Multimodal Perception of Material Properties

Rodrigo Martín * Julian Iseringhausen* Michael Weinmann* Matthias B. Hullin*
University of Bonn

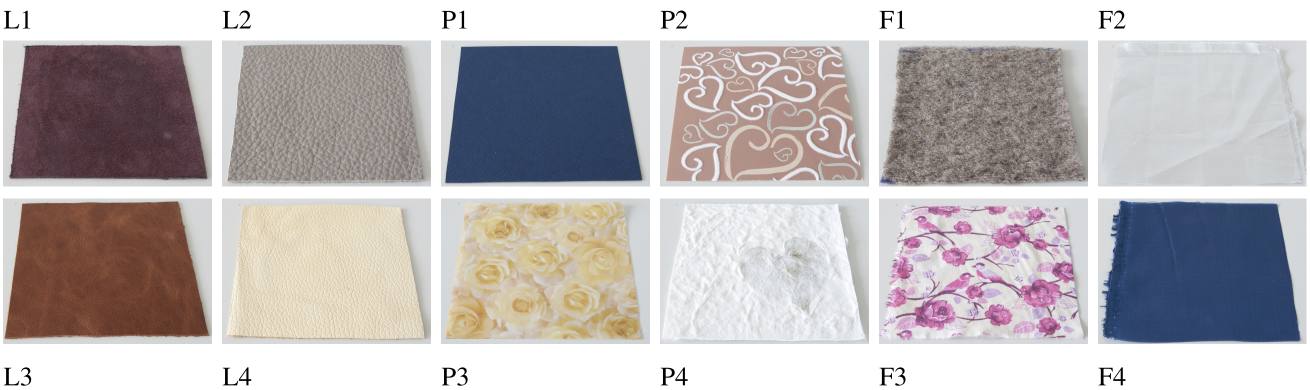


Figure 1: *Materials utilized in the experiment. Included are four leathers (L1 – L4), four papers (P1 – P4), and four fabrics (F1 – F4).*

Abstract

The human ability to perceive materials and their properties is a very intricate multisensory skill and as such not only an intriguing research subject, but also an immense challenge when creating realistic virtual presentations of materials. In this paper, our goal is to learn about how the visual and auditory channels contribute to our perception of characteristic material parameters. At the center of our work are two psychophysical experiments performed on tablet computers, where the subjects rated a set of perceptual material qualities under different stimuli. The first experiment covers a full collection of materials in different presentations (visual, auditory and audio-visual). As a point of reference, subjects also performed all ratings on physical material samples. A key result of this experiment is that auditory cues strongly benefit the perception of certain qualities that are of a tactile nature (like “hard–soft”, “rough–smooth”). The follow-up experiment demonstrates that, to a certain extent, audio cues can also be transferred to other materials, exaggerating or attenuating some of their perceived qualities. From these results, we conclude that a multimodal approach, and in particular the inclusion of sound, can greatly enhance the digital communication of material properties.

CR Categories: I.2.10 [Vision and Scene Understanding]: Perceptual reasoning; I.3.8 [Computer Graphics]: Applications.

Keywords: multisensory perception and integration, visual psychophysics

*{rodrigo|iseringhausen|mw|hullin}@cs.uni-bonn.de

SAP '15, September 13-14, 2015, Tübingen, Germany.
© 2015 ACM. ISBN 978-1-4503-3812-7/15/09…$15.00
DOI: http://dx.doi.org/10.1145/2804408.2804420

1 Introduction

Being able to perceive the materials that objects are made of, and their respective properties, is of utmost importance in our everyday human lives; yet, to this day we know very little about this skill. What makes material perception a fascinating and inexhaustible subject of investigation is that it is highly multimodal, or multisensory, by nature, combining vision, hearing, touch, smell and taste to varying extents. Consequently, recreating the intricate appearance of materials in a digital context is a very hard task. For example, even the most advanced models and methods from computer graphics have not yet managed to fully virtualize the material sampling process in product design; instead, physical samples are still the standard. In this paper, we build upon the assumption that a designer's decision for or against a material is not only based on measurable physical parameters but also on subjective or affective characteristics. Under this premise, effective communication of materials requires an understanding of how these characteristics are perceived multimodally.

The main contribution of this work are two psychophysical experiments performed to quantify the isolated and combined effect of visual and auditory stimuli on a set of material properties or qualities. This setting maps well to the capabilities of today's consumer devices, where 2D display and stereo audio are regularly available in high quality.

In our first experiment, participants rated 10 material qualities for a set of 12 different material samples, each in 3 different virtual presentations (visual, auditory, and audiovisual). Reference data was obtained by letting the subjects interact with a physical sample of each of the materials (full-modal interaction) and rating the same set of parameters. We investigated to which amounts the visual and auditory channels impact the different perceived material qualities. As a key result, we learned that the assessment of qualities that are of a tactile nature (such as “hard–soft” or “rough–smooth”) strongly benefits from auditory cues.

Following up on this insight, we performed a second experiment of similar design where images and sounds of different materials were combined. The main finding of this experiment was that by changing the auditory stimulus, the perception of the tactile qualities can be manipulated in a consistent manner. In fact, quite extreme changes can be achieved without compromising the overall realism of the experience.

From these results, we conclude that the digital presentation of materials can be improved by creating a multimodal experience. However, future research will be needed in order to explore the potential and limits of sound in material perception.

2 Related work

While the perception of objects, surfaces and color has been studied in great detail over the course of several decades, the study of material perception has gained momentum relatively recently. To this date, we still know very little about the processes that govern human perception of materials; as a consequence, applying such knowledge in the field of computation is not a straightforward procedure. For a high-level overview of problems and challenges in material perception, we refer to the excellent surveys by Adelson [2001], Maloney and Brainard [2010] and Fleming [2014].

The majority of the literature in the field is based on purely visual representations of the materials[1]. Several of these studies focused on understanding how humans perceive the luminance of the surfaces. For example, Adelson and Pentland [1996] examined the ability to judge the reflectance and the shading of the objects in three-dimensional scenes. Ho et al. [2006] researched the visual estimation of surface roughness, discovering that observers perceive surfaces to be rougher with decreasing illuminant angle. Visual perception of material glossiness has been also investigated in isolated form [Pellacini et al. 2000] and together with transparency [Cunningham et al. 2007]. Both works aimed to find perceptually meaningful reparameterizations for optical properties by exploring the relationships between physical parameters and the perceptual dimensions of glossy and transparent appearance. How the shape of materials influences the perception of reflectance properties has been analyzed by Vangorp et al. [2007]. Bouman et al. [2013] examined the human competence to estimate the stiffness and density of fabrics from video, in the context of predicting such features algorithmically. The interactions between the tasks of material classification and material judgment of a set of qualities in both the visual and semantic domains was investigated by Fleming et al. [2013]. Their studies revealed a high degree of consistency between these two assignments, suggesting that subjects access similar information about materials in both circumstances. Finally and in a similar way to our own dimensionality analysis, Rao and Lohse [1993] explored the dimensionality of a space of abstract visual textures, identifying three strong orthogonal directions.

Sound-only approaches to material perception are not very frequent, however there are some interesting studies. Carello et al. [1998] researched the capability to perceive the specific size of objects. This was one of the approaches that first addressed the assignment of judging geometrical properties of an object (length) using audition. The relation between material perception and variables that govern the synthesis of contact sounds was analyzed by Klatzky et al. [2000] and additionally by Avanzini and Rocchesso [2001]. Giordano and McAdams [2006] investigated the identification of materials from impact sounds. They shown that, while listeners performed well with respect to gross material categories, their performance degraded for materials within the same gross category. Lemaitre and Heller [2012] studied the human performance on identifying either the actions or the materials used to produce certain auditory stimuli. Also purely tactile approaches have been a matter of research, especially looking at the general dimensionality of the spaces underlying haptic interactions. In this way, Etzi et al. [2014] examined the nature of aesthetic preferences for tactile textures.

Material perception is multimodal by nature, and the interplay of the different sensing modalities is far from understood. Guest et al. [2002] explored how the tactile perception of textures can be modified by manipulating the frequency content of touch-related sounds. Tactile information has been also combined with visual stimuli by Baumgartner et al. [2013], who looked for correspondence between visual and haptic material representations, and Hope et al. [2013], who evaluated possible associations between physical and emotional material properties. Nevertheless, the combination that has gained more interest in material perception is the association of vision and sound. Bonneel et al. [2010] combined and analyzed levels of detail in audiovisual rendering; Fujisaki et al. [2014] researched the principles that govern cross-modal integration of material information. Fujisaki et al. [2015] also went a step further, investigating whether the same subjective classifications for vision, audition and touch can be found.

With this work, we aim to extend the state of the art in several regards. Starting on the frame of multimodal perception, we propose two experiments in which participants not only rated isolated and combined audiovisual stimuli, they also interacted and evaluated the physical material samples. This allows the subjects to obtain a full-modal experience. Our selection of materials covers three types or classes (leather, fabric, and paper), each composed of multiple members to represent the respective intra-class variances. The present study focuses on a set of perceptual properties that are fine-grained, strongly subjective, and not strictly aligned with class boundaries. Using vision and sound as virtual presentation modalities, our key question is which of these properties is transported through which channels, and how they play together. Our insight is that even simple auditory cues complement the visual channel quite effectively, allowing digital media to span a wider gamut of perceptual material properties.

3 Experiment 1

We conducted a psychophysical experiment in order to explore the effect of visual and auditory stimuli on the task of material property perception. Our goal was to obtain meaningful evidence supporting the influence of auditory cues in isolated form, or in addition to visual ones. Firstly, we will briefly describe the details of the experiment, which will be followed by the discussion of the results.

3.1 Methods

Selection of materials. We have collected a database of 32 flat material samples distributed along three distinct categories including 11 leathers, 10 papers, and 11 fabrics. In the election of materials we selected the specimens to be as diverse as possible in terms of their physical and aesthetic properties, in an attempt to cover the relative heterogeneity within each material class. As the sound produced by an object highly depends on its geometry, we only considered flat or nearly-flat samples, in order to avoid undesired variability.

Visual stimuli. From each of the selected specimens, we cut a sample of 12×12 cm^2, placed it on a bright background in natural illumination, and took a photograph using a digital camera (Canon PowerShot G9 in raw mode), located at approximately 25 cm from the sample under a light angle. The described illumination and viewing conditions were kept constant during the whole acquisition process. Pictures were taken at a resolution of 4000×3000 px.

[1] The technical aspects of creating and handling these representations have been researched extensively in the graphics community. Interested readers are kindly referred to the SIGGRAPH course by Weyrich et al. [2009]

Subsequently white-balance correction has been applied and the images were cropped, such that the specimen covers approximately the whole image and all images share the same aspect ratio. The resulting images are shown in Figure 1.

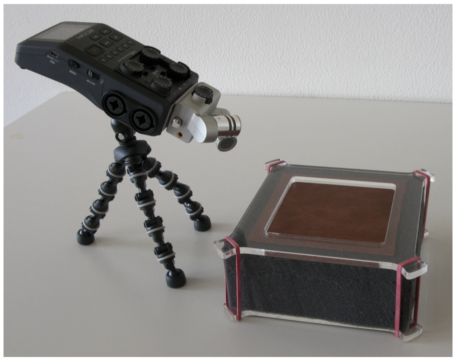


Figure 2: *View of the audio recording setup.*

Auditory stimuli. In order to record the contact sound produced by the specimens, we manufactured a special sample holder consisting of a $15 \times 15 \times 8$ cm^3 piece of polyurethane foam located between two layers of acrylic, the top one with a 10×10 cm^2 square cutout to expose the material sample. The sample is placed underneath the top acrylic layer, which gently presses it against the foam block. The entire stack is held together by four rubber bands under light tension, one in each corner. With this setup, the sounds produced by the contacts between the sample and any other adjacent surfaces can be reduced to a minimum. Sound recording was performed in an acoustically isolated room using a portable audio recorder (Zoom H6) with an X-Y pair of condenser microphones, about 10 cm away from the sample and facing towards it. Figure 2 depicts the whole setup.

With the purpose of covering a wide range of characteristic material sounds, we produced six different types of audio stimuli by touching the material with the fingertip. First, we performed four perpendicular movements, followed by four circular movements and lastly four strokes in the center of the material surface. Afterwards the same interactions were carried out using the fingernail instead of the fingertip creating one single track of sound. The length of each interaction was approximately 3 seconds, which altogether resulted in an audio track with a duration between 18 and 21 seconds. No post-processing was performed, with the exception of trimming.

After the recording step, we selected a final subset of 12 samples from the previous assortment of materials, 4 of each class, whose sound exhibited significant dissimilar characteristics.

Perceptual properties. We chose a set of 10 opposite pairs of adjectives, representing an intentionally diverse collection of perceptual properties (See Table 1). They sample the most characteristic properties from previous studies on material perception [Hope et al. 2013; Baumgartner et al. 2013; Fleming et al. 2013; Fujisaki et al. 2014; Fujisaki et al. 2015]. This assortment of qualities was conceptually organized into three groups according to the means of perception: tactile, visual and subjective. While the first two groups include properties related to physical parameters, the last group is rather associated with an emotional meaning or the user's personal preferences.

Participants. 26 subjects, gathered through our university's Laboratory of Experimental Economics (BonnEconLab), voluntarily participated in this experiment (13 females, mean age 26.46 years, standard deviation 6.39 years; 13 males, mean age 30.01 years, standard deviation 8.85 years). All the participants were naïve to the purpose of the experiment and reported normal or corrected-to-normal visual and hearing acuity. They provided informed consent and received economic compensation for their participation.

Tactile	Visual	Subjective
rough–smooth	shiny–matte	expensive–cheap
hard–soft	simple–complex	old–new
warm–cold	colorful–colorless	natural–synthetic
		beautiful–ugly

Table 1: *Set of opposite property pairs utilized in Experiment 1, grouped by type.*

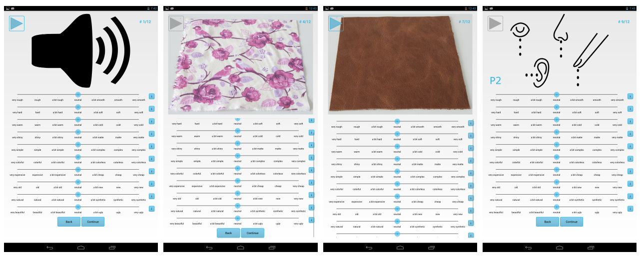
Procedure. The user study was carried out using a tablet device (Toshiba Excite Pro 10.1, 2560×1600 px resolution) running a custom Android application, shown in Figure 3, together with a set of headphones (Sony MDR-7506). With this kind of experiment, our setup is not only scalable to larger surveys but also representative of today's consumer hardware.

The experiment was conducted in a well-illuminated room with the windows and doors closed to avoid any source of external noise or disturbance. An experimenter was present during the whole course of the experiment. The number of subjects per session was limited to 6–8 to help the experimenter to control the correct realization of the experiment. An introductory presentation was provided in order to explain the procedure, clarify questions, and a video of the contact sound generation process was shown. Participants were instructed to infer or imagine properties that are not revealed in a particular presentation (e.g., purely visual properties during the auditory presentation). The application also provided a help system for definitions of each adjective in question. Upon starting the experiment, each participant had to individually set the volume to a comfortable level. It was then fixed and could not be changed during the completion of the test.

The procedure was structured into four consecutive phases, where the same materials have been presented to the subjects using different modalities. In every phase the order of materials was randomized. The tablet computer was used to generate the particular stimuli and also to conduct the questionnaire. For each combination of material and stimulus, the subjects rated the selected assortment of properties using a slider with values ranging from -3 to 3. Each of the values was consistently labeled with a term indicating the intensity of the property in both axes (e.g., very rough, rough, a bit rough, neutral, a bit smooth, smooth, very smooth). These ratings were finally interpreted as a magnitude estimation process [Stevens 1957]. The experiment was composed of the following four presentations:

- Auditory: An audio playback of prerecorded contact sounds.
- Visual: An image of the material.
- Audiovisual: Combined both image and audio playback.
- Full-modal: The participants received a 6×6 cm^2 physical material sample and were motivated to interact with it.

Moreover, the application was instrumented to identify user errors (such as skipping a material or failure to play a sound) in order to improve the reliability of the gathered data.



(a) Auditory (b) Visual (c) Audiovisual (d) Full-modal

Figure 3: *Screenshots of the Android application. Each image corresponds to one of the four presentations composing the experiment.*

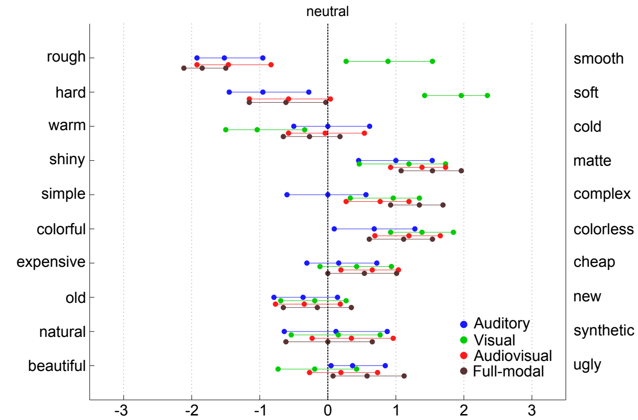


Figure 5: *Ratings for material P4. The central circles represent the participants' mean rating, the outer circles represent the bootstrapped 95% confidence interval for the mean. The figure is discussed in detail in Section 3.2.*

3.2 Results

The evaluation of Experiment 1 is structured into a study of the inter-participant correlation, an analysis of the individual material ratings as well as the participants' preferences. Finally, we also explored the dimensionality of the perceptual spaces spanned by the investigated modalities.

Inter-participant correlation. Given the broad nature of the selected properties, they are not likely to be communicated equally well along the four different types of presentations. We argue that if a property is clearly transported by a certain stimulus, the participants should generally agree on the judgment of this quality. Contrary, if the information is not well depicted by the presentation, the participants will have to use their imagination for rating and thus are expected to agree less. With this in mind, we employed an inter-participant correlation analysis in order to investigate the quality of property representations in each of the stimuli. Figure 4 plots the average correlations for each of the property pairs.

For the auditory presentation, the highest correlation has been obtained for the tactile attribute pairs "hard–soft", "rough–smooth", and "warm–cold". We deduce that, for the given set of attributes, sound is most suitable to transport tactile information. As expected, the agreement on visual properties is rather low here. The visual presentation performs exceptionally well on the adjective pairs "colorful–colorless" and "shiny–matte" which again was expected as these are purely visual properties. The agreement on the tactile properties is lower than in the auditory presentation. Examining the data for the audiovisual test, a tendency combining the previous presentations can be observed. In contrast to the preceding tests, the three most correlated adjective pairs include two visual properties and one tactile, namely "hard–soft", "shiny–matte", and "colorful–colorless". Lastly, the correlations of the full-modal presentation follow the same tendency as the audiovisual presentation, but show an overall higher correlation.

To summarize, we have found evidence that, especially for tactile properties, the participants' overall agreement in property rating rises by adding contact sounds to an image-only presentation.

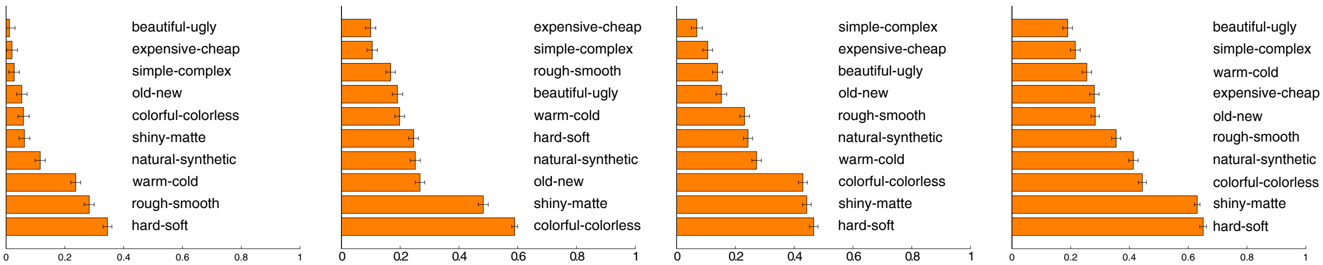
Material ratings. In this section, we intend to additionally explore whether it is possible to enrich the digital communication of material properties by adding sound to the visual representation. For this purpose, we have analyzed the average property ratings and the *confidence intervals* (CI) of the mean across all the participants, for each material independently. Given that the accuracy of this interval depends on the normality of the data and since the distribution of the material ratings is not normal, we bootstrapped the confidence intervals of the mean. Bootstrapping provides a way to construct CIs that does not rely on the normal assumption [Efron 1982; Efron and Tibshirani 1986]. Figure 5 depicts these values for one concrete material.

We make use of statistical hypothesis testing on this material to evaluate the significance of the following suppositions. Firstly, we focus on verifying whether our full-modal stimuli conveyed relevant impressions in any side of the polarity axis. This represents our alternative hypothesis H_a. Thus, our null hypothesis H_0 states that all stimuli are neutral on the polarity axis. The hypothesis H_0 would be falsified if the neutral position is not in the confidence interval for the presentation in question. Indeed, for 7 out of 10 pairs ("rough–smooth", "hard–soft", "shiny–matte", "simple–complex","colorful–colorless", "expensive–cheap" and "beautiful–ugly") we are able to reject H_0, therefore giving significant support to H_a.

Secondly, we evaluate whether the audiovisual presentation improves the isolated presentations when communicating the properties of this material (H_a). The formulation of H_0 declares that the audiovisual mean is not in the confidence interval for the full-modal presentation. Our goal again is to falsify H_0. The depiction shows that we succeed in rejecting H_0 for 8 of the 10 property pairs ("hard–soft", "warm–cold", "shiny–matte", "colorful–colorless", "expensive–cheap", "old–new", "natural–synthetic", and "beautiful–ugly"). For the remaining two pairs ("rough–smooth" and "simple–complex"), the mean is still closely located to the confidence interval boundary. Applying the same hypothesis to the visual-only test would involve rejecting H_0 only for 6 pairs of qualities.

We conclude that, for this particular material, the ratings for the full-modal presentation mainly exhibit a significant bias towards the extrema of the property pairs. Furthermore, we found significant evidence that, for most of the quality pairs, the audiovisual test is more consistent with the ratings of the full-modal test than the purely visual one.

Preference analysis. In order to gain a better and broadened understanding on the predilections of the participants for the different modalities in the task of property judgments, we performed a preference analysis independently on each of the examined properties. We consider a material presentation to be well-suited to represent a



(a) Auditory (b) Visual (c) Audiovisual (d) Full-modal

Figure 4: *Average inter-participant correlation per property, grouped by presentation and sorted in ascending order w.r.t. the correlation. Note that the addition of sound increases the the overall agreement for tactile properties.*

certain property if participants rate close to the full-modal presentation. Contrary, when the ratings are far apart, the presentation is judged to be less realistic and thus less suitable. To comprehend which is the impact of sound in these preferences, we compare the visual to the audiovisual stimulus, using a weighted voting scheme.

Let r_{fm}, r_{v} and r_{av} be the ratings for a certain combination of material and property with full-modal, visual and audiovisual stimulus respectively. The corresponding weights are defined as

$$w_{\text{v}} = \begin{cases} |r_{\text{fm}} - r_{\text{av}}| - |r_{\text{fm}} - r_{\text{v}}| & \text{if } |r_{\text{fm}} - r_{\text{v}}| < |r_{\text{fm}} - r_{\text{av}}| \\ 0 & \text{else} \end{cases}$$

and

$$w_{\text{av}} = \begin{cases} |r_{\text{fm}} - r_{\text{v}}| - |r_{\text{fm}} - r_{\text{av}}| & \text{if } |r_{\text{fm}} - r_{\text{av}}| < |r_{\text{fm}} - r_{\text{v}}| \\ 0 & \text{else} \end{cases} .$$

This means, that the weights grow with the difference of the ratings. To compute the final scores, we sum up the weights over all materials and participants, which is followed by a normalization,

$$S_{\text{v}} = \frac{\sum w_{\text{v}}}{\sum(w_{\text{v}} + w_{\text{av}})}, \quad S_{\text{av}} = \frac{\sum w_{\text{av}}}{\sum(w_{\text{v}} + w_{\text{av}})}.$$

The normalized scores, separated by properties, are shown in Figure 6. A clear preference for the audiovisual presentation for a certain property would entail that the addition of sound information augments the way we perceive materials for the given conditions. Indeed, analyzing the results reveals a meaningful enhancement for some of the properties, especially for "rough–smooth" and "hard–soft", both categorized as tactile. Substantial preferences for the audiovisual presentation can also be observed in other adjective pairs such as "simple–complex", "old–new" or "beautiful-ugly". No significant bias towards the visual presentation could be observed for any of the property pairs. This suggests that the addition of sound doesn't downgrade the representation of material properties.

Dimensionality of the perceptual property space. In the previous section we noticed that the addition of sound to a visual material presentation is able to enrich the perception of certain properties. To further confirm this insight, we analyzed the dimensionality of the perceptual property space spanned by the qualities used in this experiment.

We averaged the ratings over all participants and performed a principle component analysis (PCA) for each type of presentation on

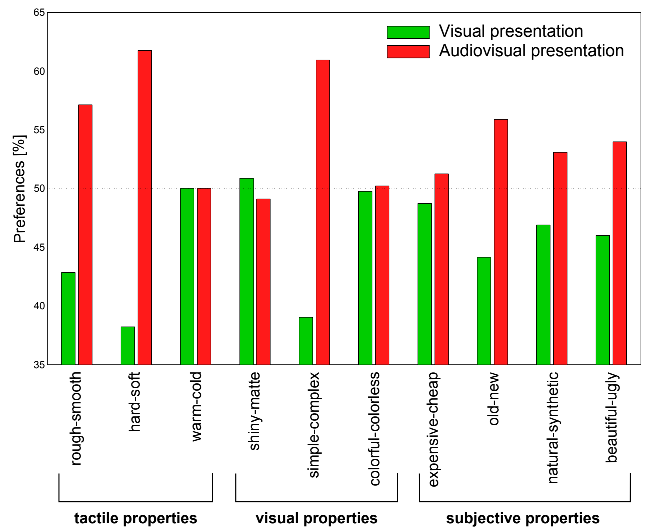


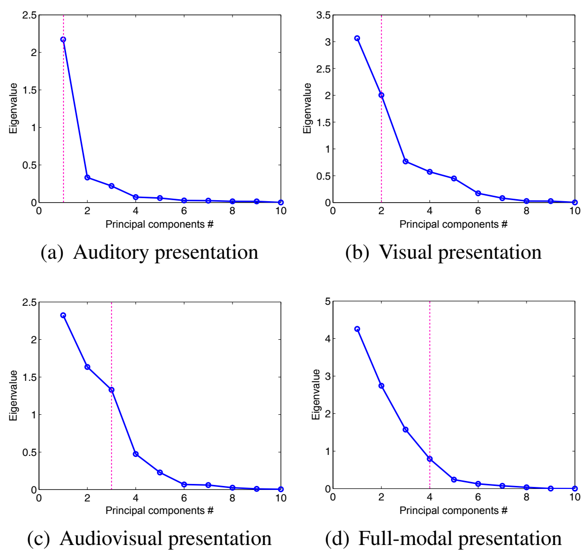
Figure 6: *Participants' preference for the type of presentation, according to our voting schema. A strong bias towards the audiovisual presentation can be observed especially for the tactile property pairs "rough–smooth" and "hard–soft", as well as for "simple–complex", "old–new" and "beautiful-ugly". There is no significant preference for the visual presentation in any pair, so the addition of sound does not deteriorate the perception.*

the mean data. The factor loadings, as well as the explained variances of the first 3 principle components for each presentation are shown in Table 2. Furthermore, Figure 7 illustrates the corresponding scree plots, with the principle components on the x-axis and corresponding eigenvalues on the y-axis. Using the scree test we determined the dimensionality of the data by looking for the point in the plots, where the graph's strong slope ceases and the remaining eigenvalues start to approximately even out on a low level. With this criterion, we found one significant dimension for the auditory presentation, two for the visual, three for the audiovisual, and four for the full-modal presentation, with the cumulative explained variance being 73.99, 70.74, 85.88, and 95.13 percent respectively. We deduce that combining auditory and visual cues increases the representable dimensionality of the perceptual property space over the visual presentation alone.

A detailed examination of the coefficients reveals that, for the auditory test, the most significant PC is dominated by the tactile qualities "hard–soft", "rough–smooth", and "warm–cold", which

	Auditory			Visual			Audiovisual			Full-modal		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
rough–smooth	**0.600**	−0.045	**−0.405**	0.022	−0.242	−0.048	0.319	**−0.351**	−0.160	−0.265	**0.466**	−0.216
hard–soft	**0.618**	0.073	0.316	0.292	0.006	0.240	**0.619**	**−0.477**	0.112	0.116	**0.680**	**0.519**
warm–cold	**−0.372**	**0.357**	−0.258	−0.330	0.151	−0.115	**−0.440**	−0.008	0.162	−0.237	−0.181	−0.020
shiny–matte	−0.076	−0.028	**0.452**	**0.653**	−0.022	0.056	**0.353**	**0.647**	−0.069	**0.611**	−0.219	−0.183
simple–complex	−0.067	0.017	**0.670**	−0.080	0.112	0.078	0.020	0.034	0.132	0.192	−0.080	0.285
colorful–colorless	−0.177	0.106	−0.039	0.153	**0.824**	**−0.355**	0.133	0.149	**0.818**	0.200	−0.120	**0.444**
expensive–cheap	−0.103	0.218	−0.035	−0.027	0.144	**0.617**	0.038	−0.033	0.205	−0.130	**−0.353**	0.274
old–new	0.171	0.171	0.101	**−0.379**	−0.056	**−0.369**	−0.333	−0.134	0.000	−0.344	−0.149	−0.130
natural–synthetic	0.162	**0.870**	0.047	**−0.451**	0.195	**0.480**	−0.252	**−0.427**	0.246	**−0.519**	−0.180	**0.370**
beautiful–ugly	−0.110	0.136	0.038	0.050	**0.407**	0.211	0.054	0.077	**0.377**	0.078	−0.200	**0.378**
Explained variance [%]	73.99	11.33	7.48	42.78	27.96	10.70	37.74	26.54	21.60	43.24	27.85	15.98

Table 2: *Factor loadings and explained variance of the first three principal components for each modality. Bold values represent the strongest factors (greater than* 0.350*) for each principal component.*



(a) Auditory presentation (b) Visual presentation

(c) Audiovisual presentation (d) Full-modal presentation

Figure 7: *Scree plots of the PCAs, showing the PCs vs. the corresponding eigenvalues. Application of the scree test, illustrated by the vertical line, reveals one significant dimension for the auditory presentation, two dimensions for the visual, three dimensions for the audiovisual, and four dimensions for the full-modal.*

is in accordance to the inter-participant correlation reported above. Moreover, tactile properties have no strong influence on the first two PCs of the visual presentation, whereas they are strongly present in the first PCs of the audiovisual presentation. This indicates that the information representable by the auditory and visual presentation is orthogonal, which explains the increase in the dimensionality. For the full-modal presentation we can observe that the first two PCs interchange w.r.t. the audiovisual presentation. Here, the first PC is dominated by vision and the second PC by tactile properties, contrary to the audiovisual stimulus.

4 Experiment 2

While the results of our first experiment indicate that augmenting the visual presentation of materials with additional sound characteristics modifies the way we perceive them, Experiment 2 explores whether it is possible to consistently manipulate the perception of a material in the audiovisual presentation by replacing its auditory stimulus. For this purpose, diverse combinations of the sounds and images acquired for the individual material samples were shown to the participants. Similar to Experiment 1, we will first describe the details of the experiment and subsequently discuss the corresponding results.

4.1 Methods

Selection of materials and properties. Based on the results of Experiment 1, we identified specimens which, for the visual and auditory modalities, elicited stronger visual and acoustic ratings for specific qualities. Additionally, we also selected those specimens whose ratings showed a certain degree of contrast between the same modalities. Our selection was then reduced to a subset of 4 materials (L2, P1, P4 and F3) plus one additional sound stimuli (L4).

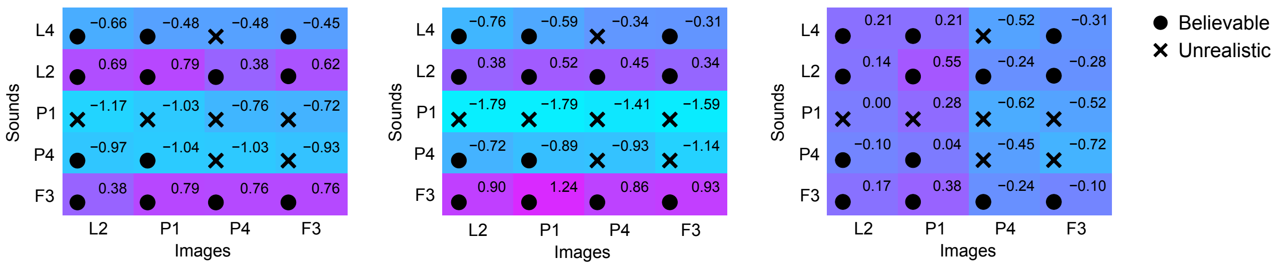
We also narrowed the selection of material qualities to the tactile ones (“rough–smooth”, “hard–soft”, “warm–cold”), complemented with the two subjective properties “old–new” and “beautiful–ugly”, which showed better audiovisual performance in the preference analysis in Section 3.2. In addition to this choice, we incorporated the pair “unrealistic–believable” in order to determine to what extent the realism of the experience was compromised.

Participants and procedure. 29 subjects (12 females, mean age 23.58, standard deviation 2.64; 17 males (mean age 27.06, standard deviation 4.77) participated in Experiment 2. The selection of the aforesaid participants was based on similar principles as in our previous experiment. Similarly, tablet devices and headphones were used for the presentation of visual and auditory information to 6-8 subjects simultaneously in a quiet, well-illuminated room. Again, instructions were given at the beginning of the procedure. In contrast to Experiment 1, in the audiovisual presentation all possible combinations between sound and image were shown to the participants for a total of 5 sounds $\times 4$ images $= 20$ stimuli.

4.2 Results

The results of Experiment 1 point out that sound contributes to the perception of material properties, especially for the tactile properties. In contrast, we now focus on obtaining insights on whether it is possible to change material property perception in a consistent and plausible way by manipulating the contact sound. In this scope, we additionally analyze whether the auditory ratings correlate to audiovisual ratings where the respective sound is combined with images of other materials.

Firstly, we investigated the image-sound interaction by exploring the mean ratings for the audiovisual presentation for each property.



(a) Rough–smooth (b) Hard–soft (c) Old–new

Figure 8: *Mean ratings for the manipulated audiovisual presentations. Rows indicate the utilized auditory stimulus and columns the visual stimulus. The mean ratings are color-coded and denoted in each cell. The realism of the combination is indicated by the symbol in the lower-left corner. A circle represents the combination was rated believable, while a cross represents it was rated unrealistic.*

Figure 8 depicts such values for the pairs "rough–smooth", "hard–soft" and "old–new". The rows of the respective matrices denote the sounds used in the evaluation while the columns denote the images. The level of acceptance (represented by the unrealistic–believable dimension) has been additionally specified with symbols. Depending on the sign of the mean ratings, we use a circle for specifying that a particular combination was rated to be believable, and a cross if it was rated to be unrealistic.

At first glance, the ratings obtained for "rough–smooth" and "hard–soft" reveal a homogeneous characteristic in the rows of the matrix. This suggests that the audiovisual perception is dominated by the characteristics of the auditory information rather than by vision for these two property pairs, i.e. the varying visual information contained in the different images does not exhibit a substantial influence. In contrast, in Figure 8(c) the row vectors of the matrix show a strong similarity among each other, i.e. the columns show a homogeneous behavior. Additional examination of the acceptance level denotes that, for most of the cases, this bias was achieved without endangering the plausibility of the experience. The sounds that were deemed largely unrealistic were principally the ones produced by paper, even for the actual sound-image pairs. We attribute this to the obvious imperfections in the audio recording and reproduction process (consumer devices). Nevertheless, these sounds can be used to affect the ratings of the other properties consistently.

To validate our observations, we considered the mean correlations between the columns and the rows respectively. The corresponding values are given in Table 3. For the pairs "rough–smooth" and "hard–soft" the mean correlations obtained for fixed auditory stimulus are significantly higher than the ones obtained with fixed visual information. For the pair "old–new", the correlation values exhibit the opposite tendency. These findings are in line with our aforementioned observations. In order to evaluate whether the audiovisual perception can be manipulated in a both predictable and consistent way, we also compared the auditory-only mean ratings to the corresponding audiovisual mean ratings. Indeed, we could find a high mean correlation here as well, being 0.97 for "rough–smooth" and 0.98 for "hard–soft".

	auditory	visual
rough–smooth	0.9736	0.1133
hard–soft	0.9807	−0.0684
warm–cold	0.8671	−0.1532
old-new	0.6490	0.9209
beautiful–ugly	0.9403	0.3547

Table 3: *Mean correlations between the ratings with fixed auditory and visual stimulus respectively. High correlations for the fixed audio are found especially for the pairs "rough–smooth" and "hard–soft", indicating a dominant influence of the auditory stimulus.*

5 Discussion and future work

The findings of our investigation are in line with previous work as they confirm that sound is indeed an important factor for the perception of material properties. We found that even simple contact sounds as the ones offered in our experiments can support the judgment of properties that are of a tactile nature, and, hence, offer an orthogonal complement to the visual channel. Even more, we can use sounds as a tool to achieve deliberate biases and manipulate the perception of those properties almost independently of the visually transmitted ones.

The sound presentation was limited to playing back pre-recorded sounds of a default sequence of touch activities. Observing that sound is strongly linked to haptic experience, it would be consequent to develop a synthesis scheme that would allow users to "scratch" a surface by touch, and listen to the resulting sounds in real time. We expect a significant increase in realism from a more immediate mode of interaction. A further avenue of future research could be the analysis of the connections between the space spanned by the perceptual qualities and the frequency spectrum of the audio signals.

By and large, the subspace made accessible by sound appears to be one-dimensional, and it remains unclear to which extent this is due to the scale of our experiments. In order to keep the overall size of the study manageable, we had curated 3 classes of materials and 10 pairs of opposite adjectives where there could have been many more of each. As a result, even the variance of the full-modal experience is represented to 95% by only 4 principal components. Fleming et al.'s [2013] 42-dimensional ratings space, on the other hand, exhibits a much more gently decaying eigenvalue spectrum, requiring 7 principal components to explain 50% of the total variance. We imagine that a scaled-up version of our experiments would reveal additional structure within the space of perceptual parameters and shed more light on how they are linked to the various sensing modalities.

Finally, we acknowledge that the visual stimuli used in this study were static, whereas the auditory stimuli were dynamic and the full-modal presentations even fully interactive. This fact may have caused bias in favor of auditory and audio-visual presentations. For future iterations of the study, we project to include dynamic visual models including animated objects and/or light sources to level the playing field.

6 Conclusion

We have found evidence that the addition of sound benefits the perception of digital materials, particularly for tactile qualities. Additionally we identified a way of manipulating the judgments of such properties in a consistent way. We believe that most of these findings can immediately be put to practice in product design and visualization. At the same time, it is clear that many questions on multimodal perception of materials remain to be answered; our results provide strong directions for deeper research.

Acknowledgments

This work was developed in the X-Rite Graduate School on Digital Material Appearance at the University of Bonn. Michael Weinmann was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 323567 (Harvest4D); 2013-2016. We would further like to thank Marc Ellens, James Ferwerda, Roland Fleming, Reinhard Klein, Adrian Kohlbrenner and Matthias Scheller Lichtenauer for valuable discussion. The experiments were conducted in collaboration with the Laboratory of Experimental Economics (BonnEconLab).

References

ADELSON, E. H., AND PENTLAND, A. P. 1996. Perception as bayesian inference. Cambridge University Press, New York, NY, USA, ch. The Perception of Shading and Reflectance, 409–423.

ADELSON, E. H. 2001. On seeing stuff: the perception of materials by humans and machines. *Proc. SPIE 4299*, 1–12.

AVANZINI, F., AND ROCCHESSO, D. 2001. Controlling material properties in physical models of sounding objects. In *Proc. Int. Computer Music Conf., La Habana*, 91–94.

BAUMGARTNER, E., WIEBEL, C. B., AND GEGENFURTNER, K. R. 2013. Visual and haptic representations of material properties. *Multisensory Research 26*, 5, 429–455.

BONNEEL, N., SUIED, C., VIAUD-DELMON, I., AND DRETTAKIS, G. 2010. Bimodal perception of audio-visual material properties for virtual environments. *ACM Trans. Appl. Percept. 7*, 1 (Jan.), 1:1–1:16.

BOUMAN, K. L., XIAO, B., BATTAGLIA, P., AND FREEMAN, W. T. 2013. Estimating the material properties of fabric from video. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, IEEE, 1984–1991.

CARELLO, C., ANDERSON, K. L., AND KUNKLER-PECK, A. J. 1998. Perception of object length by sound. *Psychological Science 9*, 3, 211–214.

CUNNINGHAM, D. W., WALLRAVEN, C., FLEMING, R. W., AND STRASSER, W. 2007. Perceptual reparameterization of material properties. In *Proceedings of the Third Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging (Computational Aesthetics '07)*, 89–96.

EFRON, B., AND TIBSHIRANI, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 54–75.

EFRON, B. 1982. *The jackknife, the bootstrap and other resampling plans*, vol. 38. SIAM.

ETZI, R., SPENCE, C., AND GALLACE, A. 2014. Textures that we like to touch: An experimental study of aesthetic preferences for tactile stimuli. *Consciousness and Cognition 29*, 178–188.

FLEMING, R. W., WIEBEL, C., AND GEGENFURTNER, K. 2013. Perceptual qualities and material classes. *Journal of Vision 13*, 8.

FLEMING, R. W. 2014. Visual perception of materials and their properties. *Vision Research 94*, 62–75.

FUJISAKI, W., GODA, N., MOTOYOSHI, I., KOMATSU, H., AND NISHIDA, S. 2014. Audiovisual integration in the human perception of materials. *Journal of Vision 14*, 4.

FUJISAKI, W., TOKITA, M., AND KARIYA, K. 2015. Perception of the material properties of wood based on vision, audition, and touch. *Vision Research 109, Part B*, 185–200. Perception of Material Properties (Part I).

GIORDANO, B. L., AND MCADAMS, S. 2006. Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *The Journal of the Acoustical Society of America 119*, 2, 1171–1181.

GUEST, S., CATMUR, C., LLOYD, D., AND CHARLES, S. 2002. Audiotactile interactions in roughness perception. *Experimental Brain Research 146*, 2, 161–171.

HO, Y.-X., LANDY, M. S., AND MALONEY, L. T. 2006. How direction of illumination affects visually perceived surface roughness. *Journal of Vision 6*, 5, 8.

HOPE, A. D., JONES, M., AND ZUO, H. 2013. Sensory perception in materials selection for industrial/product design. *The International Journal of Designed Objects 6*, 3, 17–31.

KLATZKY, R. L., PAI, D. K., AND KROTKOV, E. P. 2000. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments 9*, 4, 399–410.

LEMAITRE, G., AND HELLER, L. M. 2012. Auditory perception of material is fragile while action is strikingly robust. *Acoustical Society of America Journal 131*, 1337.

MALONEY, L. T., AND BRAINARD, D. H. 2010. Color and material perception: Achievements and challenges. *Journal of Vision 10*, 9.

PELLACINI, F., FERWERDA, J. A., AND GREENBERG, D. P. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of SIGGRAPH 2000*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 55–64.

RAO, A. R., AND LOHSE, G. L. 1993. Towards a texture naming system: Identifying relevant dimensions of texture. In *Proceedings of the 4th Conference on Visualization '93*, IEEE Computer Society, Washington, DC, USA, VIS '93, 220–227.

STEVENS, S. S. 1957. On the psychophysical law. *Psychological Review 64*, 3, 153.

VANGORP, P., LAURIJSSEN, J., AND DUTRÉ, P. 2007. The influence of shape on the perception of material reflectance. In *ACM SIGGRAPH 2007 Papers*, ACM, New York, NY, USA, SIGGRAPH '07.

WEYRICH, T., LAWRENCE, J., LENSCH, H. P. A., RUSINKIEWICZ, S., AND ZICKLER, T. 2009. Principles of appearance acquisition and representation. *Foundations and Trends in Computer Graphics and Vision 4*, 2 (Feb.), 75–191.