THE INFORMATION CAPACITY OF PLAUSIBLE AUDITORY AUGMENTATIONS: PERCUSSION OF RECTANGULAR PLATES

Marian Weger[1], *Michael Aurenhammer*[1], *Thomas Hermann*[2], *Robert Höldrich*[1]

[1] IEM, University of Music and Performing Arts, Graz, Austria
[2] Ambient Intelligence Group, CITEC, Bielefeld University, Bielefeld, Germany
weger@iem.at

ABSTRACT

For unknown physical objects, we often infer the "ground truth" (e.g., on material, hollowness, or thickness) that is hidden below the visual appearance by percussion, i.e., by knocking on them. Auditory augmentations embed digital information into physical objects by modulating their auditory feedback in a plausible way. Based on physically justified sound models, we are able to design auditory displays that blend seamlessly into the acoustic environment. We assume that magnitude estimations are easier for physical parameters than for abstract sound parameters, at least for untrained users. In two experiments we measured how listeners extract physical information (size, aspect ratio, material) from the sound of impacted rectangular plates. First, participants actively explored an augmented table through a ballpoint pen. The second experiment evaluated only unisensory auditory identification of physical parameters. The results let us estimate the total information capacity of such interactive sonifications with multidimensional (2D and 3D) parameter mappings. Despite using only natural sounds, the information capacity of both mappings was on par with a comparable 1D auditory augmentation and only slightly below 1D auditory displays based on more salient but implausible sound parameters. The results additionally allow a better understanding of human sound source identification in general.

1. INTRODUCTION

Nothing is more annoying than a bad sonification, not adapted to the specific environment and users it aims for. If providing critical information to a single professional, e.g., performing a surgery or controlling an airplane, the information is best mapped to the most salient sound parameters in the given environment (e.g., hospital or cockpit). If uncritical information should be provided ambiently to multiple users in a shared office, without disturbing others, an entirely different approach is necessary; e.g., *auditory augmentation* [1, 2, 3]. In its strict sense, the original auditory feedback that results from physical interaction is modulated for conveying additional information. Sonifications in general need time to be perceived by living beings. Interactive sonifications such as auditory augmentations are living themselves: they cannot be consumed passively but require active participation. In order to stay calm and unobtrusive, the augmented auditory feedback stays in a plausible but usable range, with respect to the physical object and the performed action.

Besides auditory augmentation, several closely related concepts have been proposed. Gaver [4] parameterized auditory icons based on physical models, whereas Barrass [5] fabricated solid physical from model-based sonifications of digital data [6]. Mauney and

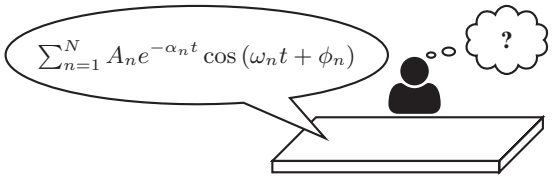


Figure 1: *How many bits fit in a rectangular plate?*

Walker [7] used soundscapes for unobtrusive monitoring of peripheral data. Ferguson [8] introduced *ambient sonification systems* for providing an invisible interface to ambient data by augmenting interactions with physical objects in a domestic environment by sound. Tünnermann et al. [9] created *blended sonifications* which "blend into the users' environment" without confronting them with any explicit technology, always ready to hand if needed. Blended sonifications manipulate interaction sounds or environmental sounds in such a way "that the resulting sound signal carries additional information of interest while the formed auditory gestalt is still perceived as coherent auditory event" [9]. We assume that due to their restrictions, such interactive sonifications are rather limited in information capacity.

Pollack [10, 11] measured the information capacity of perceptual parameters such as pitch or loudness based on absolute magnitude estimations. Participants estimated the value of a given sound parameter on an ordinal scale, while parameter range and the number of discrete levels were varied in several conditions. The received information of each condition is then the average number of correctly identified levels. The information capacity is the maximum of all conditions, i.e., partitionings. Instead of levels L, it is usually given in bits: $C = \log_2 L$. If plotted against sent information (i.e., number of discrete levels), the received information usually starts as a straight line (1 level sent, 1 level received), rising with increasing sent information, but successively approaches a maximum that is never exceeded, no matter how much is sent: the information capacity. This pattern is similar for any perceptual parameter in any sensory modality; and the information capacity is always in the region between 2 and 3 bit or around 7 levels: the capacity of our short-term memory [12]. Due to this strict limitation of human perception, the saliency of the parameter or its range play only a minor role [11]. However, there is a way out of this dilemma: combining several parameters to a multidimensional auditory display.

If combining the two presumably orthogonal perceptual parameters of pitch and loudness to a 2D auditory display, the number of combined discriminable levels doubles from 5 to 10 [13]. But wait, shouldn't it rather be $5 \times 5 = 25$ levels, i.e., 2×2.3 bit? That

DOI: 10.5281/zenodo.7552234

is the question. And what if we chose less annoying, less salient, less orthogonal sound parameters? In fact, we did not have this prior work on information capacity in mind when designing the experiments that are presented here.

The research question we want to shed light on now is: "what is the information capacity of a rectangular thin plate?" We chose this very specific physical object for several reasons. Aiming at plausible auditory augmentations, it suits many everyday objects in our rather rectangular world: tables, walls, windows, houses, computers, even this article. We are already accustomed to the sound of rectangular objects; maybe we are even experts in auditory perception of their sound. The sound of rectangular plates can be synthesized by simple physical models, e.g., [14, 15, 16]. Modal synthesis is established in psychoacoustic experiments, e.g., [17, 18], and cannot be discriminated from real recordings [19], even if applying crude simplifications [20].

In order to answer our research questions, we will first briefly summarize how the physical properties of a rectangular plate are encoded in its sound, and how this physical information may be extracted by human listeners (Sec. 2). A 2D auditory augmentation of a table with varying length and aspect ratio is evaluated in Sec. 3. A follow-up listening experiment using pre-rendered sounds then investigates a 3D display that employs aspect ratio, metallicity, and rigidity (Sec. 4). The results are discussed with respect to our research question, i.e., the information capacity of plausible auditory augmentations of rectangular plates, in Sec. 5. General conclusions are drawn in Sec. 6.

2. FROM PHYSICAL PARAMETERS TO SOUND PARAMETERS AND VICE VERSA

Our personal experiences suggest that the average person is able to distinguish between different materials and shapes of rigid physical objects by exploring the auditory feedback through tapping, or scratching. On the basis of listening experiments from the literature[1], with synthesized and also with physically struck objects, we feel safe to say that humans can almost perfectly discriminate between gross material categories (glass/metal vs. wood/plastic, at least for small damping) [22, 23], and to some extent even between materials within categories (e.g., glass vs. metal) [17] or between different sizes and shapes (e.g., small vs. large [24, 18, 25], or plate vs. bar [26, 27, 28]). In addition, we know that perception of physical parameters benefits from combining different sensory modalities such as audition, vision, and touch [29, 30, 31, 32].

According to a modal synthesis model, we assume that the sound of impacted rigid objects is a sum of N exponentially decaying sinusoids, corresponding to the objects' so-called modes. The simplified impulse response, i.e., response to an ideal impact, is given in Fig. 1. Each individual mode n is defined by the following sound parameters: its starting amplitude A_n (including also frequency-dependent sound radiation), its frequency f_n or angular frequency $\omega_n = 2\pi f_n$, and its decay factor α_n.[2] The mode with lowest frequency corresponds to the object's base frequency. Concerning its sound, the most important physical parameters of a rectangular plate are the plate dimensions (thickness h, length l_x, width l_y, aspect ratio $r_\mathrm{a} = l_x/l_y$, area $S = l_x l_y$), density ρ, and elastic material constants (either 4 rigidities D_i or Young's modulus E, Poisson's ratio ν, shear modulus G_{xy}, and orthotropy $\Omega = \sqrt[4]{D_1/D_3}$) [33, 15]. Additional meta-parameters include longitudinal wave velocity $c_\mathrm{L} = \sqrt{12D/\rho}$ and rigidity $D = \sqrt{D_1 D_3} = E/[12(1-\nu^2)]$. The natural frequencies $\omega_0 = 2\pi f_0$ connect to these via

$$\omega_0 = \frac{\pi h}{2\rho S}\left[\frac{D_1}{r_\mathrm{a}^2}G_x^4 + D_2 H_x H_y + D_3 r_\mathrm{a}^2 G_y^4 + D_4 J_x J_y\right]^{1/2} \quad (1)$$

or

$$\omega_0 = \frac{\pi}{\sqrt{48}}\frac{h c_\mathrm{L}}{S}\left[\frac{\Omega^2}{r_\mathrm{a}^2}G_x^4 + 2\nu H_x H_y + \frac{r_\mathrm{a}^2}{\Omega^2}G_y^4 + \frac{G_{xy}}{3D}J_x J_y\right]^{1/2}, \quad (2)$$

with G_x, G_y, H_x, H_y, J_x, and J_y given in [34]. The (frequency-dependent) damping is expressed by decay factor[3] α and includes loss due to viscoelasticity (α_v) [14, 35, 36], thermoelasticity (α_t) [14, 37], viscosity (α_f) [14], and radiation (α_r) [14]. These either sum up directly or blend between non-metallic ($\alpha_{\mathrm{v},\cancel{\mathrm{M}}}$) and metallic ($\alpha_{\mathrm{v,M}} + \alpha_\mathrm{t}$) via metallicity H [17]:

$$\alpha = (1-H)\alpha_{\mathrm{v},\cancel{\mathrm{M}}} + H\,(\alpha_{\mathrm{v,M}} + \alpha_\mathrm{t}) + \alpha_\mathrm{r} + \alpha_\mathrm{f}\ . \quad (3)$$

While α_v are proportional to frequency, α_t are approximately constant over frequency, but weighted for each mode individually, depending on the mode shapes [14, 35]. In our simplified model, the amplitude of a given mode depends on its radiation efficiency (usually low for low frequencies, i.e., some kind of high-pass filter, based on an empirical model [38]), its shape at the excitation position (zero on nodal lines, positive/negative at peaks/troughs [16, 34]), and the plate's indentation hardness (temporal Hann window modeled as 3rd-order low-pass filter at cutoff frequency f_cH [39, 40, 16]). To some extent, it is possible to reverse this process of physical modeling, in order to derive physical parameters from measured sound parameters [41, 42, 28, 43, 15]. This direction, however, exhibits ambiguities, e.g., with size and material both affecting the base frequency and thus pitch. In such cases we tend to base our judgments on expectations due to our everyday acoustic environment (e.g., small metal bars and large glass plates) [22].

In addition, our perceptual resolution differs across sound parameters and thus also physical parameters.

Decay factor. We employ an empirical formula for the just-noticeable difference (JND) in time constant [18], based on data from [44, 45]. Its valid range between $2\,\mathrm{ms}$ and $200\,\mathrm{ms}$ fits 45 JNDs. $\tau = 20\,\mathrm{ms}$ yields a Weber fraction of $35\,\%$. Based on synthesized plucked strings, a Weber fraction of $40\,\%$ was measured [46].

Amplitude. An empirical formula for the JND in amplitude was derived [18], based on measurements with pulsed sinusoids by [47] and the frequency-dependent threshold of hearing [48]. A plausible range of $60\,\mathrm{dB}$, roughly equaling the range of music at $1\,\mathrm{kHz}$ [49, 17], is divided into 17 JNDs.

Missing partials. A missing lower partial within the first 3 to 6 partials is easily detected (sensitivity d' between 2 and 8), contrary to missing higher partials ($d' < 2$) [50].

Base frequency. Within the valid frequency range between $0.2\,\mathrm{kHz}$ and $8\,\mathrm{kHz}$ (which usually suffices for everyday objects), the JND in frequency follows an empirical formula [18]. It fits 1452 JNDs.[4] In practice, the discrimination of fundamental frequency f_1 of complex tones depends on the signal duration and the number of partials

[1] See [21] for a comprehensive literature review on auditory and multisensory perception of physical information.

[2] Subscript indices are omitted from now on for better readability.

[3] Decay factor α connects to loss factor η, Q-factor, time constant τ and $-60\,\mathrm{dB}$ reverberation time T_{60} via: $\eta = \frac{1}{Q} = \frac{2\alpha}{\omega} = \frac{2}{\omega\tau} = \frac{2\ln(1000)}{\omega T_{60}}$.

[4] Note that the underlying JND describes the minimum absolute frequency difference that is needed to identify the sign of the frequency difference between two pulsed sinusoids [51].

DOI: 10.5281/zenodo.7552234

Table 1: *Model coefficients of the rendered plates in experiment 1.*

		glass	wood	
thickness	h	10	12	mm
viscoelastic loss	η_v or η_{iv}	0.001	[0.0051 0 0.0216 0.0164]	
density	ρ	2550	415	$\mathrm{kg\,m^{-3}}$
rigidities	D_i	[6700 – – 10 270]	[1320 77 82 227]	MPa
orthotropy (fibers in x-direction)	Ω	1	2	
viscous loss	α_f	5.8		Hz
upper cutoff frequency (indentation hardness)	f_cH	>20	2.41	kHz
length	l_x	{0.30 , 0.34 (small) , 0.40 , 0.46 (medium) , 0.53 , 0.61 (large) , 0.70}		m
aspect ratio	r_a	{1.10 , 1.42 (compact) , 1.82 , 2.35 (longish) , 3.02 , 3.88 (bar-shaped) , 5.00}		

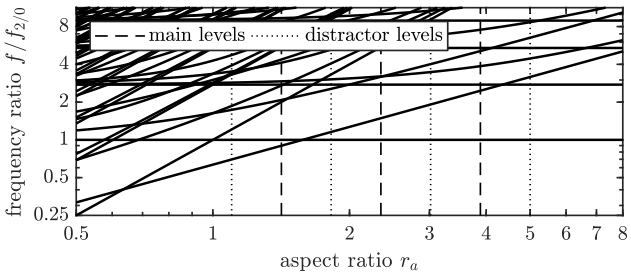


Figure 2: *Frequencies of a rectangular plate, relative to mode* $2/0$*, as a function of aspect ratio; including the levels of experiment 1.*

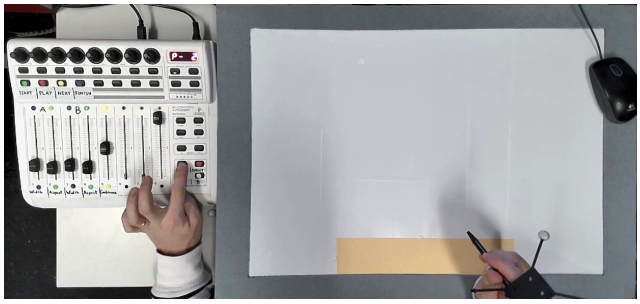


Figure 3: *Apparatus of experiment 1, including the interface plate, tracked ballpoint pen, and MIDI controller (captured via webcam during the experiment).*

[52, 53].

Frequency ratios / intervals. If resonances are far enough apart from each other ($\geq$10 %), they are heard as individual pitches [54]. For complex tones, the JND in frequency ratio is about 1.24 % of the base ratio [55]. One octave (ratio of 2), thus holds 57 JNDs.

Modal density DM. Defined as the average number of partials per Hz, its JND is about $0.3DM$ for low to moderate values [56]. A realistic range between 0.001 (1 mode per kHz) and 0.1 (1 mode each 10 Hz), fits 18 JNDs.

Upper cutoff frequency / low-pass filtering. At high frequencies, we assume a dense spectrum, so that the principles of perception of low-pass filtered noise can be applied, at least at low damping. Independent of base cutoff frequency, JNDs are approximately 25 % for 1st order and 4 % for 4th-order low-pass filters [57], in line with previous studies [58]. A low-pass filter with f_c between 0.1 and 10 kHz thus fits 18 JNDs in case of 1st order and 100 JNDs in case of 4th order.

3. MULTISENSORY DISCRIMINATION OF SIZE AND ASPECT RATIO

Incorporating the perceptual aspects from above, we designed an experiment to explore the perception of size and shape of rectangular plates in an ecological scenario of percussion. We want to investigate to what extent participants are able to distinguish between size (here in the form of length) and shape (here in the form of aspect ratio), and additionally, with what precision participants are able to estimate the size and shape of rectangular plates of different materials. The technical setup is based on the AltAR/table platform [59].[5] During the experiment, participants directly interacted with the interface plate and identified an unknown plate's length and aspect ratio in direct comparison to a reference plate with known dimensions. The experiment was performed separately for the two materials glass and wood.

3.1. Stimuli and apparatus

The model parameters of the rendered plates are given in Tab. 1. Length and aspect ratio took 7 levels each, evenly spaced on a logarithmic scale, including 4 distractor levels for concealing the discrete levels and total range. The resulting frequency ratios are depicted in Fig. 2. Both length and aspect ratio were jittered randomly to pretend an interval scale and to prevent participants from simply remembering distinct sounds. While main levels were jittered uniformly within $\pm$30 % of the logarithmic increment between adjacent levels, distractor levels were jittered within $\pm$40 %, with the exception of the outmost distractor levels which were only jittered inwards. Viscous damping α_f was tuned by ear to equalize the overall decay across materials and to shift the unfamiliar low damping of freely vibrating plates to a more ecologically valid range. The effect of radiation efficiency was attenuated by taking its 4th root in order to model a more ecologically valid near-field behavior.

In an acoustically treated room, participants sat in front of the table and interacted with the interface plate through a ballpoint pen that was equipped with infrared markers of an OptiTrack motion capture system (see Fig. 3). Tracking data was recorded for later analysis. Interaction was only possible within an active region of 297 mm $\times$ 59.4 mm which equaled the dimensions of the smallest plate that was modeled during the experiment. Within this region, a paper overlay of the same dimensions was placed. The experiment software was implemented in Pd, ran on a separate computer, and sent control data to the auditory augmentation system. A graphical representation was shown on screen; however, participants responded by using a Behringer BCF2000 MIDI controller. The motorized sliders copied those on screen.

[5] Demo video of AltAR/table:
`https://phaidra.kug.ac.at/o:126460`

DOI: 10.5281/zenodo.7552234

ɔ. Procedure and participants

e experiment was structured into 4 parts. First, participants fa-liarized with the influence of excitation position on the resulting ɹnd by tapping on prepared physical plates of glass and wood. en, participants took a seat at the augmented table. During pas-e training, the effect of length and aspect ratio was demonstrated playing pre-recorded sounds through the interface. Both param-rs stepped through the 7 levels, i.e., the whole range of the slider, ɯ low to high and back, while the other parameter was set to a ɹstant medium value, respectively. Participants listened to all 4 ɯbinations of parameter and material at least once.

Parts 3 (active training) and 4 (test) shared a similar procedure, ɕh less trials for training. Active training (part 3) and the actual t (part 4) were performed separately for both materials (glass d wood), with balanced order across participants. At the start of ɕh trial, the augmentation was set to the reference plate A whose ɪgth and aspect ratio were given by two sliders. Participants ɹld switch to the unknown plate B through a button. The state length and aspect ratio was copied from the reference plate at rt. Participants were asked to identify the parameter that differed ːween A and B, and set the corresponding slider to an estimated ue. If one slider was moved, the other was instantly reset to ː value of the reference plate to ensure an answer in only one ːameter. Participants could change between both plates at will fore submitting their answer and proceeding to the next trial. participants decided for the wrong parameter, the background lor switched to red, and participants were asked to correct their lgment. After responding in the correct parameter dimension, the ɕkground switched back to green and the next trial was presented.

To speed up the experiment, the unknown plate B of a trial yed as reference plate A in the next trial. This included the reso-ion of the previous trial as feedback, and let participants directly ɔceed to the unknown plate B as they were already familiar with The test was organized in a series of 48 trials per material, which ɯed a trajectory through the 2D parameter space where only e parameter changed between successive trials (i.e., between A d B). Active training included only 8 trials per material. The jectories were pre-computed in Matlab so that each of the 9 com-ɪations of main levels was reached as the unknown plate by all 4 ɯbinations of main levels that were possible for the correspond-ː reference plate. This led to a total of 36 main trials per material. addition, distractor trials were generated which appeared always pairs so that a parameter changed to a random distractor level d then returned back to a random main level. Six such pairs distractor trials were inserted at random positions within the jectory, but exactly once in a row of 9 main trials. There was ways at least one main trial between two pairs of distractor trials.

A total of 14 participants (8 female, 6 male) were recruited to ɯ a diverse mix of experts (4 colleagues, 4 graduate students in ɹnd design) and non-experts (4 undergraduate students, family ːmbers, and friends). They received no compensation for their ːticipation; all reported normal hearing.

ɜ. Results

r sonification, parameter confusion and direction confusion are ɜumed to be the most critical. The respective accuracies are ɔtted against each other in Fig. 4. Average accuracies are 0.888 $D = 0.093$) for direction confusion, 0.818 ($SD = 0.065$) for el confusion (confusion between the 3 parameter levels), and ′63 ($SD = 0.100$) for parameter confusion. Participants 7, 9, and

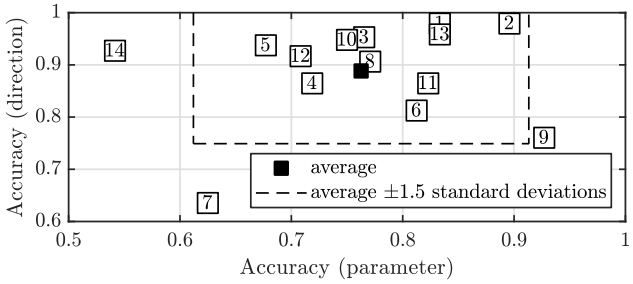


Figure 4: *The individual participants' performance: accuracy in direction (increase vs. decrease) and parameter (length vs. aspect ratio) discrimination.*

14 (all more than 1.5 standard deviations away from the average in at least one of the 3 categories of accuracy) are considered as outliers and are excluded from further analysis.

A fundamental task for the participants was to decide which of the two parameter dimensions (length or aspect ratio) had changed between A and B. Each trial can be attributed to one of 4 fields in the confusion matrix of true parameter and selected parameter. Overall accuracies (i.e., probabilities for choosing the correct parameter) were 0.75 for glass and 0.81 for wood.

For comparison between main levels of length and aspect ratio, we exclude trials with a distractor as plate B, take only the last answer into account (either correct or corrected), and round jitter-corrected estimated values to the nearest main level. The overall accuracy for length in case of wood is a bit higher ($Acc = 0.87$) than for all other combinations (between 0.83 and 0.84).

As we cannot assume normally distributed data, Mann-Whitney U tests were used for pairwise comparisons between median values. If not stated differently, any given p-values are Bonferroni-Holm adjusted, and a threshold for statistical significance of $5\,\%$ is employed. A larger value of true aspect ratio always led to a significantly larger estimated aspect ratio. In particular, longish and bar-shaped were perceived significantly more elongated than compact, and bar-shaped was perceived significantly more elongated than longish (all $p < 0.001$, respectively). Length follows the same trend: medium and large were perceived significantly larger than small ($p < 0.001$, respectively, for both materials), and large was perceived significantly larger than medium ($p = 0.026$ for glass, $p < 0.001$ for wood). For the discrimination of the direction (i.e., sign) of a parameter change, accuracies were generally high. In particular, accuracy for length in case of wood (0.97) was significantly higher than for the other combinations (0.89 and 0.92).

Figure 5 shows estimated lengths and aspect ratios, pooled over participants. The second main level (medium length or longish aspect ratio), can be reached by either a parameter increase (upward jump) or decrease (downward jump), if only main levels are considered. First and third main level can be reached from only one direction, but with different step size (one or two levels). For wood, the judgments of length following an increase were significantly larger than those following a decrease ($p < 0.001$), which leads to hysteresis. In all other cases, the difference between increase and decrease towards the unknown plate was not significant.

A sonification designer might seek to know the number of discriminable levels for each parameter. It obviously depends on the amount of error we tolerate. Such a function is derived via the effect size Cohen's d (see also [60]). For both steps between adjacent main levels, d is calculated, and the number of discriminable levels

DOI: 10.5281/zenodo.7552234

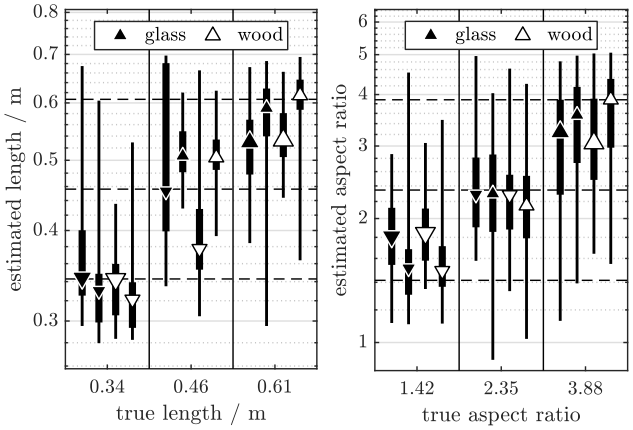


Figure 5: *Estimated vs. true value for all combinations of material and parameter. Triangles are medians, error bars are* 5*-,* 25*-,* 75*-, and* 95*-percentiles.* △ *= increase by 1 level,* △ *= increase by 2 levels,* ▽ *= decrease by 1 level, and* ▽ *= decrease by 2 levels. Dashed lines are the true values.*

is obtained with respect to a desired threshold d_t:

$$D = \frac{\sum_{i=1}^{2}(d_{i,i+1})}{d_\mathrm{t}} + 1 \ , \quad d = \sqrt{2}\Phi^{-1}(P_\mathrm{s}) \ . \tag{4}$$

Via the cumulative standard normal distribution Φ, any d can also be expressed in terms of probability of superiority P_s, i.e., the probability that a larger true value leads to a larger estimated value [61]. The resulting estimates are shown in Fig. 6.

The participants' answers can be interpreted as a confusion between the estimated plate and the true plate. If only main levels are considered, and estimated values are rounded to main values, each pair of true and answered plate is interpreted as a confused pair. Based on the frequency for each pair (the order of reference and unknown plate doesn't matter), a probability of confusion is constructed. The top 10 most confused pairs of plates exhibit a confusion probability larger than $5\,\%$ and differ in area by less than factor 3. We could observe that plates of equal area are likely to be confused. This may be attributed to the assumption that participants sometimes tend to answer in terms of area instead of the demanded parameter. Note that any pair of the four parameters length, width, aspect ratio, and area is sufficient to describe the 2D plate dimensions. If estimated areas are correlated with the true areas of the modeled plates, we could observe that judgments in the correct parameter dimension (glass: $R^2 = 0.69$, wood: $R^2 = 0.80$) led to much stronger agreement with the true area than those judgments where participants changed the wrong parameter (glass: $R^2 = -0.21$, wood: $R^2 = 0.17$).[6]

Although all participants received the same introductions, including the recommendation to tap everywhere within the tapping region, they developed quite different tapping strategies to explore the rendered plates. Especially interesting were the patterns of participants 6, 8, and 11, who independently from each other concentrated on three distinct spots for tapping. As their performance was pretty average (see Fig. 4), no conclusions can be drawn about possible benefits or drawbacks of this strategy. On average, participants took $15.0\,\mathrm{s}$ to form an answer, measured from the time they first switched to plate B.

[6]A negative R^2 in this case means that true areas perform worse than their geometric mean when predicting estimated areas.

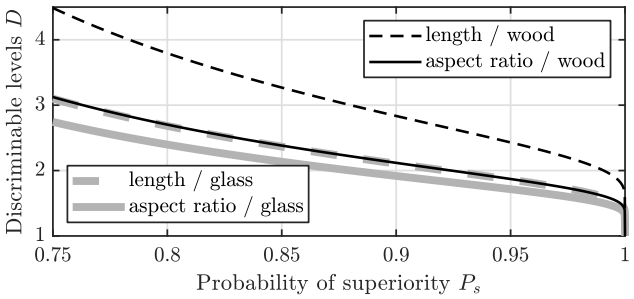


Figure 6: *The number of discriminable levels, plotted against the probability of superiority.*

3.4. Discussion

The two parameters length and aspect ratio are mainly connected to absolute frequency factor (and thus pitch) and relative frequency ratios between modes (and thus intervals and modal density), respectively. The range of length corresponds to a frequency ratio of 1.77 ($0.607/0.343$, at constant aspect ratio). If mode $3/0$ is interpreted as base frequency, then the whole frequency range, averaged across materials, takes about 558 JNDs.

The range of aspect ratio defines a ratio of frequency ratios, i.e., the frequency ratio between modes $n/0$ and $0/n$. Between bar-shaped and compact, it equaled 2.74. Some of the lower resonant frequencies are more than $10\,\%$ apart from each other and can thus be perceived as individual pitches [54]. This way, the range of intervals holds about $\log_{1.012}(2.74) = 84.5$ JNDs. The ratio between modal densities of bar-shaped and compact aspect ratios is almost constant across lengths and materials, on average 2.22. The range of modal densities that is given through the ratio of aspect ratios fits about $\log_{1.3}(2.22) = 3.0$ JNDs. This equals the about 3 discriminable levels of aspect ratio at $P_\mathrm{s} = 0.75$. It seems likely that participants exploited mainly modal density when judging aspect ratio. Figure 2 visualizes the frequency ratios (relative to the frequency of mode $2/0$ for a quadratic plate) as a function of aspect ratio, for isotropic materials. For orthotropic plates, the effective aspect ratio is divided by Ω ($=2$ in our case). If the lowest modes cross each other, their frequency ratios are ambiguous. In case of glass ($\Omega = 1$), modes $1/1$ and $2/0$ cross just above the lowest main level (compact). In case of wood ($\Omega = 2$), this effect is even stronger. Surprisingly, the accuracies of level and direction discrimination of aspect ratio did not significantly differ between materials. Due to crossing modes, aspect ratio may also significantly affect pitch, i.e., perceived length. One may argue that the lower modes are anyway barely radiated, especially for bar-shaped plates, so that only higher modes are evaluated for estimating the aspect ratio. The radiation efficiency can be roughly approximated by a 1st-order high-pass filter with cutoff at the critical frequency f_cr of radiation damping [14, 38]. It equals $1162\,\mathrm{Hz}$ for glass and $1763\,\mathrm{Hz}$ for wood. As mode $3/0$ is mostly below this frequency (see above), participants are barely able to utilize interval relationships of single modes for estimating aspect ratios. The slightly better performance in parameter identification for wooden plates might even be attributed to this attenuation of lower modes which otherwise confound the perceptual descriptors of length and aspect ratio. While inaudible lower partials can be reconstructed computationally by matching a theoretical model of a rectangular plate [15], participants failed to

DOI: 10.5281/zenodo.7552234

Table 2: *Model coefficients of the rendered plates in experiment 2.*

		non-metal			metal			
		plastic	wood	glass	gold	brass	aluminum	
thickness	h	8.557	10.602	8.000	6.659	7.105	7.707	mm
density	ρ	1150	590	2550	19 300	8500	2700	$\mathrm{kg\,m^{-3}}$
Young's modulus	E	3.20	3.29	66.90	80.00	95.00	72.00	GPa
Poisson's ratio	ν	0.300	0.100	0.250	0.423	0.330	0.340	
upper cutoff frequency (indentation hardness)	f_{cH}	9.37	6.67	>20	7.82	14.68	9.60	kHz
thermoelastic constants [14, 37]	$R_{1\mathrm{t}}$	64.31	22.42	24.84	64.31	22.42	24.84	$10^{-3}\,\mathrm{rad\,m^2\,s^{-1}}$
	$c_{1\mathrm{t}}$	1.251	0.489	0.977	1.251	0.489	0.977	$10^{-3}\,\mathrm{rad\,s^{-1}}$
metallicity [17]	H		0			1		
viscoelastic loss factor [36, 62]	η_{v}		$5.7/c_{\mathrm{L}}$			$0.57/c_{\mathrm{L}}$		
longitudinal wave velocity	c_{L}	1748.7	2373.6	5290.0	2246.9	3541.5	5491.1	$\mathrm{m\,s^{-1}}$
length	l_x			0.420				m
aspect ratio	r_{a}			$\{2, 4, 8\}$				

evaluate this information.

In summary, participants encountered several difficulties when performing the demanded task. A follow-up experiment is therefore designed, based on these considerations.

4. AUDITORY DISCRIMINATION OF MATERIAL AND ASPECT RATIO

We have learned from the first experiment that size and aspect ratio can indeed be employed as carrier parameters of a 2D auditory display. Aspect ratio and orthotropy, however, confound each other and their parameter range should be carefully chosen to avoid crossing partials. According to Fig. 2, $r_{\mathrm{a}} \geq 2$ seems appropriate. Instead of the length (i.e., surface area) from experiment 1, we try another approach based on material perception, using a 3D parameter space of aspect ratio, metallicity, and density. As density mainly affects pitch, it seems convenient to employ a more high-level meta-parameter such as longitudinal wave velocity c_{L} for this purpose — we call it rigidity to underline its physical meaning. Due to the common effect on pitch, length is set constant. While the participants of experiment 1 actively explored the model plates through a physical interface, experiment 2 tested pure auditory perception of sounds based on the same physical model.

4.1. Stimuli

While all three meta-parameters affect the plate on a continuous interval scale, only discrete levels are used during the experiment. Metallicity takes 2 levels (non-metallic, metallic). Aspect ratio takes 3 levels (compact, longish, and bar-shaped). Rigidity takes 3 levels, with labels for individual material categories that depend on the state of metallicity. For non-metals, these are plastic, wood, and glass. For metals, these are gold, brass, and aluminum. Metallicity blends between the pair of non-metal and metal material of same rigidity.

For blending between non-metals and metals, we chose pairs of materials with approximately equal longitudinal wave velocity c_{L} (and thus base frequency). c_{L} can be only perceived in combination with thickness h [15]; their product hc_{L} scales the overall frequency. We therefore use the freely adjustable thickness to align selected material categories with their corresponding value of c_{L} to an equally-spaced grid of base frequencies (on a logarithmic scale). Glass is selected as reference with a plausible thickness of $8\,\mathrm{mm}$ for a table. Rigidity levels below are tuned to 0.75 and 1.5 octaves below. Metals are equalized likewise.

Length is set constant to $0.42\,\mathrm{m}$, large enough for a table, but sufficiently high in pitch. Aspect ratio (length : width) ranges from 2 (compact) via 4 (longish) to 8 (bar-shaped). A plausible physical interpretation might be that the table is made from many narrow planks in the latter case. Excitation position is set constant, on the edge of the plate, so that as many modes are excited as possible[7], but in the maximum of mode $3/0$, at normalized position $[0.3083, 0]$ in order to stabilize pitch. Possible disturbing modes ($1/1$ and $2/0$) are thereby attenuated while mode $3/0$ is boosted.

In the experiment, metallicity takes only the extreme values 0 (non-metallic) and 1 (metallic). In addition to damping, H blends between non-metallic and metallic material constants on an exponential scale. The actual constants are given in Tab. 2. Thermoelastic constants $R_{1\mathrm{t}}$ and $c_{1\mathrm{t}}$ are pre-computed via the underlying physical constants [14]. We argue that almost any frequency-dependent damping higher than that of the ideal free plate can be achieved by physical suspension or dampers. The decay factors are therefore equalized by an individual (positive) bias so that mode $3/0$ matches a desired (shorter) decay time. Mode $3/0$ roughly controls pitch and is usually the longest decaying mode if those below are neglected due to excitation in its maximum. Within the experiment, 2 decay times T_{60} of $0.15\,\mathrm{s}$ and $0.45\,\mathrm{s}$ are used as different conditions.

Stimuli are pre-rendered by using the a physical model similar to the ones described by [14, 16, 15, 17]. As excitation signal for a single impact, a Hann window of constant length $0.5\,\mathrm{ms}$ is used. The duration is chosen to perceptually match the pen excitation through a paper overlay in experiment 1. Each stimulus consists of the same model plate that is excited by 4 successive impacts in time intervals of $200\,\mathrm{ms}$. A natural rhythm is achieved by randomization of the onset time by $\pm 10\,\mathrm{ms}$. All combinations of levels of metallicity, rigidity, and aspect ratio lead to 18 model plates for each of the 2 damping conditions, or 36 different stimuli in total. To prevent participants from directly memorizing individual sounds, the 3 meta-parameters (normalized between 0 and 1) are jittered uniformly by ± 0.05. Each individual impact is jittered by $\pm 3\,\mathrm{dB}$ in amplitude and $\pm 10\,\%$ in duration. Each of the 36 stimuli is rendered in 4 different variations, two for training and two for the two repetitions in the test.

4.2. Apparatus, procedure, and participants

The experiment was implemented in form of a web page, with the help of the open-source JavaScript library `jsPsych` which is designed for running cognitive experiments online in the web browser [63]. The experiment was hosted free of charge on the

[7] Note that free boundary conditions are used on all edges.

DOI: 10.5281/zenodo.7552234

related Cognition[8] platform. Its source code including the rendered sound files is available online.[9]

Participants had to indicate if they would classify themselves as trained listeners (e.g., due to musical training or professional background). In addition, they were asked to put on the best headphones available. During a passive introductory phase, all parameter dimensions were explained individually with the help of sound examples. During an active training phase, participants compared all 18 combinations within the current damping condition via two 9×9 matrices of play buttons for non-metals and metals, respectively. They were free to decide when ready to start the test for this condition. Both conditions (each consisting of active training and subsequent test) were presented one after another in randomized order. Each test contained 2 repetitions of 18 stimuli in random order. The possibility to take a break was announced in between. During the test, they had to identify the given stimuli in all three parameter dimensions, with the possibility of replay.

20 anonymous participants (recruited among students and personal acquaintances) finished the experiment. As remuneration, a lottery of 2 vouchers worth 50 EUR each was initiated. 14 of them classified themselves as trained listeners.

4.3. Results

From all stimuli, 16 % were correctly identified in all three parameter dimensions (95 % confidence interval CI_{95} between 13 % and 18 %). Note the chance level of $1/18 = 5.6\,\%$. Damping had no significant effect on this result. As the number of trials was balanced across classes, this simple metric of percent correct classifications, i.e., accuracy, is an appropriate measure of the participants' performance. When discriminating between non-metal and metal, pooled accuracy was 0.80. For rigidity (plastic/gold vs. wood/brass vs. glass/aluminum), pooled accuracy was 0.55. For aspect ratio (compact vs. longish vs. bar-shaped), pooled accuracy was 0.36. As the participants' individual pooled accuracies spread rather symmetrically around their average, with low standard deviation (0.07 for metallicity, 0.12 for rigidity, and 0.06 for aspect ratio), none were excluded as outliers. Neither the average trial duration nor the average number of replays had a significant effect on the average accuracy. Trained participants achieved a slightly (but not significantly) higher accuracy than untrained participants (0.61 vs. 0.57).

One-tailed Wilcoxon signed-ranks tests were used for pairwise comparisons between adjacent levels of rigidity and aspect ratio, taking ordinal values between 1 and 3, with a 5 % threshold for statistical significance. For rigidity, level 2 (wood/brass) was judged significantly higher than level 1 (plastic/gold) ($Z = 30142$, $p < 0.001$), and level 3 (glass/aluminum) was judged significantly higher than level 1 (wood/brass) ($Z = 50806$, $p < 0.001$). For aspect ratio, level 2 (longish) was judged significantly higher than level 1 (compact) ($Z = 31539$, $p < 0.001$), but the difference between level 3 (bar-shaped) and 2 (longish) was not significant. Metallicity included only 2 categorical levels (1 and 2). According to Fisher's exact test, metals were judged significantly different to non-metals ($p < 0.001$, odds ratio: 15.6).

In special cases, even two different plates may sound similar. For each combination of two stimuli, the probability that one was identified as the other was computed. The 16 most confused pairs had confusion probabilities above 0.13. 12 of them involve confusion between aspect ratios (top 5: longish vs. bar-shaped), 12 involve a bar-shaped plate, and 5 refer to rigidity confusions.

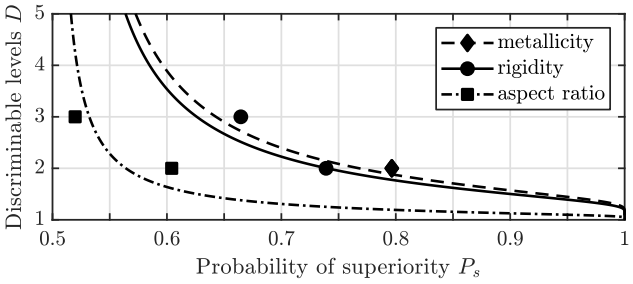


Figure 7: *The number of discriminable levels of the three parameters (metallicity, material, aspect ratio), plotted against the probability of superiority P_s. Markers show P_s computed from raw responses. Lines show estimates based on effect size.*

When discriminating non-metals and metals, a significantly higher accuracy (0.82) was achieved for weakly damped plates than for strongly damped plates ($Acc = 0.77$, $CI_{95} = [0.74, 0.80]$). This is mainly attributed to false negatives for metal in case of strong damping. While accuracy was equally high for plastic/gold and wood/brass (0.91), it was significantly lower for glass/aluminum ($Acc = 0.57$, $CI_{95} = [0.53, 0.62]$). While accuracy of metallicity identification was equally high for compact and longish plates (0.81 and 0.82, respectively), it was significantly lower for bar-shaped plates ($Acc = 0.76$, $CI_{95} = [0.72, 0.80]$).

Concerning rigidity identification, participants were significantly better in discriminating non-metals ($Acc = 0.62$) than metals ($Acc = 0.49$, $CI_{95} = [0.45, 0.52]$). While discrimination between rigidities was equal for longish and bar-shaped plates (on average, $Acc = 0.57$), accuracy was lower for compact plates (0.51, $CI_{95} = [0.47, 0.56]$).

When estimating aspect ratio, despite low accuracy, the answers were significantly different from chance (weak damping: $\chi^2(4) = 82.7$; strong damping: $\chi^2(4) = 32.3$; both $p \leq 0.001$). While accuracy was similar for compact and longish plates (on average: 0.38), it was significantly lower (equal to chance) for bar-shaped plates (0.32, $CI_{95} = [0.28, 0.36]$).

Other than in experiment 1, the responses are only ordinal. Nevertheless, we can get a rough estimate of the number of discriminable levels as a function of the probability of superiority, based on Cohen's d (see Fig. 7). In addition, we have one reference data point for each parameter: the probability of superiority that is achieved for the number of levels that were tested in the experiment. In case of 2 levels (metallicity), P_s is equal to the accuracy. In case of 3 levels, P_s of both adjacent pairs are averaged. For material and aspect ratio, the responses are additionally reduced to two levels, in order to also obtain P_s for that case. This can be done in two ways, by combining either levels 1 and 2, or 2 and 3, referring to different decision thresholds. Assuming that participants choose the best decision threshold, the maximum of both values is selected. The resulting reference values are marked in Fig. 7.

4.4. Discussion

Combined identification of metallicity, rigidity, and aspect ratio was demanding for the participants. For metallicity and rigidity identification, accuracy was acceptable (0.80 and 0.55, respectively),

[8] Cognition platform: `https://www.cognition.run`

[9] Experiment source code: `https://github.com/m---w/experiment_listening_to_rectangular_plates`

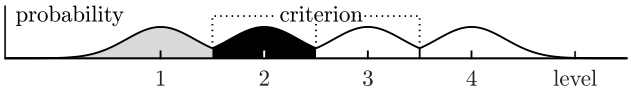


Figure 8: *Gaussian model for estimating information capacity.*

but incorporated a negative effect of damping. Metallicity is perceptually expressed through (a) damping of some lower partials due to thermoelasticity, and (b) the amount of frequency-dependent damping due to viscoelasticity. (a) is almost completely masked by the small amount of damping that is included, even in the condition with weak damping, while (b) affects only higher modes beyond the critical frequency. Participants could hardly discriminate between glass and aluminum ($Acc = 0.57$) while plastic/gold and wood/brass discrimination was excellent ($Acc = 0.91$). This may be attributed to the high base frequency of glass and aluminum plates, leading to a small number of audible partials due to strong damping above the critical frequency. Overall accuracy for aspect ratio discrimination was only slightly better than chance (0.36). However, it was still better for plastic/gold and wood/brass (0.36) than for glass/aluminum (0.32). The latter case was not significantly different from chance performance. Similar to metallicity, the low number of audible partials made it difficult to judge aspect ratio. The results for the 3 parameters show that an actual sonification would require different parameter ranges and segmentations. While experiment 1 yielded similar accuracy in both dimensions, aspect ratio identification in experiment 2 is unacceptable for sonification. We assume, however, that a reduced parameter range in combination with decimation to 2 levels per parameter would yield almost perfect identification, with theoretical information capacity of 3 bit.

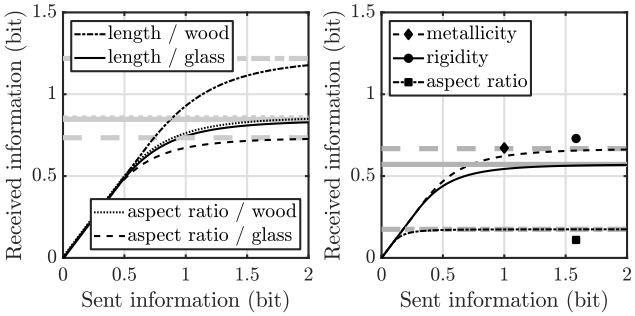
5. INFORMATION IN AUDITORY AUGMENTATIONS

In order to compare the results with others from the literature, we need to transform them to a common domain based on information theory. In the literature, the information capacity of an auditory display is obtained by performing the same experiment with different resolutions, i.e., numbers of discrete levels that partition the total parameter range [12, 10]. Contrary to this straightforward approach, we performed our experiments only at one single partitioning for each parameter — obviously not enough for fitting a curve and finding a maximum. Under some assumptions of signal detection theory, however, we are able to predict the results of other parameter partitionings, based on measured effect sizes.

Similar to before, we assume a normal distribution of parameter estimations on a continuous scale, centered around the true value, with equal variances. This leads to overlapping normal distributions, as visualized for 4 levels in Fig. 8. We further assume a perfect decision criterion midway between adjacent levels. The tails of the distributions that exceed the criteria thus quantify wrong answers, i.e., identifications as different level. The probability of correct identification is the area below the whole probability distribution ($=1$) minus the exceeding tail(s). For the lowest and highest level (P_1 in Eq. 5, gray region in Fig. 8), it is larger than that of the in-between levels (P_2 in Eq. 5, black region):

$$P_1 = 1 - \Phi(-|d|/2) \quad , \quad P_2 = 1 - 2\Phi(-|d|/2) \quad . \tag{5}$$

Assuming that all levels occur equally often (as in our experiments) the weighted average of probabilities, for the given number of levels



(a) experiment 1 *(b) experiment 2*

Figure 9: *Received vs. sent information per parameter.*

D, yield the probability of correct identification:

$$P_\mathrm{c} = \frac{(D-2)P_2 + 2P_1}{D} \quad . \tag{6}$$

Under the given assumptions, for $D = \{2, 3, ...\}$, the computed values for P_c are exact. The number of levels D actually represents the amount of transmitted information $I_\mathrm{sent} = \log_2(D)\,\mathrm{bit}$. The amount of received information I_rec is thus:

$$I_\mathrm{rec} = \log_2(D \cdot P_\mathrm{c})\,\mathrm{bit} = I_\mathrm{sent} + \log_2(P_\mathrm{c})\,\mathrm{bit} \quad . \tag{7}$$

Received vs. sent information is plotted in Fig. 9 for both experiments. While the underlying effect sizes of experiment 2 are only approximate, the actual measurements marked in the plot suggest a good fit. For the 2D sonification in experiment 1, we achieve an information capacity of about 1.3 bit for length and 0.9 bit for aspect ratio, in case of a wooden plate. The total information capacity with two dimensions A and B and confusion accuracy Acc_{AB} is

$$I_\mathrm{rec} = I_{\mathrm{rec},A} + I_{\mathrm{rec},B} + \log_2(Acc_{AB})\,\mathrm{bit} \quad . \tag{8}$$

In case of wood, we obtain 1.9 bit. For experiment 2, parameter confusion is already contained in the estimate effect size. The 3D information capacity is the sum of the individual dimensions, i.e., 1.5 bit. How good is that in comparison to the literature? We didn't even reach the 2.5 bit of 1D pitch or amplitude identification, but that was not expected anyway, due to our choice of parameters. Even a 1D auditory augmentation based on virtual room acoustics yielded only 2 bit [60]. The logical consequence is to circumvent the hard limit of 1D displays by adding more dimensions. Pollack and Ficks already took it to the extreme with an 8D display that used 8 independent sound parameters of 2 levels or 1 bit resolution each [13]: listeners perceived about about 7 bit of the 8 bit sent.

6. CONCLUSIONS

Our approach to use two or three sound parameters of low resolution was already the right choice, but not yet enough. Note that the physically-inspired parameter dimensions partly interfere with each other, so that participants were possibly forced to distinctly remember every single parameter combination. Furthermore, we investigated only absolute identification without reference, similar to absolute pitch perception. Even a lower information capacity would therefore seem acceptable for interactive sonifications, as most information is conveyed relatively, as a change over time. Absolute identification of the extreme levels in each parameter within a multi-dimensional sonification may then serve as guidance to facilitate the correct identification of relative parameter changes.

DOI: 10.5281/zenodo.7552234

7. REFERENCES

[1] Till Bovermann, René Tünnermann, and Thomas Hermann, "Auditory Augmentation," *Int. J. Ambient. Comput. Intell.*, vol. 2, no. 2, pp. 27–41, 2010.

[2] Katharina Groß-Vogt, Marian Weger, and Robert Höldrich, "Exploration of Auditory Augmentation in an Interdisciplinary Prototyping Workshop," St. Pölten, Austria, 2018, pp. 10–16.

[3] Marian Weger, Thomas Hermann, and Robert Höldrich, "Plausible Auditory Augmentation of Physical Interaction," in *International Conference on Auditory Display (ICAD)*, Houghton, Michigan, 2018, pp. 97–104.

[4] William W. Gaver, "Synthesizing auditory icons," in *Conference on Human factors in computing systems (CHI)*, Amsterdam, The Netherlands, 1993, pp. 228–235, ACM SIGCHI.

[5] Stephen Barrass, "Digital Fabrication of Acoustic Sonifications," *J. Audio Eng. Soc.*, vol. 60, no. 9, pp. 7, 2012.

[6] Thomas Hermann and Helge Ritter, "Listen to your data: Model-based sonification for data analysis," *Advances in intelligent computing and multimedia systems*, 1999.

[7] Bradley S. Mauney and Bruce N. Walker, "Creating functional and livable soundscapes for peripheral monitoring of dynamic data," in *International Conference on Auditory Display (ICAD)*, Sydney, Australia, 2004.

[8] Sam Ferguson, "Sonifying every day: activating everyday interactions for ambient sonification systems," in *International Conference on Auditory Display (ICAD)*, Łódź, Poland, 2013, pp. 77–84.

[9] René Tünnermann, J. Hammerschmidt, and Thomas Hermann, "Blended sonification–sonification for casual information interaction," in *International Conference on Auditory Display (ICAD)*, Lodz, Poland, 2013.

[10] Irwin Pollack, "The Information of Elementary Auditory Displays," *J. Acoust. Soc. Am.*, vol. 24, no. 6, pp. 745–749, 1952.

[11] Irwin Pollack, "The Information of Elementary Auditory Displays. II," *J. Acoust. Soc. Am.*, vol. 25, no. 4, pp. 765–769, 1953.

[12] George A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information.," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[13] Irwin Pollack and Lawrence Ficks, "Information of Elementary Multidimensional Auditory Displays," *J. Acoust. Soc. Am.*, vol. 26, no. 2, pp. 155–158, 1954.

[14] Antoine Chaigne and Christophe Lambourg, "Time-domain simulation of damped impacted plates. I. Theory and experiments," *J. Acoust. Soc. Am.*, vol. 109, no. 4, pp. 1422–1432, 2001.

[15] Martin Czuka, Marian Weger, and Robert Höldrich, "Klangsynthese und akustische Erkennung rechteckiger Platten," in *DAGA Jahrestagung für Akustik*, Vienna, Austria, 2021.

[16] P. Troccaz, R. Woodcock, and F. Laville, "Acoustic radiation due to the inelastic impact of a sphere on a rectangular plate," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2197–2202, 2000.

[17] S. McAdams, V. Roussarie, A. Chaigne, and B. L. Giordano, "The psychomechanics of simulated sound sources: Material properties of impacted thin plates," *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. 1401–1413, 2010.

[18] R. A. Lutfi and C. N. J. Stoelinga, "Sensory constraints on auditory identification of the material and geometric properties of struck bars," *J. Acoust. Soc. Am.*, vol. 127, no. 1, pp. 350–360, 2010.

[19] R. A. Lutfi, E. Oh, E. Storm, and J. M. Alexander, "Classification and identification of recorded and synthesized impact sounds by practiced listeners, musicians, and nonmusicians," *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 393–404, 2005.

[20] J. Traer, M. Cusimano, and J. H. McDermott, "A perceptually inspired generative model of rigid-body contact sounds," in *DAFx*, Birmingham, UK, 2019.

[21] Marian Weger, *Plausible auditory augmentation of physical interaction*, Ph.D. thesis, University of Music and Performing Arts, Graz, Austria, 2022.

[22] B. L. Giordano and S. McAdams, "Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 1171–1181, 2006.

[23] G. Lemaitre and L. M. Heller, "Auditory perception of material is fragile while action is strikingly robust," *J. Acoust. Soc. Am.*, vol. 131, no. 2, pp. 1337–1348, 2012.

[24] C. Carello, K. L. Anderson, and A. J. Kunkler-Peck, "Perception of Object Length by Sound," *Psychological Science*, vol. 9, no. 3, pp. 211–214, 1998.

[25] S. Tucker and G. J. Brown, "Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds," in *AES Convention*, Amsterdam, Netherlands, 2003.

[26] A. J. Kunkler-Peck and M. T. Turvey, "Hearing Shape," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 1, pp. 279–294, 2000.

[27] S. Lakatos, S. Mcadams, and R. Caussé, "The representation of auditory source characteristics: Simple geometric form," *Perception & Psychophysics*, vol. 59, no. 8, pp. 1180–1190, 1997.

[28] Z. Zhang, J. Wu, Q. Li, J. B. Tenenbaum, Z. Huang, and W. T. Freeman, "Shape and Material from Sound," in *Conference on Neural Information Processing Systems (NIPS)*, 2017.

[29] R. Martín, J. Iseringhausen, M. Weinmann, and M. B. Hullin, "Multimodal perception of material properties," in *ACM SIGGRAPH Symposium on Applied Perception*, Tübingen, Germany, 2015, pp. 33–40.

[30] Waka Fujisaki and Shin'ya Nishida, "Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception," *Experimental Brain Research*, vol. 198, no. 2-3, pp. 245–259, 2009.

[31] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu, and S. Nishida, "Audiovisual integration in the human perception of materials," *Journal of Vision*, vol. 14, no. 4, pp. 12–12, 2014.

[32] Waka Fujisaki, Midori Tokita, and Kenji Kariya, "Perception of the material properties of wood based on vision, audition, and touch," *Vision Research*, vol. 109, pp. 185–200, 2015.

DOI: 10.5281/zenodo.7552234

[33] M. E. McIntyre and J. Woodhouse, "On measuring the elastic and damping constants of orthotropic sheet materials," *Acta Metallurgica*, vol. 36, no. 6, pp. 1397–1416, 1988.

[34] G. B. Warburton, "The Vibration of Rectangular Plates," *Proceedings of the Institution of Mechanical Engineers*, vol. 168, no. 1, pp. 371–384, 1954.

[35] Lionel Zoghaib and Pierre-Olivier Mattei, "Damping analysis of a free aluminum plate," *Journal of Vibration and Control*, vol. 21, no. 11, pp. 2083–2098, 2015.

[36] Lothar Cremer, M. Heckl, and B. A. T. Petersson, *Structure-borne sound: structural vibrations and sound radiation at audio frequencies*, Springer, 3rd edition, 2005.

[37] Antoine Chaigne and Jean Kergomard, *Acoustics of Musical Instruments*, Springer, 2016.

[38] A. Putra and D. J. Thompson, "Sound radiation from rectangular baffled and unbaffled plates," *Applied Acoustics*, vol. 71, no. 12, pp. 1113–1125, 2010.

[39] S. Schedin, C. Lambourg, and A. Chaigne, "Transient sound fields from impacted plates: comparison between numerical simulations and experiments," *Journal of Sound and Vibration*, vol. 221, no. 3, pp. 471–490, 1999.

[40] Antoine Chaigne and Vincent Doutaut, "Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars," *J. Acoust. Soc. Am.*, vol. 101, no. 1, pp. 539–557, 1997.

[41] Eric Krotkov, "Robotic perception of material," in *Int. Joint Conference on Action and Perception*, 1995, pp. 88–95.

[42] E. Krotkov, R. Klatzky, and N. Zumel, "Robotic perception of material: Experiments with shape-invariant acoustic measures of material type," in *Experimental Robotics IV*, vol. 223, pp. 204–211. Springer, London, 1997.

[43] Z. Zhang, J. Wu, Q. Li, Z. Huang, J. B. Tenenbaum, and W. T. Freeman, "Inverting Audio-Visual Simulation for Shape and Material Perception.," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 2536–2538, IEEE.

[44] R. S. Schlauch, D. T. Ries, and J. J. DiGiovanni, "Duration discrimination and subjective duration for ramped and damped sounds," *J. Acoust. Soc. Am.*, vol. 109, no. 6, pp. 2880–2887, 2001.

[45] Sharon M. Abel, "Duration Discrimination of Noise and Tone Bursts," *J. Acoust. Soc. Am.*, vol. 51, no. 4B, pp. 1219–1223, 1972.

[46] Hanna Järveläinen and Tero Tolonen, "Perceptual Tolerances for Decay Parameters in Plucked String Synthesis," *AES Journal*, vol. 49, no. 11, pp. 1049–1059, 2001.

[47] Walt Jesteadt, Craig C. Wier, and David M. Green, "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.*, vol. 61, no. 1, pp. 169–177, 1977.

[48] D. W. Robinson and R. S. Dadson, "A re-determination of the equal-loudness relations for pure tones," *British Journal of Applied Physics*, vol. 7, no. 5, pp. 166–181, 1956.

[49] Hugo Fastl and Eberhard Zwicker, *Psychoacoustics: facts and models*, Springer, 3rd edition, 2007.

[50] C.N.J. Stoelinga and R.A. Lutfi, "Detection of missing modal frequencies," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3415–3415, 2008.

[51] Craig C. Wier, Walt Jesteadt, and David M. Green, "Frequency discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.*, vol. 61, no. 1, pp. 178–184, 1977.

[52] H.E. Gockel, B.C.J. Moore, R.P. Carlyon, and C.J. Plack, "Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency," *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 373–382, 2007.

[53] H.E. Gockel, R.P. Carlyon, and C.J. Plack, "Combining information across frequency regions in fundamental frequency discriminationa)," *J. Acoust. Soc. Am.*, vol. 127, no. 4, 2010.

[54] W.R. Thurlow and S. Bernstein, "Simultaneous Two-Tone Pitch Discrimination," *J. Acoust. Soc. Am.*, vol. 29, no. 4, pp. 515–519, 1957.

[55] D. A. Fantini and N. F. Viemeister, "Discrimination of frequency ratios," in *Auditory processing of complex sounds*, pp. 47–56. Routledge, 1987.

[56] Christophe N. J. Stoelinga and Robert A. Lutfi, "Discrimination of the spectral density of multitone complexes," *J. Acoust. Soc. Am.*, vol. 130, no. 5, pp. 2882–2890, 2011.

[57] Shari L. Campbell, "Uni- and multidimensional identification of rise time, spectral slope, and cutoff frequency," *J. Acoust. Soc. Am.*, vol. 96, no. 3, pp. 1380–1387, 1994.

[58] J. M. Pickett, R. L. Daly, and S. L. Brand, "Discrimination of Spectral Cutoff Frequency in Residual Hearing and in Normal Hearing," *J. Acoust. Soc. Am.*, vol. 38, no. 5, pp. 923–923, 1965.

[59] Marian Weger, Thomas Hermann, and Robert Höldrich, "AltAR/table: a platform for plausible auditory augmentation," in *International Conference on Auditory Display (ICAD)*, Virtual Conference, 2022.

[60] Katharina Groß-Vogt, Marian Weger, Matthias Frank, and Robert Höldrich, "Peripheral Sonification by Means of Virtual Room Acoustics," *Computer Music Journal*, vol. 44, no. 1, pp. 71–88, 2021.

[61] John Ruscio, "A probability-based measure of effect size: Robustness to base rates and other factors.," *Psychological Methods*, vol. 13, no. 1, pp. 19–30, 2008.

[62] Teruaki Ono and Misato Norimoto, "Study on Young's Modulus and Internal Friction of Wood in Relation to the Evaluation of Wood for Musical Instruments," *Japanese Journal of Applied Physics*, vol. 22, no. Part 1, No. 4, pp. 611–614, 1983.

[63] Joshua R. de Leeuw, "jsPsych: A JavaScript library for creating behavioral experiments in a Web browser," *Behavior Research Methods*, vol. 47, no. 1, pp. 1–12, 2015.

 DOI: 10.5281/zenodo.7552234