Effect of acoustic low frequencies on vibro-tactile perception during an augmented walking task

...

By making use of instrumented shoes provided with mini-loudspeakers and in-sole force sensors and vibrotactile actuators, we investigated the influence of auditory cues on the perception of underfoot vibrations during a walking task. A real-time system was set up which acquires force data measured by the sensors and generates auditory and vibrotactile feedback, simulating a ground covered with underbrush. This feedback was sent to the shoes and, limitedly to sound, to a loudspeaker set. Seventeen subjects were asked to wear the augmented footwear and walk along a path across different trials, during which the intensity of low frequencies from the loudspeakers could be varied midway along that path. At the end of each trial, subjects had to differentiate between changes in the vibrotactile feedback by answering "yes" or "no", and to associate their judgement to a corresponding level of confidence. Results show a significant percentage of correct answers, suggesting the existence of a cross-modal effect of sound changes on vibrotactile perception with decreasing and increasing low frequency levels, both in the order of 6 dB or more.

Categories and Subject Descriptors: H.1.2 [**Information Systems**]: User/Machine Systems—*human factors; human information processing*; H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces evaluation—*haptic I/O; input devices and strategies; TODO*

General Terms: Experimentation, Measurement, Human Factors, TODO

Additional Key Words and Phrases: tactile illusion, cross-modality, ecological perception, interactive shoes

1. INTRODUCTION

In recent years, multimodality has gained interest in the field of human-computer interaction. Besides traditional and novel visualization paradigms, alternative haptic and auditory design concepts have become relevant in an effort toward the goal of achieving seamless interaction with computers, smart objects, virtual and augmented environments. Often following an *ecological* approach, several researchers now take into consideration all sensory modalities that humans employ during their interactive experience of the surrounding environment [CITARE QUALCOSA SU ECOLOGY OF INTERACTION].

Concerning in particular the act of walking, starting from the elementary obser-

The research leading to these results has received funding from the European Community's Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking. TODO Author(s) address(es).

vation that ground perception affects all everyday tasks in which locomotion and self-equilibrium are involved, interaction designers have begun to consider foot-based machine interfaces as means to exchange information with the user [?; ?; ?; ?]. Notably, this information is carried by visual, auditory, tactile, vestibular and proprioceptive cues [?; ?].

Investigations on the relationships between perception and action in walking exist, which analyze the role of footstep sounds on posture [?] and gender [?] recognition, as well as their potential affective characters [?]. In parallel, there is substantial literature investigating walking perception in the audio-visual [?] and visual-tactile [?] domains.

Conversely, to our knowledge little attention has been paid to audio-tactile phenomena linked to walking. Such phenomena are especially important, since they bring a subjective picture of the ground in all cases when the visual attention is focalized elsewhere. Exceptions exist, dealing with the role of haptic and/or auditory information in the identification of ground materials [?; ?]. Such exceptions become less sparse if we consider the diverse receptive areas, such as the hand or whole body. In this broader domain, studies can be found which take into account the cross-modal effects of audio stimuli on vibrotactile perception [?; ?; ?]. Explorations in the opposite direction are fairly common as well: for instance, studies on the effects of vibrotactile stimuli in auditory perception [?] or in auditory-induced phenomena [?].

To complete the picture, some works on audio-tactile perception investigate cross-modal integration/segregation by temporal, spatial, frequency, and intensity cues [?], dealing for example with bias in tactile perception by incongruities in audio and haptic stimuli [?; ?], and with effects of synchronicity [?] or spatial distribution of audio and tactile stimuli (called ventriloquism in the audio-visual case) [?].

In this paper we focus on the effect of auditory stimuli on vibrotactile perception during walking tasks. For some aspects similarly to what has been researched for the hands [?], we have focused our attention on the role of spectral and intensity cues: Specifically, on the effect of loudness changes in low frequency.

2. METHOD

2.1 Apparatus

The setup used in the experiment is a sophisticated software/hardware system, realizing a footwear-based walking interface that has been developed for augmented-reality applications [?]. A pair of sport sandals with rubber soles have been equipped with force-sensing resistors (FSRs), vibrotactile transducers (haptuators) and mini-loudspeakers. The sandals used are of size 44 of the European scale (10 1/2 in U.S. standard male's size), but thanks to Velcro straps they can comfortably fit smaller foot sizes as well. In detail, each sandal is provided with a couple of Interlink 402 FSRs and TactileLabs TL002-14-A haptuators – placed respectively under the insole and inside the sole, one by the heel and one by the forefoot – and a mini battery-powered loudspeaker (Goobay Soundball) attached next to the instep. Figure 0?? shows the positioning of the devices. Moreover, a set of four Genelec 8020A bi-amplified speakers are placed at the four corners of the experiment room, pointing outward and directed towards reflective panels which allow to diffuse sound in a non-directional manner.

The system is able to detect and track the gesture of walking – making use of the contacts heel–ground and forefoot–ground – thanks to the FSRs which provide force signals to an Arduino Duemilanove acquisition board connected via USB to a Mac Pro desktop computer. The data are sampled on four channels (two FSRs per sandal) with a uniform sampling rate of 1470.5 Hz and 10 bit of resolution [?]. This continuous flow of force data is processed and used to drive a physically-based sound engine [?] which allows to synthesize ecological sounds as those of footsteps on creaking floors or aggregate (that is, non homogeneous) grounds, like icy snow and brushwood. All the processing and synthesis are done in the real-time programming environment Pure Data.

Before being sent to a RME Fireface 800 professional multi-channel audio interface, the audio signals are processed and routed along a threefold path: two audio signal paths and one path going to the haptuators (see Section 2.1.1 for details). Moreover, because of the haptuators' input specifications, the signals feeding them are amplified by means of a couple of T.AMP S-75 stereo amplifiers. All audio and haptic signals are rendered at 44.1 kHz.

The experiment was conducted in a 4.7×8 m sized and quiet (see Section 2.4 for details) room with good sound attenuation qualities: the walls are either covered with furniture or curtains, thus avoiding unwanted reverberation, while the floor is made of PVC, hence resulting acoustically neutral (almost silent) when walking over it with rubber sole shoes. TODO schema of the experiment room / path Thanks to the room's properties, the actual sound made by the user while walking is completely surpassed and substituted by the synthetic sound.

2.1.1 *Details and remarks.* Considering that the force signal useful for detecting basic gesture of walking is the ground reaction force (GRF), which is essentially responsible for the center of mass movement of the walker, we are interested in a frequency range up to about 300 Hz [?]. Moreover, the chosen FSRs have a response time of about 2 ms, which limits the bandwidth of the acquisition components to 0-500 Hz [?]. Therefore a sampling rate of 1470.5 Hz is well beyond the requirements for such configuration.

Because of the set up requirements (e.g. the use of audio amplifiers) and in order to minimize the latency, all the connections are wired. However we managed to minimize most encumbrances by gathering all the wiring inside a couple of Ethernet cables (one per sandal), each coming out from the sole's heel. The cables are then connected to the data acquisition board (which receives force signals coming from the FSRs) and to both the audio interface (that sends audio signals to the loudspeakers) and amplifiers (which provide signals for the haptuators). The cables were fastened at the participants' waist making use of a couple of snap-hooks, and then directed to a pulley attached at the ceiling, in this way optimizing the user's freedom of movement. TODO picture of a subject walking?

As anticipated above, the synthesized signals are used to feed both the loudspeakers and haptuators. With regard to the audio paths, the signal is split into two frequency bands by means of a 24 dB/octave cross-over filter (realized in Pure Data) set to around 162 Hz, and the upper frequency band is routed to the mini-speakers mounted on the sandals (referred as "mini-speakers" from now on), while the complementary lower band is routed to the four broadband loudspeakers at the room's corners (referred as "speakers" from now on). We set up the cross-over frequency in order to take advantage of the *precedence effect* [?] – commonly exploited in hi-fi and home theater configurations using a subwoofer and mid/high-range satellite speakers. In this way, while walking with the augmented sandals, the user hears an ecologically consistent sound of footsteps coming exclusively from her feet. As for the vibrotactile feedback, before being routed to the amplifiers and then the haptuators, the signals are first equalized: in order to better fit the haptuators' bandwidth we decided to filter out frequencies below 20 Hz by means of 12dB/octave hi-pass filters, and to slightly attenuate the signal amplitude in the vicinity the haptuators' main resonance frequency. All the filtering was done in Pure Data. The haptuators vibrate essentially along a widthwise direction relative to the sandals. However, due to the coupling with the sandals' sole and insole, vibrations propagate also longitudinally and with direction normal to the sole of the foot (see Section 2.4.3 for details).

Since in humans the two-point discrimination threshold (that is, the maximum spatial resolution) of the sole varies from around 1.5 cm at the big toe to 3.4 cm at the heel [?], it is clear that even only two haptuators (exposed surface: 2.9×1.4 cm) – especially considering their coupling with the insole (see Figure 0??) – can give the user the sensation of a fairly diffuse vibration underfoot at the heel and forefoot.

As a result, the system is capable of providing a realistic audio-tactile experience of walking on different grounds, depending on the sound engine's settings.

2.2 Participants

Seventeen subjects, twelve male and five female aged between 20 and 26 years, voluntarily participated in the experiment. All of them had normal hearing capability, while eight of them were or used to be musicians. Also, all the participants reported to usually wear snickers or other rubber-sole shoes. Their foot size ranged from 36 to 44 of the European scale (4 1/2 to 10 1/2 in U.S. standard male's size).

2.3 Stimuli

2.3.1 *Soundscape.* Taking inspiration from the results reported in [?], we decided to play back an appropriate soundscape during the experiment. In particular, we used the recording of environmental sounds of a forest (especially representing wind in the trees, birds singing), which we mixed together with the recording of a river flowing. We prepared a looped sample avoiding excessive loudness peaks, and this way we were able to provide a quiet, uniform enough but effective soundscape, which filled a large part of the sound spectrum, thus offering an enhanced sense of presence and isolating the user from possible external interferences.

As anticipated in 2.1 the soundscape was played through the four speakers.

2.3.2 *Interactive stimuli.* Making use of a physically-based sound model, we were able to simulate the sound of footsteps through underbrush. Since the model generating such sounds is based on a stochastic process [?], the synthesized sounds are quite varied, as those resulting from walking on real irregular ground. However, in order to provide the subjects with fully controllable stimuli, we decided not to synthesize the audio in real-time, but rather to play back two pre-recorded samples corresponding to the contacts heel–ground and forefoot–ground, as soon as the heel or forefoot of the sandals came into contact with ground. In this way we were able to provide a controlled audio-tactile feedback, constant in terms of frequency content and intensity. Moreover, in order to enrich the frequency spectrum of the stimuli – especially for the auditory feedback – we decided to mix two additional recordings (as before, one corresponding to the heel and one to the forefoot) of footsteps on dry leaves, together with the synthesized signals. The spectrum of the signals obtained exhibits a marked low-pass character.

After an informal evaluation process provided by people non participating in the experiment, we found a suitable level for the haptic feedback, avoiding distortion or excessive noise from the haptuators, which resulted in a light vibrotactile sensation for most of the subjects.

Considering that the skin of the hand – which is the most sensitive part of our body – can perceive vibrations of frequency up to 1 kHz [**?**] it is clear that our signals are suitable for providing effective vibrotactile stimulations (see Section 2.4.3 for actual measurements of the rendered vibration's frequency and intensity).

After some informal judgments given by people non participating in the experiment, we fixed the relative and absolute loudness (see Section 2.4) of the environmental and footsteps sounds in order to provide the subjects with realistic ecological sound feedback. Since the goal of the experiment was to investigate the effect of auditory feedback – particularly in the lower frequency band – on the perception of vibrotactile stimulations, we payed special attention in setting a reference level w for the low-frequency signal (LFS) feeding the speakers in order to match the vibrotactile stimulus (which was kept to a fixed level along the whole experiment).

Finally, we implemented a switch in our Pure Data program which allowed to choose among three levels of the LFS: w, that is the reference level, $w^{+} = w + 6$ dB and $w^{++} = w + 12$ dB. See Section 3 for more insights on the utilization of these stimuli.

As a final remark, the soundscape and LFS (both feeding the four speakers) were sent as mono signals, the upper-band audio signals (feeding the mini-speakers) were sent as stereo signals gathering the sounds of both the heel and forefoot of each foot, while the haptic signals were treated as four separate channels corresponding to the heel and forefoot of each foot (respectively feeding a different haptuator).

2.4 Measurements

2.4.1 *Latency measurement.* We measured the input-output latency of our system as the delay between the onset of an impulse (simulating a change in the force signal) sent to the acquisition board and its counterpart taken from the audio interface's outputs. The measures were done by means of an oscilloscope and were found to be equal to about 15 ms. To this delay, one should add the response time of the devices driven by the computer-generated signals (2 ms for the haptuators; unknown, but likely very low, for the speakers) and the latency due to the distance of the speakers from the ears (the speed of sound in air is 343 m/s). As for the mini-speakers attached to the sandals, the latency due to their distance is perfectly ecological and related to the height of the subject, while the latency due to the distance of the speakers at the room's corners was found to help the precedence effect (sound localization), while not adding any troublesome delay. Indeed, the system is very reactive and very little or no latency can be noticed, thus giving rise to a credible, unitary percept.

2.4.2 *Acoustic measurement.* All the acoustic measurement were made with a digital phonometer Cesva SC-2c which was set to average the sound pressure over 1 second. Since the experiment involved to walk along a path in the room, the measurements were done in two positions: at one extreme of the path and halfway (see Figure 0**??**), at a height averaging the subjects' listening point (about 160 cm).

The noise floor measured in the middle of the path was 32.2 dB(A).

The sound pressure measured while playing back only the soundscape was most of the times in the range 43.5–47.5 dB(A) halfway and 43.5–46 dB(A) at one extreme of the path. Such counterintuitive intensity difference between the two positions is likely due to phase cancellations.

The sound pressures while walking and playing back the soundscape and footsteps sound through both the speakers and mini-speakers were measured for the three LFS levels used in the experiment (w, w^{+}, w^{++}, see Section 2.3.2). During the measurements, the footsteps sounds were played continuously, independently of the gait, in order to obtain consistent sound pressure values from the integrator of the phonometer. The resulting measures were: for w and w^{+}, average pressure between 45–47 dB(A) both halfway and at one extreme; for w^{++}, average pressure between 45–48 dB(A) both halfway and at one extreme. Indeed the A-frequency-weighting justifies such small variations in the sound level when increasing the level in the lower-frequency band (< 162 Hz) of 6 and 12 dB.

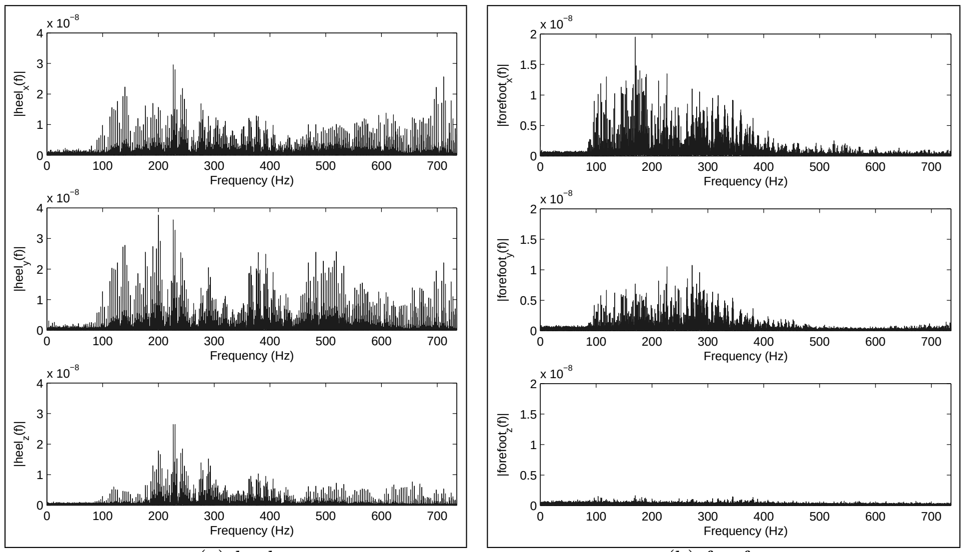
2.4.3 *Vibration measurement.* Making use of 3-axes accelerometers (Freescale MMA7260Q) fixed to the insole next to the haptuators, we measured the intensity of the vibrotactile feedback at the heel and forefoot. As anticipated in 2.1, due to their construction and placement the haptuators vibrate essentially along a widthwise direction relative to the sandals. However, owing to the coupling with the sandals' sole and insole, vibrations propagate also longitudinally and with direction normal to the sole of the foot.

The positioning of the accelerometers prevented to do the measurements during the experiment, without interfering with the latter. However, in order to better simulate the experimental conditions, the measurements were done while an average-weight subject was standing upright and put his weight first on the heel and then on the forefoot. In order to facilitate the analysis, the vibrotactile signals were played back continuously, avoiding silent passages. As a result, we obtained the accelerations on three axis – namely widthwise (x-axis), lengthwise (y-axis), and perpendicular to the sandals (z-axis). We removed the DC component, and after a double integration (using the trapezoidal rule) we obtained the displacement vectors on the three axis. In particular, the measured signals have the following peak amplitudes (in absolute value):

—heel: x-axis, 1.5 μm; y-axis, 1.7 μm; z-axis, 0.9 μm. The vectorial sum has maximum amplitude 2.4 μm.

—forefoot: x-axis, 0.7 μm; y-axis, 0.5 μm; z-axis, 0.1 μm. The vectorial sum has maximum amplitude 0.8 μm.

The single-sided amplitude spectrum of the vibrotactile signals is shown in Figure 1.



(a) heel (b) forefoot

Fig. 1. Single-sided amplitude spectrum of the vibrotactile signals measured at the heel (left) and forefoot (right). From top to bottom, x, y and z axes are represented.

In literature, sensitivity thresholds are commonly expressed as vibration amplitudes relative to both the location and area of contact, and the frequency of the stimulus. Moreover, stimuli are generally provided as vibrations perpendicular to the skin. For example, the smallest thresholds (highest sensitivity) for the hand were found in the frequency range 200-300 Hz, and in particular for a 250 Hz sinusoidal stimulus the hand is sensitive to vibrations of amplitude ranging from 0.01 to 2 μm, respectively for contact areas from 5.1 to 0.08 cm^2 [**?**]. In general, the smaller the contact area, the higher the threshold (that is, the lower the sensitivity).

As for the foot, we could not find equivalent investigations of its sensitivity for a contact area covering the whole foot and for broadband stimuli. However, it is generally acknowledged that – on equal terms – the foot is much less sensitive than the hand. For instance, for a contact area of 0.79 cm^2 and sinusoidal stimuli with frequency of 250 Hz, thresholds of 2 μm at the heel and 5 μm at the big toe were found [**?**]; in a different study, for a contact area of 0.5 cm^2 and sinusoidal stimuli with frequency of 240 Hz, thresholds of about 1 μm at fore- and mid-foot and 2 μm at the heel were found [**?**].

In our case several aspects are worth noticing:

—the contact area is much larger than that used in the experiments cited above (the haptuators alone cover an area of 4 cm^2 each);

—the stimuli are broadband, almost noisy signals;

—the stimulation is not provided along a direction perpendicular to the skin;

—during the experiment, the whole foot is always yet variably in contact with the insole;

—the vibrotactile augmentation only occurs while the foot depresses the ground, that is when subjects are performing an action. The subjective thresholds of sensitivity to floor textures and the size of their patterns during walking tasks are yet to be systematically investigated.

All this likely justifies the fact that in experiment even vibrations of such small amplitude can exceed the sensitivity threshold of the foot.

3. PROCEDURE

The subjects were asked to wear the augmented sandals and to walk with a regular rhythm along a predefined path (see Figure 0**??**) while – starting from halfway – the intensity of the LFS could be in/decreased or left unaltered. Subjects had been preliminarily informed that a change in the system's feedback could happen once during each take, halfway in the path, however they were not aware that only the audio feedback could actually be altered.

On the whole, the experiment lasted about 45 minutes and consisted of twelve experimental configurations corresponding to several combinations of LFS levels: a set of six "increments" (namely, three no-variation conditions and three actual increments) plus a set of six "decrements" (again the same three no-variation conditions as above, and three actual decrements). In detail, every condition was made up of a pair of LFS levels: one for the first half of the path and a second one for the second part. Table I shows the LFS level pairs used in the experiment. The pairs S1, S2 and S3 (no-variation) are the same for both increments and decrements. On the other hand, the pairs S4, S5 and S6 (corresponding to LFS variation) are in reverse order for increments and decrements, however they are labeled in the same way as they represent the same two LFS levels.

	Increments		Decrements	
	1st	2nd	1st	2nd
S1	w	w	w	w
S2	w^+	w^+	w^+	w^+
S3	w^{++}	w^{++}	w^{++}	w^{++}
S4	w	w^+	w^+	w
S5	w	w^{++}	w^{++}	w
S6	w^+	w^{++}	w^{++}	w^+

Table I. Legenda of sound stimuli and sequences of LFS level increments and decrements. Note that the conditions S1, S2 and S3 (no-variation) are the same for both increments and decrements, while the conditions S4, S5 and S6 (variation) are in reverse order for increments and decrements.

Each condition was repeated four times, for a total of 48 trials. Moreover the decrements and increments were presented separately, that is half subjects started with a randomized sequence of 24 "decrements" (actually, twelve decrements and twelve no-variation trials) and half with a randomized sequence of 24 "increments" (twelve increments and twelve no-variation trials).

At the end of each take, they had to answer by marking a preprinted form whether they had felt any change in the vibrotactile feedback under their feet (answer: yes/no) and the corresponding confidence rating (from 1 to 7, where 1: "absolutely not confident", 4: "moderately confident", 7: "very confident"). As a security measure, following each take the subjects had also to tell the experimenter their yes/no answer. In this way we were able to double-check the results after each experiment by comparing verbal and written answers.

In order to help subjects to focus on their sense of hearing and touch, artificial lights in the experiment room were turned off, so that light could only enter from a full-length window covered by a curtain. Nevertheless, there was always enough light for them to walk safely and confidently.

Before starting the experimental session, each subject was instructed about the path and had two minutes of free walk around the room for becoming familiar with the equipment. In particular, during this phase the system was set as in the experiment except for three features:

(1) the forest soundscape was turned off. The training was not intended to provide sense of presence, but only to help participants to focus on the auditory and vibrotactile stimuli;
(2) the LFS was turned completely off. This helped avoiding that, while walking close to the speakers and thus canceling the precedence effect, subjects could discover that part of the footstep sounds came from there instead of the sandals;
(3) a metronome-like signal was played in order to provide the subjects with a reference walking speed. The temporal interval was set to 720 ms, resulting in a medium In this way, we were able to obtain very similar gait among all the participants. This requirement also allowed to avoid that subjects started "investigating" how the system worked, for example by excessively slowing down their gait.

During the experiment, the computer managing the whole system was hidden and its screen turned off. Two experimenter were always present in the room, one of them being responsible of switching the LFS level by means of a USB MIDI controller (Novation ReMOTE ZeRO SL, providing visual feedback to experimenter via backlit LCDs) interfaced with Pure Data, while watching the participants walking.

4. RESULTS

For each participant, the percentages of "yes" responses were calculated for the twelve experimental conditions. For the different analysis, the difference from random percentage (50%) was tested using one-proportion (two-tailed) z-tests, and we used two-proportion (two-tailed) z-tests in order to check the differences between the experimental conditions. An alpha-level of 0.05 was adopted.

A first global analysis considered the percentages of "yes" responses in the stimulus conditions "*without variation* vs *with variation*" and in the perceptual conditions "*increase* vs *decrease*". The results are presented in Figures 2 and 3. The analysis revealed that the percentages of "yes" responses in the *without variation*

(39.46%) and *with variation* (62.25%) conditions were significantly different from random ($z = -4.26$, $p < 0.001$ and $z = 4.95$, $p < 0.001$, respectively) and these two conditions were significantly different the one from the other ($z = -6.51$, $p < 0.001$). By contrast, the analysis showed that the percentages of "yes" responses in the *increase* (50.25%) and *decrease* (51.47%) conditions were not significantly different from random ($z = 0.10$, $p = 0.92$ and $z = 0.59$, $p = 0.55$, respectively) and these two conditions were not significantly different ($z = -0.35$, $p = 0.73$).

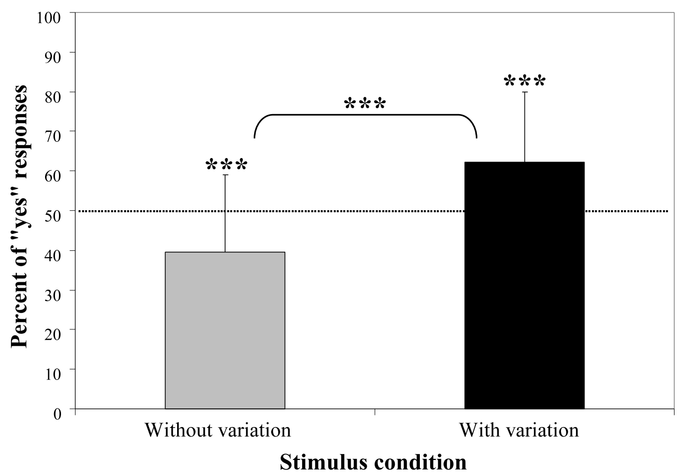


Fig. 2. Mean percentage of "yes" responses (bars represent std) for the stimulus conditions "*without variation* vs *with variation*". The difference from random (line at 50%) was tested using one-proportion (two-tailed) z-tests. The difference between the two stimulus conditions was tested with a two-proportion z-test. Legenda: ***: $p < 0.001$.

In a detailed analysis, we considered the percentages of "yes" responses for the two stimulus conditions "*without variation* vs *with variation*" as a function of the perceptual conditions "*increase* vs *decrease*". The results are presented in Figure 4. For the *without variation* condition the percentages of "yes" responses in the *increase* (41.18%) and *decrease* (37.75%) conditions were significantly different from random ($z = -2.52$, $p < 0.05$ and $z = -3.5$, $p < 0.001$, respectively). Similarly, for the *with variation* condition the percentages of "yes" responses in the *increase* (59.31%) and *decrease* (65.20%) conditions were also significantly different from random ($z = 2.66$, $p < 0.01$ and $z = 4.34$, $p < 0.001$, respectively). The differences between the four conditions were tested with a Bonferroni-adjusted alpha-level (p divided by the number of comparisons), thus the alpha-level was adjusted to $p = 0.05/4 = 0.0125$. For both the *without variation* and *with variation* conditions the analysis showed that the difference between the *increase* and *decrease* conditions was not significant ($z = 0.71$, $p = 0.48$ and $z = -1.23$, $p = 0.22$, respectively). On the other hand, for the *increase* and *decrease* conditions the analysis indicated that the difference between the *without variation* and *with variation* conditions was significant ($z = -3.66$, $p < 0.05$ and $z = -5.55$, $p < 0.05$, respectively).

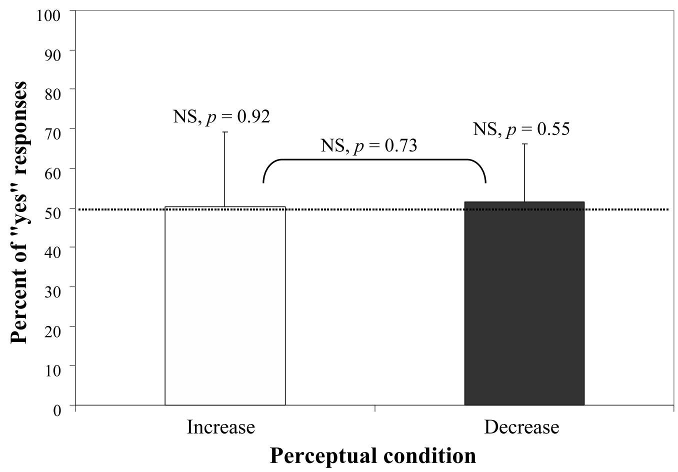


Fig. 3. Mean percentage of "yes" responses (bars represent std) for the stimulus conditions "*increase* vs *decrease*". The difference from random (line at 50%) was tested using one-proportion (two-tailed) z-tests. The difference between the two perceptual conditions was tested with a two-proportion z-test. Legenda: NS: not significant.

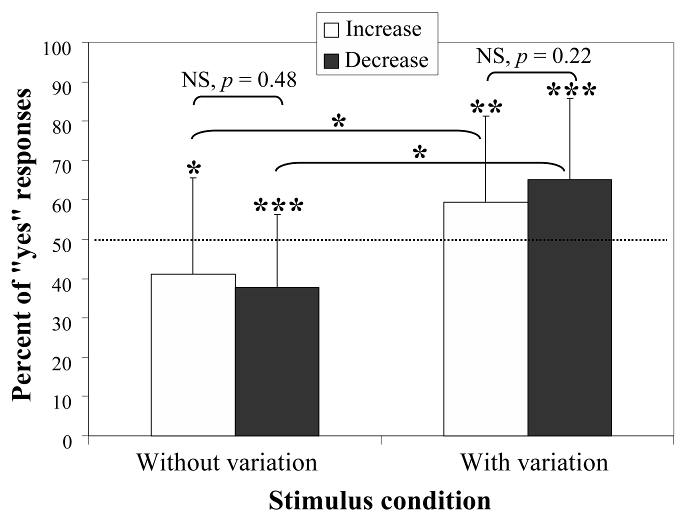


Fig. 4. Mean percentage of "yes" responses (bars represent std) for the stimulus conditions "*without variation* vs *with variation*" as a function of the perceptual conditions "*increase* vs *decrease*". The difference from random (line at 50%) was tested using one-proportion (two-tailed) z-tests. The differences between the four conditions were tested with two-proportion z-tests (two-tailed and Bonferroni-adjusted alpha-level with $p = 0.05/4 = 0.0125$). Legenda: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, NS: not significant.

In another detailed analysis, we considered the percentages of "yes" responses for the two stimulus conditions "*without variation* vs *with variation*" as a function of the sound configurations listed in Table I. The results are presented in Figures 5 and 6. For the *without variation* condition the percentage of "yes" responses

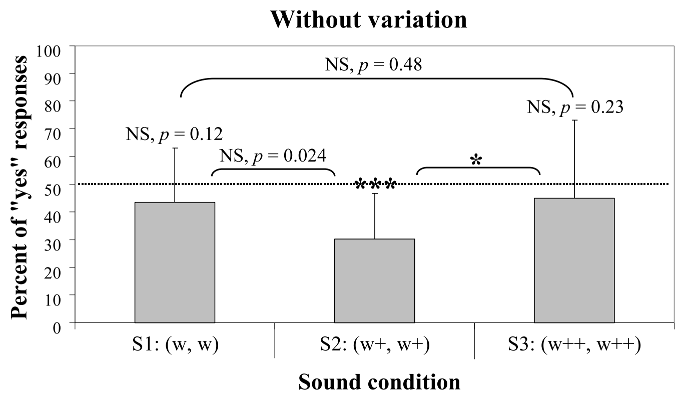


Fig. 5. Mean percentage of "yes" responses (bars represent std) for the *without variation* condition as a function of the sound configurations S1: $(w,\ w)$, S2: $(w^+,\ w^+)$, S3: $(w^{++},\ w^{++})$. The difference from random (line at 50%) was tested using one-proportion (two-tailed) z-tests. The differences between the three sound conditions were tested with two-proportion z-tests (two-tailed and Bonferroni-adjusted alpha-level with p = 0.05/3 = 0.0167). Legenda: *: p < 0.05, ***: p < 0.001, NS: not significant.

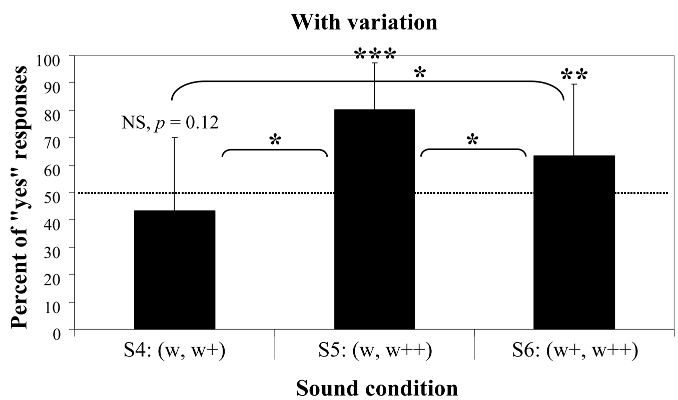


Fig. 6. Mean percentage of "yes" responses (bars represent std) for the *with variation* condition as a function of the sound configurations S4: $(w,\ w^+)$, S5: $(w,\ w^{++})$, S6: $(w^+,\ w^{++})$. The difference from random (line at 50%) was tested using one-proportion (two-tailed) z-tests. The differences between the three sound conditions were tested with two-proportion z-tests (two-tailed and Bonferroni-adjusted alpha-level with p = 0.05/3 = 0.0167). Legenda: *: p < 0.05, **: p < 0.01, ***: p < 0.001, NS: not significant.

was significantly different from random in the S2 (30.15%) condition ($z = -4.63$, $p < 0.001$) but not in the S1 (43.38%) and S3 (44.85%) conditions ($z = -1.54$, $p = 0.12$ and $z = -1.20$, $p = 0.23$, respectively). For the *with variation* condition the percentage of "yes" responses was significantly different from random in the S5

(80.15%) and S6 (63.24%) conditions ($z = 7.03$, $p < 0.001$ and $z = 3.09$, $p < 0.01$, respectively) but not in the S4 (43.38%) condition ($z = -1.54$, $p = 0.12$).

The differences between the three sound conditions in each stimulus condition "*without variation* vs *with variation*" were tested with a Bonferroni-adjusted alpha-level ($p = 0.05/3 = 0.0167$). For the *without variation* condition, the analysis indicated that the S1 condition was not significantly different from the S2 ($z = 2.26$, $p = 0.024$) and S3 ($z = -0.24$, $p = 0.48$) conditions. By contrast, the S2 and S3 conditions were significantly different ($z = -2.50$, $p < 0.05$). For the *with variation* condition the analysis revealed that the S4 condition was significantly different from the S5 ($z = -6.24$, $p < 0.05$) and S6 ($z = -3.28$, $p < 0.05$) conditions. Similarly, the S5 and S6 conditions were significantly different ($z = 3.10$, $p < 0.05$).

Finally, for each participant, the mean confidence ratings were calculated for the four conditions resulting from the crossing of the stimulus conditions "*without variation* vs *with variation*" and the perceptual conditions "*increase* vs *decrease*". A 2 *stimulus conditions* × 2 *perceptual conditions* repeated-measures ANOVA on mean confidence ratings revealed that there was no significant effect for the stimulus condition ($F(1, 16) < 1$). Thus, confidence ratings in the *without variation* ($M = 4.52$) and *with variation* ($M = 4.48$) conditions were not significantly different. Similarly, there was no significant effect for the perceptual condition ($F(1, 16) < 1$). In other words, confidence ratings in the *increase* ($M = 4.53$) and *decrease* ($M = 4.47$) conditions were not significantly different. The *stimulus condition* × *perceptual condition* interaction was not significant ($F(1, 16) < 1$).

5. DISCUSSION

The results summarized in Figure 2 confirm that a cross-modal interference is actually present, and manifests itself as an audio-tactile illusion where audio (LFS) can affect the vibrotactile perception. The significant differences from random (50%) for the two conditions *without variation* and *with variation* can be interpreted like a double proof of such audio-tactile illusion. In other words, we have a clear cause/effect relationship, where the cause is the intensity variation of the LFS and the effect is the illusion: when the cause is present the effect is present, when the cause is absent the effect is absent.

In addition, Figure 3 shows how the perception of participants is not significantly affected by the perceptual condition "*increase* vs *decrease*". This result is consistent with those shown in Figure 4 which backs up the fact that an audio-tactile illusion occurs in both *increase* and *decrease* conditions (double proof), and this audio-tactile illusion is not significantly different between such conditions.

The results pictured in Figures 5 and 6 explain that the association of LFS levels (i.e., the pairs represented in Table I) is crucial, and that there is an optimal association (sweet spot) for the audio-tactile illusion. In this regard, it is an expected outcome that the largest LFS variation (i.e., 12 dB) – represented by S5: (w, w^{++}) – results in a strong illusion. On the other hand, it is found that the same relative LFS variation (6 dB) provided starting from different absolute levels – namely S4: (w, w^{+}) and S6: (w^{+}, w^{++}) – result in markedly different effects. In particular, a sweet spot is found for higher and more clearly perceivable absolute LFS levels (see Section 2.4.2).

As for the confidence ratings given by the participants, the results show a homogeneous behavior across the different conditions. Also, the mean confidence ratings for the different conditions (value around 4, that is an intermediate confidence) indicate that there are no ceiling or floor effects. In other words, the experiment was neither too difficult nor too easy.

6. CONCLUSIONS

The results of the experiment show that subjects feel a change in the vibration underfoot when modifying only the intensity of footsteps sound in the lower-frequency band.

Several subjects interpreted such change as corresponding to a different level of sinking in a virtual ground. This is even more surprising considering that in such scenario the proprioceptive information due to compliant materials – absent in our case – should play a substantial role. Nevertheless this result are in accordance with what described in **?**; **?**]: the authors found that, for solid ground surfaces, the vibrotactile and auditory channels are likely more important as conveyance of information for the identification of materials than proprioceptive information. This suggests that, in simulations, the identity of such materials may be preserved to an acceptable level of accuracy even without the rendering of proprioceptive information.

Received Month Year; revised Month Year; accepted Month Year