004**
8-trig	0.65(0.53, 0.75)	8-env	0.8(0.51, 0.95)	stat=69.5 p=0.197
S-trig	0.5(0.31, 0.69)	**S-env**	0.7(0.51, 0.8)	**stat=537 p=0.002**
R-trig	0.65(0.41, 0.8)	**R-env**	0.8(0.55, 0.9)	**stat=1920 p=0.027**
S-all	0.63(0.39, 0.75)	**R-all**	0.7(0.49, 0.9)	**stat=4639 p=0.008**

Table 2: Results of the experiment, each line compares different conditions using the Mann-Whitney U test with significance level $\alpha = 0.05$. Abbreviations stand for: trig..*trig* method; env..*env* method; S..synthetic sounds; R..realistic sounds; 1-8..sound samples (presets) with ID 1-8 as referred to in Tab.1. Significant differences are plotted in bold face, i.e. where $p < 0.05$.

As for the qualitative results from the questionnaire, many details support the quantitative results and make their interpretation clearer. When asked about the criteria they used to assess the sounds, most participants responded *responsiveness* (which was mentioned 9 times in a free verbalization) and/ or the *dynamic response* (6 times). Further, the *perceived immersion, realism, and response time* where called three times each. Other reported criteria were various sound features, such as *the sound itself, its quality, clarity or texture.* Also *plausibility* was mentioned once. The large majority stated that they had perceived differences in the responsiveness of the floor (13 "yes", 1 "no"). All of these 13 described the differences correctly, focusing on the responsiveness in terms of changing *dynamics*, *sensitivity*, or *latency*. A few times, different response behavior of the floor depending on the region (middle/ borders) was mentioned as well, which was a real but minimal effect of the system and had therefore been considered negligible by the developers.

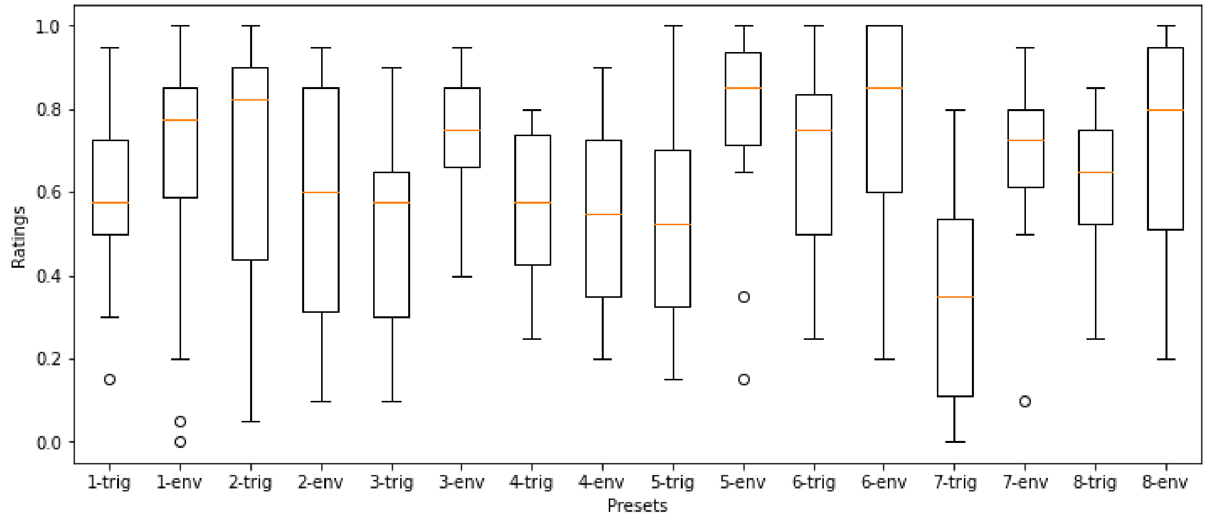
We asked which sound people could remember, first as an open verbalization, followed by a list that participants could choose from and comment on. Clearly, water was the sound best remembered, mentioned even 13 times in the free enumeration part. This was followed by mostly categories of sounds, as probably the other individual sounds were not so clearly recognized. Named categories were for instance *synthesized sounds* (8 times), *snow & ice* (7), or *gravel* (6). Twice, sounds were described that could not directly be related to any of the stimuli (for instance, *"giant steps"* and *"walking on a large tambour"* could mean the same sound but do not have to; the later could also refer to the *gong* sound). When the list was given to choose from, results were similar, with realistic sounds being most often remembered, again water outstanding in number of remembrances.

The SAM manikin rating revealed also very uniform answers for the three criteria, see Fig. 7: for valence (range: very angry=-2 to very joyful=+2) most participants chose *joyful* (mean 1), for arousal (very tired=-2 to very exciting=+2) the mean was 0.8, which corresponds closely to *excited*, and for the dominance (very controlled=-2 to very dominating=+2) the mean was 0.6 which could be interpreted as *rather dominating*.

3.3 Discussion

The quantitative data show that our hypothesis is unrefuted: there seems to be a perceivable difference in sonic interaction when the played sound is linked closer to details of the input signal. This was true in general over all sounds, even if the difference in rating was not very big. We argue that the *trig* condition was implemented so thoroughly, with little latency, and the sound samples so carefully mimicking real steps, that even the developers were not always sure themselves which condition was playing. Still, the differences show that an even higher degree of correspondence between the interaction and the sound has measurable effects.

When considering single sounds, only for three sounds the differences in rating were significant, ID 3 ("phaser"), ID 5 ("icy snow"), and ID 7 ("gong"). We do not see obvious similarities between these three. It may be argued that it is a matter of number of participants or number of trials to get statistically significant results for the other pre-sets. We also see a tendency, for only two sounds, that the onset triggering was rated better: ID 2 ("giant steps") and ID 4 ("out of engine"). Both of these sounds were low-frequency and very reverberant, therefore the subtle pressure changes of the steps did not make much of an audible difference. For these kinds of sounds there is no perceivable difference between our two algorithms, and this leads to our concluding remark: it may be argued that the reason for only three significantly different comparisons lies in their



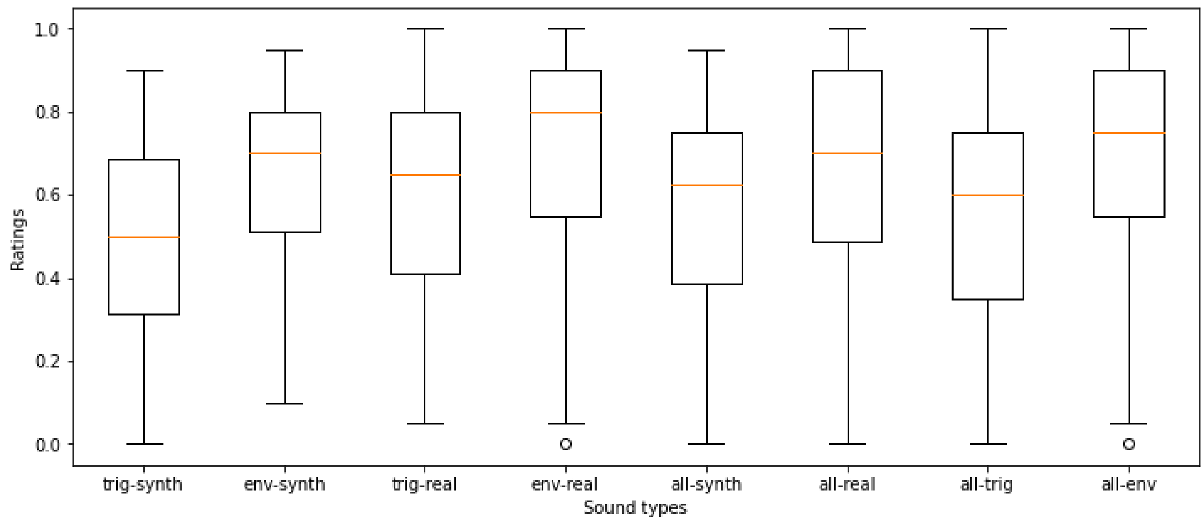


Figure 6: (a) Boxplot of ratings for all eight sounds in two conditions (trig, env), i.e. 16 presets, on a scale from 0/"very poor" to 1/"excellent". For three sounds, there was a significant difference, (3, 5, and 7). (b) compares different conditions and shows that, overall, *env* is rated better than *trig*, and realistic sounds (real), better than synthetic ones (synth).

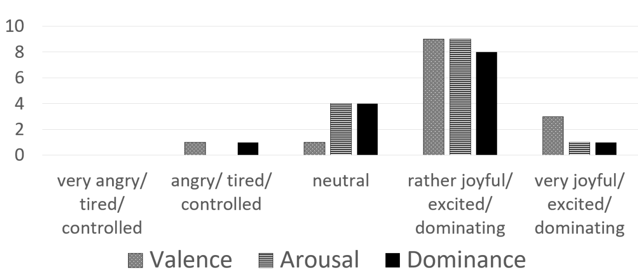


Figure 7: Histogram of the number of ratings concerning the affective reaction towards the Augmented Floor following the SAM criteria [16].

sound design. This factor is hard to systematize – further research could incorporate more detailed psycho-acoustic modeling, which was out of the scope of this work.

As a side finding, it was interesting to us that realistic sounds were rated higher than the synthetic ones; it might be because we are more used to real step-sounds and perceive more differences there. Admittedly, when working with synthetic sounds, matters of taste come into play. The participants may just not have liked the synthetic sounds we chose. A separate experiment rating the sounds without interacting, as a pure listening test, would be necessary to understand the underlying motifs . Furthermore, some sounds were rated better than others, either because they could be recognized more easily and/ or because they fit well to both implemented algorithms. ID 7 ("Water") and ID 6 ("Icy snow") were rated best overall. Also the qualitative comments supported this. Water is a special case, because the sound of wading into a basin with water produces a sound when the water surface is hit – rather than the ground of the basin. Therefore timing is less sensitive in this case. Partly, this might also apply for the icy snow, and might be a reason for the high ratings.

From the SAM rating, we can conclude that experiences with the system were similar amongst participants, and rated rather positively, as joyful, exciting, and participants felt rather dominating (rather than being controlled by the floor's sounds).

4 CONCLUSION

We presented a prototypic system that allows to play step-sounds in real-time following people walking on the Augmented Floor. The step-sounds were rendered starting from the pressure signal of the integrated sensors of the floor, with two different DSP processes, evaluating two methods of auditory augmentation. One was a carefully implemented onset detection of steps, followed by the triggering of sound samples. The other used more direct information of the original signal, as its envelope was as used an envelope for the playback sound.

The presented work was meant to assess auditory augmentation, and achieved to find differences in the perception of the two methods. The better the interaction was coupled to the sound, even on a very detailed level, the better the sonic interaction was rated. As auditory augmentation is a method where the playback sound always stays close to the original signal, we may conclude that the method in general is a beneficial way of sonic interaction.

The sound design for the 8 chosen pre-sets was rather overarticulated. In a real-world implementation of an Augmented Floor we would use the sounds much more subtly, as was also suggested by comments of the participants. The implementation of the software, especially the envelope reusing, can be used in cases where no hi-fi signal is available.

Even if about half of the participants were not sound experts, we were astonished how clearly the differences in our two DSP processes have been described. We conclude that the perception of the closed interaction loop is a highly developed, subconscious virtue that auditory augmentation can make use of.

ABBREVIATIONS

DFT	Discrete Fourier transform
DSP	Digital signal processing
env	envelope reusing (DSP method II)
IEM	Institute for Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria
OSC	Open sound protocol
PET	Polyethylene terephthalate
PhISM	Physically Informed Sonic Modeling
SAM	Self-assessment manikin
SNR	Signal-to-noise ratio
trig	step-sound triggering (DSP method I)

ACKNOWLEDGMENTS

Thanks to all participants of the experiment, and to Robert Höldrich for fruitful feedback.

REFERENCES

[1] Ibrahim Al-Naimi and Chi Biu Wong. 2017. Indoor human detection and tracking using advanced smart floor. In *2017 8th International Conference on Information and Communication Systems (ICICS)*. IEEE, 34–39.

[2] Stephen Barrass and Tim Barrass. 2013. Embedding sonifications in things. In *Proc. of the International Conference on Auditory Display (ICAD)*. Proc. of the ICAD.

[3] Till Bovermann, René Tünnermann, and Thomas Hermann. 2012. Auditory augmentation. In *Innovative Applications of Ambient Intelligence: Advances in Smart Systems*. IGI Global, 98–112.

[4] Perry R Cook. 1997. Physically informed sonic modeling (phism): Synthesis of percussive sounds. *Computer Music Journal* 21, 3 (1997), 38–49.

[5] Guodong Feng, Jiechao Mai, Zhen Ban, Xuemei Guo, and Guoli Wang. 2016. Floor pressure imaging for fall detection with fiber-optic sensors. *IEEE Pervasive Computing* 15, 2 (2016), 40–47.

[6] Sam Ferguson. 2013. Sonifying every day: Activating everyday interactions for ambient sonification systems. Proc. of the ICAD.

[7] Katharina Groß-Vogt, Marian Weger, and Robert Höldrich. 2018. Exploration of Auditory Augmentation in an Interdisciplinary Prototyping Workshop. In *FMT*. 10–16.

[8] Katharina Groß-Vogt, Marian Weger, Robert Höldrich, Thomas Hermann, Till Bovermann, and Stefan Reichmann. 2018. Augmentation of an institute's kitchen: An ambient auditory display of electric power consumption. Proc. of the ICAD.

[9] Robert H Jack, Adib Mehrabi, Tony Stockman, and Andrew McPherson. 2018. Action-sound latency and the perceived quality of digital musical instruments: Comparing professional percussionists and amateur musicians. *Music Perception: An Interdisciplinary Journal* 36, 1 (2018), 109–128.

[10] Bryan J Kemp. 1973. Reaction time of young and elderly subjects in relation to perceptual deprivation and signal-on versus signal-off conditions. *Developmental Psychology* 8, 2 (1973), 268.

[11] Justyna Maculewicz, Cumhur Erkut, and Stefania Serafin. 2016. An investigation on the impact of auditory and haptic feedback on rhythmic walking interactions. *International Journal of Human-Computer Studies* 85 (2016), 40–46.

[12] Paul Marshall, Alissa Antle, Elise Van Den Hoven, and Yvonne Rogers. 2013. Introduction to the special issue on the theory and practice of embodied interaction in HCI and interaction design. , 3 pages.

[13] Abu Zaher Md Faridee, Sreenivasan Ramasamy Ramamurthy, and Nirmalya Roy. 2019. Happyfeet: Challenges in building an automated dance recognition and assessment tool. *GetMobile: Mobile Computing and Communications* 22, 3 (2019), 10–16.

[14] Zachary Pousman and John Stasko. 2006. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the working conference on Advanced visual interfaces*. 67–74.

[15] Barbara Stadlober, Martin Zirkl, and Mihai Irimia-Vladu. 2019. Route towards sustainable smart sensors: ferroelectric polyvinylidene fluoride-based materials and their integration in flexible electronics. *Chemical Society Reviews* 48, 6 (2019), 1787–1825.

[16] F Stevens, D Murphy, and S Smith. 2016. Emotion and soundscape preference rating: using semantic differential pairs and the self-assessment manikin. In *Sound and Music Computing conference, Hamburg*, Vol. 2016.

[17] Ana Tajadura-Jiménez, Joseph Newbold, Linge Zhang, Patricia Rick, and Nadia Bianchi-Berthouze. 2019. As light as you aspire to be: changing body perception with sound to support physical activity. (2019), 1–14.

[18] Luca Turchet, Stefania Serafin, Smilen Dimitrov, and Rolf Nordahl. 2010. Physically based sound synthesis and control of footsteps sounds. In *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx-10)*. 161–168.

[19] I.; Höldrich R. Weger, M.; SvoronosKanavas. 2022. Schrödinger's Box: an artifact to study the limits of plausibility in auditory augmentations. *To be published at: Audio Mostly* (2022).

[20] Martin Zirkl, G. Scheipl, Barbara Stadlober, Christian Rendl, P Greindl, Michael Haller, and P Hartmann. 2013. Pyzo Flex ®: A printed piezoelectric pressure sensing foil for human machine interfaces. *Proceedings of SPIE - The International Society for Optical Engineering* 8831. https://doi.org/10.1117/12.2025235