Communication of Digital Material Appearance Based on Human Perception

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Rodrigo Martín Calderón, M.-Sc.
aus
Toro (Zamora, Spanien)

Bonn, Juli 2018

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Matthias Hullin
2. Gutachter: Prof. Dr. Reinhard Klein

Tag der Promotion: 12. April 2019
Erscheinungsjahr: 2019

Contents

Abstract

In daily life, we encounter digital materials and interact with them in numerous situations, for instance when we play computer games, watch a movie, see billboard in the metro station or buy new clothes online. While some of these virtual materials are given by computational models that describe the appearance of a particular surface based on its material and the illumination conditions, some others are presented as simple digital photographs of real materials, as is usually the case for material samples from online retailing stores. The utilization of computer-generated materials entails significant advantages over plain images as they allow realistic experiences in virtual scenarios, cooperative product design, advertising in prototype phase or exhibition of furniture and wearables in specific environments. However, even though exceptional material reproduction quality has been achieved in the domain of computer graphics, current technology is still far away from highly accurate photo-realistic virtual material reproductions for the wide range of existing categories and, for this reason, many material catalogs still use pictures or even physical material samples to illustrate their collections.

An important reason for this gap between digital and real material appearance is that the connections between physical material characteristics and the visual quality perceived by humans are far from well-understood. Our investigations intend to shed some light in this direction. Concretely, we explore the ability of state-of-the-art digital material models in communicating physical and subjective material qualities, observing that part of the tactile/haptic information (e.g. thickness, hardness) is missing due to the geometric abstractions intrinsic to the model. Consequently, in order to account for the information deteriorated during the digitization process, we investigate the interplay between different sensing modalities (vision and hearing) and discover that particular sound cues, in combination with visual information, facilitate the estimation of such tactile material qualities.

One of the shortcomings when studying material appearance is the lack of perceptually-derived metrics able to answer questions like *"are materials A and B more similar than C and D?"*, which arise in many computer graphics applications. In the absence of such metrics, our studies compare different appearance models in terms of how capable are they to depict/transmit a collection of meaningful perceptual qualities. To address this problem, we introduce a methodology to compute the perceived pairwise similarity between textures from material samples that makes use of patch-based texture synthesis algorithms and is inspired on the notion of Just-Noticeable Differences. Our technique is able to overcome some of the issues posed by previous texture similarity collection methods and produces meaningful distances between samples.

In summary, with the contents presented in this thesis we are able to delve deeply in how humans perceive digital and real materials through different senses, acquire a better understanding of texture similarity by developing a perceptually-based metric and provide a groundwork for further investigations in the perception of digital materials.

Zusammenfassung

Im alltägliche Leben begegnen wir digitalen Materialien in einer Vielzahl von Situationen wie beispielsweise bei Computerspielen, Filmen, Reklamewänden in z.B. U-Bahn Stationen oder beim Online-Kauf von Kleidungen. Während einige dieser Materialien durch digitale Modelle repräsentiert werden, welche das Aussehen einer bestimmten Oberfläche in Abhängigkeit des Materials der Fläche sowie den Beleuchtungsbedingungen beschreiben, basieren andere digitale Darstellungen auf der simplen Verwendung von Fotos der realen Materialien, was z.B. bei Online-Shopping häufig verwendet wird. Die Verwendung von computer-generierten Materialien ist im Vergleich zu einzelnen Fotos besonders vorteilhaft, da diese realistische Erfahrungen im Rahmen von virtuellen Szenarien, kooperativem Produkt-Design, Marketing während der prototypischen Entwicklungsphase oder der Ausstellung von Möbeln oder Accesoires in spezifischen Umgebungen erlauben. Während mittels aktueller Digitalisierungsmethoden bereits eine beeindruckende Reproduktionsqualität erzielt wird, wird eine hochpräzise photorealistische digitale Reproduktion von Materialien für die große Vielfalt von Materialtypen nicht erreicht. Daher verwenden viele Materialkataloge immer noch Fotos oder sogar physikalische Materialproben um ihre Kollektionen zu repräsentieren.

Ein wichtiger Grund für diese Lücke in der Genauigkeit des Aussehens von digitalen zu echten Materialien liegt darin, dass die Zusammenhänge zwischen physikalischen Materialeigenschaften und der vom Menschen wahrgenommenen visuellen Qualität noch weitgehend unbekannt sind. Die im Rahmen dieser Arbeit durchgeführten Untersuchungen adressieren diesen Aspekt. Zu diesem Zweck werden etablierte digitalie Materialmodellen bezüglich ihrer Eignung zur Kommunikation von physikalischen und sujektiven Materialeigenschaften untersucht, wobei Beobachtungen darauf hinweisen, dass ein Teil der fühlbaren/haptischen Informationen wie z.B. Materialstärke oder Härtegrad aufgrund der dem Modell anhaftenden geometrische Abstraktion verloren gehen. Folglich

wird im Rahmen der Arbeit das Zusammenspiel der verschiedenen Sinneswahrnehmungen (mit Fokus auf die visuellen und akustischen Modalitäten) untersucht um festzustellen, welche Informationen während des Digitalisierungsprozesses verloren gehen. Es zeigt sich, dass insbesondere akustische Informationen in Kombination mit der visuellen Wahrnehmung die Einschätzung fühlbarer Materialeigenschaften erleichtert.

Eines der Defizite bei der Analyse des Aussehens von Materialien ist der Mangel bezüglich sich an der Wahnehmung richtenden Metriken die eine Beantwortung von Fragen wie z.B. "*Sind die Materialien A und B sich ähnlicher als die Materialien C und D?*" erlauben, wie sie in vielen Anwendungen der Computergrafik auftreten. Daher widmen sich die im Rahmen dieser Arbeit durchgeführten Studien auch dem Vergleich von unterschiedlichen Materialrepräsentationen im Hinblick auf. Zu diesem Zweck wird eine Methodik zur Berechnung der wahrgenommenen paarweisen Ähnlichkeit von Material-Texturen eingeführt, welche auf der Verwendung von Textursyntheseverfahren beruht und sich an der Idee/dem Begriff der geradenoch-wahrnehmbaren Unterschiede orientiert. Der vorgeschlagene Ansatz erlaubt das Überwinden einiger Probleme zuvor veröffentlichter Methoden zur Bestimmung der Änhlichkeit von Texturen und führt zu sinnvollen/plausiblen Distanzen von Materialprobem.

Zusammenfassend führen die im Rahmen dieser Dissertation dargestellten Inhalte/Verfahren zu einem tieferen Verständnis bezüglich der menschlichen Wahnehmung von digitalen bzw. realen Materialien über unterschiedliche Sinne, einem besseren Verständnis bzgl. der Bewertung der Ähnlichkeit von Texturen durch die Entwicklung einer neuen perzeptuellen Metrik und liefern grundlegende Einsichten für zukünftige Untersuchungen im Bereich der Perzeption von digitalen Materialien.

List of Abbreviations

Abbreviation	Definition
ALOT	Amsterdam Library of Textures
AMT	Amazon Mechanical Turk
ANOVA	Analysis of Variance
AR	Augmented Reality
AxF	Appearance Exchange Format
BRDF	Bidirectional Reflectance Distribution Function
BSSRDF	Scattering-Surface BRDF
BSP Tree	Binary Space Partitioning Tree
BTF	Bidirectional Texture Function
CG	Computer Graphics
CI	Confidence Interval
CNN	Convolutional Neural Network
CUReT	Columbia-Utrecht Reflectance and Texture Database
H_0	Null Hypothesis
H_a	Alternative Hypothesis
HDR	High Dynamic Range
HMD	Head-mounted displays
IM	Image Melding
IoT	Internet of Things
JND	Just-Noticeable Differences
LDR	Low Dynamic Range
LED	Light Emitting Diode
MDS	Multidimensional Scaling
MOS	Mean Opinion Score
MR	Mixed Reality
MRF	Markov Random Fields
PC	Principal Component
PCA	Principal Component Analysis

Pd	Pure Data
PSM	Perceptual Similarity Matrix
PTS	Perceptual Texture Space
RMS	Root Mean Square
RMSE	Root Mean Squared Error
SoE	Sense of Embodiment
SSE	Sum of Squared Errors
SSIM	Structural Similarity Index
SVM	Support Vector Machine
SVBRDF	Spatially Varying BRDF
VDP	Visible Differences Predictor
VR	Virtual Reality

Acknowledgments

The realization of this thesis would not have been possible without the help from various people. Firstly, I would like to thank Prof. Dr. Matthias Hullin who not only gave me the opportunity to research under his supervision but also assisted me with his guidance and advice, which was of invaluable help during my PhD.

I am also very grateful to my co-authors Prof. Dr. Matthias Hullin, Prof. Dr. Reinhard Klein, Michael Weinmann, Julian Iseringhausen, Christian Steinhausen, Dennis den Brok and Min Xue for their assistance and support regarding the publications that compose this thesis and other publications achieved during my time as a PhD student at the University of Bonn.

Additionally, I would like to express gratitude to all my office mates and the rest of my colleagues from the Computer Science II department. They were always kindly open for discussion inside the office and leisure time outside of it. Also, I learned constantly from them and they helped me to develop a broader view of different research domains.

Finally yet most importantly, I would like to thank my parents, Jesús and Sagrario, my brother Jesús, my partner Anna, the rest of my family and my friends. All of them were always there for me when I needed it and supported me through these years.

Part I

Introduction

CHAPTER 1

Introduction

1.1 Motivation

Digital representations of real world materials are present in a growing range of every-day situations, including online shopping, video games, product design, digital movie contents, advertising and many more. The continuous exposition to digital contents and, in particular, digital materials highlights the importance of developing material models that are not only physically accurate but also plausible from the point of view of human perception. In addition, the boom of devices and applications associated to the Internet of Things (IoT) demands interactions with these commodities that go beyond the passive product visualization.

Although research in computer graphics has continuously struggled to achieve higher accuracy and photo-realism at the expense of computational efficiency, the connections between measurable physical material properties and the visual quality perceived by humans is very subtle and far from well-understood. Indeed, despite the great deal of realism achieved by virtual/digitized materials, there is still an appearance gap between the latter and their physical counterparts that, in practice, may distort the perception of the final content. In particular, the precise representation of fine effects of material surfaces under varying viewing and illumination conditions remains a challenging task. For this reason, regardless of the benefits of using virtual surrogates, many material catalogs and online fashion still illustrate their collections through pictures or even physical samples, in order to avoid possible miscommunications of the material properties.

An interesting fact is that the perception of material appearance in

real-world is highly multimodal, or multisensory, by nature. In other words, while identifying the actual materials objects are made of, our vision, hearing or touch collaborate to varying extents. In spite of this evidence, the presentation of materials in virtual environments is based almost uniquely in visual models, and the rest of the senses have been largely neglected. One of the reasons for this lack of multimodal interaction is that, even though humans rely to a great extent on the tactile input in order to extract information from objects and materials, high-level, haptic interactions are still not accessible by the current digital technology. Nevertheless, high-quality stereo audio is regularly available in nowadays consumer devices and it may be employed to overcome the limitations given by the absence of tactile data. Some of these limitations relate to the fact that a consumer's decision for or against a material does not only involve the perception of physical properties but also an affective experience, which is greatly reduced in the case of passive visual representations. Considering this, the interplay between the visual and the auditory channels would provide interesting insights about how these two different senses collaborate in perceiving certain physical and affective material qualities.

Undoubtedly, this affective or emotional experience is closely related with the bandwidth of interactions allowed by a given product. However, either the visual or audiovisual representations of digital materials are typically passive (static images and pre-recorded sounds), which limits the expressivity of the final model. In this regard, a feasible way to enhance the level of interaction with digital materials could be the utilization of sonification systems. This set of techniques have demonstrated to compensate the absence of tactile information, to some degree, and to have a significant influence in the perception of product quality and efficiency.

Another interesting problem that faces the digital communication of material appearance is the absence of reliable similarity metrics. Questions like *"are materials A and B more similar than materials C and D?"* arise in many applications and could greatly facilitate the development of effective user interfaces for retailing web stores in agreement with human perception, among other applications. Yet, in the absence of such metrics, this query can be answered only partially, in terms of measured dimensions (i.e. glossiness, transparency) of parametric material models or the perceived realism according to human observers. Defining a meaningful metric for materials in accordance with human perception is a highly complex task due to the high dimensionality of material appearance. For this reason, one way to address this question is to consider simple tex-

tures as the easiest way of representing a concrete material.

The perception of textures is an essential field of research with a large range of applications in numerous vision problems. For the challenging task of estimating texture similarity as perceived by humans, former investigations have proposed diverse experimental designs to collect perceptual data. Most of these procedures become impractical for large datasets and pose further problems regarding the complexity and parallelizability of the assignment and the quality of the gathered user data. One of the worries regarding the quality of such perceptual measurements is that they have noise in them. However, this is hard to quantify as there is no ground-truth for questions like *"How similar are materials A and B?"*. Alternative approaches, like those based on the notion of just-noticeable differences (JND), allow a more direct measurement of the amount of noise which, in practice, could be employed as a mechanism to estimate the perceptual similarities/distances between material textures.

This investigation intends to shed further light on the previous inquiries and, in a more general way, to propose novel methods to communicate and measure the appearance of digital materials in accordance with human perception.

1.2 Main contributions and applications

With the work at hand, a series of experiments, methodology and data analyses is presented in order to gather deeper knowledge on some of the main unanswered questions regarding the perception of digital material appearance. In the following, we summarize the fundamental challenges and contributions on this subject introduced by the present document.

Initially, we study the multisensory nature of material perception by augmenting purely visual presentation of materials with audio characteristics. Such auditory information consists of recordings from rubbing and tapping material sounds. Thereby, we arrange a set of user studies to investigate how the visual and auditory channels contribute to the perception of material qualities. We conclude that a multisensory approach is not only able to enhance the digital communication of material properties without compromising the overall material perception, but also to achieve a deliberate bias for certain tactile qualities (e.g. roughness or hardness).

Following the same trend, we employ granular synthesis techniques to build a sonification system for tactile devices that delivers real-time contact material sound upon touch interactions. Then, we study the im-

pact of such system in the perception of physical and affective material qualities by means of a psychophysical study. The analysis of the experimental data reveals that the proposed audio cues alone (finger rubbing sounds) do not entail additional information to the perception of material attributes. We therefore suggest complementary directions of research in interactive material sonification that may lead to profitable results.

The election of photos from real samples as our stimuli in the previous investigations arises from the fact that the ability of digitized materials to communicate finer appearance effects is yet to be explored. Thus, we study the performance of state-of-the-art material appearance models in transmitting a set of subjective qualities through the visual channel, in comparison to photographs from real samples. By performing statistical analysis on the data from user experiments, we determine that digitized materials, in their current state, are still not fully capable of transmitting certain material qualities (mainly tactile ones) as good as simple photos. Subsequently, we investigate and confirm the hypothesis that important visual cues are destroyed when abstracting volumetric materials into flat digital representations, fact that is particularly noticeable at grazing angles.

To conclude, we focus on the problem of perceptually plausible metrics for material appearance. Concretely, we consider an existing dataset of textures from real fabrics samples taken under controlled conditions to examine their perceived pairwise texture similarity/dissimilarity. In this regard, we establish a novel methodology for the estimation of perceptual similarities between textures from materials based on the notion of just-noticeable differences and which relies on generating intermediate stimuli through texture synthesis techniques. Our proposal is well-suited to address fine-grain similarities, is convenient for crowdsourcing platforms and overcomes some of the issues existing in previous experimental approaches. This technique is then consistently exploited to construct a meaningful, low-dimensional and perceptually uniform space of textures from fabrics whose underlying interpretation of the main dimensions is in line with previous research.

The main purpose of our research is to find more effective ways to communicate materials in digital devices. This topic has a broad range of possible uses, but perhaps the application that could primarily benefit from the outcome of this thesis is the field of e-commerce or online shopping. In this context, our findings regarding the multimodal perception of materials and similarity metrics could be exploited in the development of potentially efficient user interfaces for retailing web-stores (see example web-stores in Figure 1.1), which integrate information from different

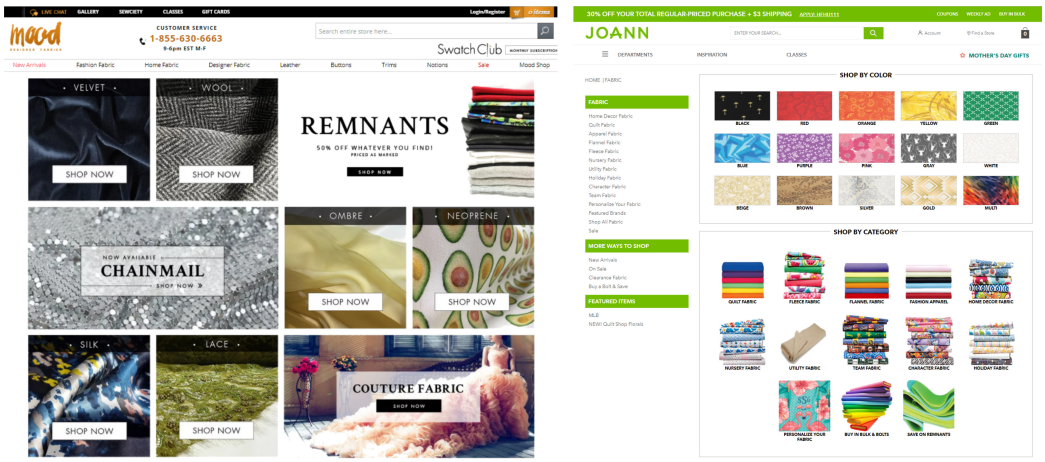


Figure 1.1: *Two examples of interfaces from online fabric stores. 'Mood Designer Fabrics' [Mood18] in the left, 'Joann' [Joann18] in the right. Both websites organize their products by color and application, independently of the material characteristics or their perceived similarity.*

senses and are more consistent with human perception. In addition, the field of product design could also draw important advantages from these conclusions e.g. when finding suitable material substitutes for unavailable samples or replacing computationally expensive models with more efficient ones among other uses. Finally, the outcome regarding the audiovisual perception of materials can be applied in the domain of virtual reality environments to better understand object properties in the absence of visual and tactile feedback in applications like clinical surgery, physical rehabilitation and sensory substitution for the impaired.

1.3 Publications

As a part of the present work, the technical chapters 3, 4, 5 and 6 are founded in publications from different computer graphics conferences and journals, which have successfully passed a peer-review process:

- R. Martín, J. Iseringhausen, M. Weinmann, and M. Hullin. Multimodal perception of material properties. In *ACM SIGGRAPH Symposium on Applied Perception*, SAP '15, pages 33–40, New York, NY, USA, Sept. 2015. ACM.

- R. Martín, M. Weinmann, and M. Hullin. Digital transmission of subjective material appearance. *Journal of WSCG*, 25(2):57–66, June

2017.

- R. Martín, M. Weinmann, and M. B. Hullin. A study of material sonification in touchscreen devices. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS '18, pages 305–310, New York, NY, USA, 2018. ACM.
- R. Martín, M. Xue, R. Klein, M. B. Hullin, and M. Weinmann. Using patch-based image synthesis to measure perceptual texture similarity. *Computers and Graphics*, 81:104 – 116, 2019.

Further closely related work with contributions from the author has been published in:

- H. Steinhausen, R. Martín, D. den Brok, M. Hullin, and R. Klein. Extrapolation of bidirectional texture functions using texture synthesis guided by photometric normals. In *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)*, volume 9398, Feb. 2015.

However, the contents of the latter publication are not part of this thesis.

1.4 Thesis outline

This document is organized in four thematic parts, each of them divided in several chapters which contain individual experimental studies, evaluations, results and contributions. Each chapter constitutes an indissoluble pipeline, in which later chapters utilize knowledge, algorithms or techniques that have been introduced in previous ones. However, in order to improve the coherence between the different chapters, an additional paragraph introducing each chapter and relating it to the previous investigation has been also included.

Thus, Part I provides a general introduction to the tasks approached in this thesis. In particular, Chapter 1 includes descriptions of the problems that motivate the present work, along with an outline of the contributions introduced as well as a list of the publications in which this essay is based. The subsequent Chapter 2 gives a necessary background on the fundamental topics on which this thesis is founded. This chapter includes a summary of the role of perception in computer graphics (Section 2.1), an overview of the main concepts regarding material appearance models and acquisition of material reflectance (Section 2.2) and a synopsis regarding the importance of textures in computer graphics along with the basic texture analysis and synthesis concepts and techniques (Section 2.3).

The technical contributions are introduced in Part II and Part III. Specifically, Part II addresses the topic of the perception of physical and affective material qualities through three chapters. In this regard, Chapter 3 presents the experimental results concerning the perception of material qualities through various senses or modalities (vision and hearing). Chapter 4 evaluates the perceptual effects of an interactive material sonification system developed for tactile devices. To conclude this part, Chapter 5 studies the ability of a concrete digital material representation (the SVBRDF model) to transmit physical and subjective physical material appearance characteristics in comparison to pictures and real materials.

Furthermore, part III explains our latest methodology for the collection of measurements of perceived similarity between material textures (Chapter 6). From the resulting similarity data, a perceptual space of textures is built and analyzed in detail.

Lastly, Part IV provides closure. In this section we summarize the ensemble of investigations presented in this thesis and suggest a landscape of profitable directions for future research.

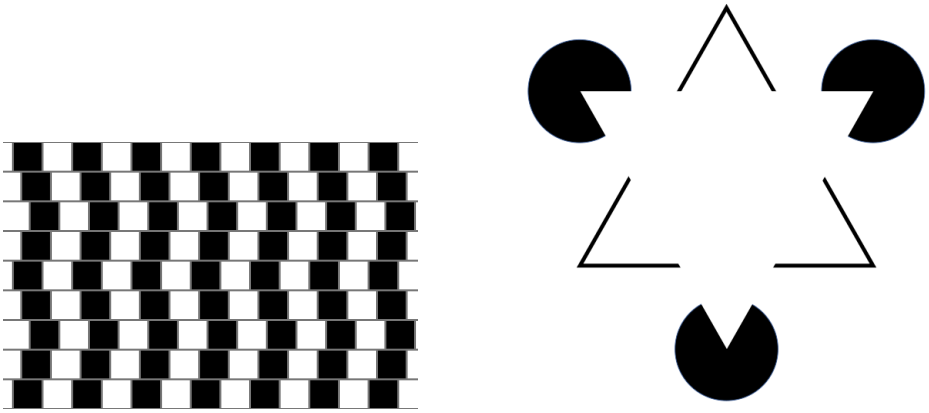
CHAPTER 2

Preliminary Knowledge

Before going into detail regarding the contributions of this thesis, an analysis of the relevant preliminary knowledge is presented. This analysis comprises the fundamentals of perception applied to computer graphics as well as a background in the principles of material appearance modeling and acquisition and a basic overview of visual textures, their analysis and synthesis.

2.1 Perception for computer graphics

Traditionally, research in computer graphics has made large efforts to achieve the technically best possible depiction of real world scenarios. Depending on the sub-field of interest, these endeavors may lead to representations based on the physics of light (*rendering*), real-time interactivity (*virtual reality*) or maximization of the information content in a scene (*visualization*). Lately, perceptual research has been increasingly integrated into computer graphics, thus providing new ways to solve existing problems by incorporating human knowledge obtained from user studies, commonly known as *psychophyisical experiments*. In this direction, the report from Bartz et al. [BCFW08] provides an exhaustive overview of the role and contributions of perceptual research in computer graphics. However, the fundamental theory basic for the understanding of the investigations presented in this manuscript will be given in the present section.



(a) *The café wall illusion, are the lines parallel?*

(b) *The Kanizsa triangle, note that the triangle edges are visible, despite not being physically present.*

Figure 2.1: *Two classical examples of visual illusions.*

2.1.1 Fundamentals of perception

In a general way, the goal of research in perception is to understand how living organisms process the physical information of the surrounding environment. From this observation, particular patterns and information are extracted to later be interpreted and utilized. In this regard, visual illusions have been often employed to investigate the shortcuts and assumptions that the visual system uses. The main purpose of such illusions is to highlight how perceptual and physical reality differ. Classical examples illustrating this contrast between perceptual and physical reality are the Café Wall Illusion [GP79] (Figure 2.1a), and the illusory contours [Kan79] (Figure 2.1b). In the first illusion, most people perceive the horizontal lines are not parallel although they actually are. In the second, a triangle can be seen in front of the three circles, when the truth is that only three partial circles are shown.

There is a growing number of visual illusions which are specifically designed to emphasize different heuristics and assumptions of our visual system. Disregarding the particular characteristics of the perceptual system, will most likely assure unwanted effects in the image or simulation under study. Taking advantage of them instead, will lead to more efficient algorithms for wide ranging topics and applications in computer graphics and other fields, as it will be discussed later.

2.1.2 From psychophysics to computer graphics

One question that may arise at this point is that, given the subjective nature of human perception, how can it be studied? The answer is that, for perception to be observed and analyzed, it must influence an organism's behaviour (humans in the case of the experiments regarding computer graphics). The methodology that allows the objective study of the underlying mechanisms of perception is known as *"psychophysics"*, and was developed by the physicist Gustaf Theodor Fechner in the mid 1800' [Fec60]. Such methods have as the main purpose *'the empirical study and mathematical formulation of the functional relationships between physical stimulation and sensory or perceptual responses'* [BCFW08]. To that end, psychophysics require total control over many factors including the actual stimuli to be presented as well as when, how and to whom they are presented. Thus, due to these constraints, perceptual research has commonly been conducted using rather abstract stimuli (lines, geometric figures or simple objects). Although such elementary stimuli present clear advantages regarding their reproduction, mathematical description and systematical variation, it is not clear whether the respective findings would generalize to more complex real-world scenarios [Gib79].

In modern psychophysics, the increasing ambition to study human perception in more real-world-like situations faces the traditional rigor of experimental methodologies. The problem with real-world stimuli is that they cannot be always exactly reproduced, mathematically described or systematically varied, which are fundamental requirements for perceptual research. The exceptional advances in computer simulations of reality, however, offer an elegant compromise to overcome preceding problem as they meet the generalization criteria as well as being representative from real-world cases.

2.1.3 Experimental design

There are five specific aspects that any perceptual experiment has to take into account: 1) what is shown (*stimulus*), 2) who gets to see it (*participants*), 3) how they get to see it (*stimulus presentation*) 4) what do they do with it (*task design*) and 5) analysis of the responses (*data analysis*). A comprehensive discussion on each of these aspects is beyond the scope of this manuscript and we kindly refer the reader to the book from Gescheider [Ges97] and the survey from Cunningham and Wallraven [CW11] for that matter. Still, the fundamental theory concerning the design of the experimental task will be reviewed here in further detail, as during our

investigations we invested an important effort in this particular aspect.

Despite of the vast amount of existing alternatives, experimental tasks can be usually classified into five rough categories: Questionnaires, free descriptions, scales, forced choice and physiology. The general idea behind free descriptions is to simply request participants to describe something using their own words. For instance, if investigating whether subjects are able to recognize what is being shown in an image (i.e. a cat, flowers, buildings), we would as them to verbally describe the contents of the picture. This task is highly subjective and arises the question of how to map the responses to a certain answer category. Questionnaires instead, ask people how they might react in a specific type of situation. For example, if we were to measure the visual quality of rendered images, a set of renderings would be presented to the participants along with the question *'which of the following images is more realistic?'*. With this design, different question formulations may lead to completely different answers. The use of rating or *Likert* scales is able to overcome some of the issues posed by two previous methods. According to this technique, when, for instance, investigating the aesthetic properties or an image we would pose the question like *'on a scale from 1 to 7, how beautiful is the following image?'*, together with a 7-point rating scale. There are several important concerns about rating scales regarding the anchoring (i.e. how beautiful is a 7?), and regarding the *level of measurement* [Ste46], which refers to the relationship between the values assigned to a variable (which is the difference between a 5-level beauty and a 6?). On this subject, research has shown that *'using appropriate adjectives for each scale point ensures that a proper interval is created'* [CW11]. Additionally, an interesting variation to Likert scales are *semantic differentials* introduced by Osgood et al. [Osg52], in which the ends of the scale are represented by opposite terms (e.g. realistic–unrealistic). In the forced choice alternative, a limited range of stimuli options is given as possible answers, and the participant is "forced" to select one of them. Finally, physiological experiments overcome the need for language by finding physiological factors that vary along with the stimuli (e.g. heart rate, blood pressure, brain waves, etc).

2.1.4 Research directions

The content of this dissertation partially lives in the ecosystem of both human perception and computer graphics research. According to this, we summarize next some relevant investigations regarding the synergies between these two fields, starting from their applications in virtual and augmented environments. The technology in Virtual Reality (VR) and

its mixed and augmented reality extensions (MR/AR) have experience a renaissance over the last years. The proliferation of consumer head-mounted displays (HMD) is parallel to the arrival of multiple specialized applications such as medical visualization, industrial design or video games, all of which have been used as a tool for the study of human perception. One concrete line of research that has an important influence in the cited applications is the sense of embodiment (SoE) in immersive environments described by Kilteni and Groten [KGS12], which relates to the feeling of owning and controlling a virtual body. According to the authors, the SoE comprises three sub-components: the sense of self-location, the sense of agency and the sense of body ownership, having each of them been subject of individual studies. On the other hand, perceptual insights have been used to evaluate the fidelity of virtual environments, including size and distance perception [CRST+15], sensation of walking [WBN+11] and visual realism among other aspects. Beyond perceptual research, the findings from these studies are also employed to learn which characteristics of a VR/AR setup have an impact on the user experience.

Visualization is a field of study that aims at the better comprehension of information and scientific data. Although there have been a considerable number of proposed techniques, only few of them have been evaluated from a perceptual point of view, in order to measure their potential benefits for the final user. Among the basic features that have been examined by means of psychophysical experiments are the choice of color, size and orientation of features and the density and regularity of texture elements. An overview on some of the predominant perceptual issues in visualization is discussed in the panel from House et al. [HIL+05] and the experimental design in the context of information visualization is further analyzed in the tutorial from Swan II [SI06].

In a similar vein, the perceived realism of digital scenes has also received a great deal of attention, focusing mainly in what is needed to produce perceptually realistic environments. In this context, the development of perceptually-driven fidelity metrics has achieved substantial importance. Most of the existing algorithms, such as the Visible Differences Predictor (VDP) [Dal93] which describes the human visual response, or the Structural Similarity Index (SSIM) [WBSS04] that measures image quality, result generally in a very conservative estimation of image/rendering fidelity [BCFW08]. The later notion of *visual equivalence* [RFWB07] proposes instead that two images are visually equivalent if *"they convey the same impressions of scene appearance, even if they are visibly different"*. Through their studies, the authors derive visual equivalence predictors for the object's shape, different illumination techniques and

employed materials.

When it comes to realism in computer generated animations, the human face is one of the best showcases, as it is capable of producing a large variety of expressions challenging to model in a digital environment. In a investigation from Wallraven et al. [WBCB08], psychophysical experiments are used to thoroughly evaluate the perceived quality of computer-generated facial animations w.r.t. real-world video sequences. In the same way, rendering algorithms also benefit from perceptual research, concretely its performance in terms of their rendering parameters and the resulting image fidelity have been also examined via psychophysical studies [KFB10].

Finally, yet importantly in the context of this thesis, human perception has a decisive role in the appearance of digital materials. The constant search in computer graphics for realistic ways to represent the interaction between light and materials has produced a large collection of digital material models which are, in general, approximations of physical laws. However, the importance of human perception in this pursiut of greater material realism has been often minimized or even disregarded. One exception is the investigation from Meseth et al. [MMK+06], which verified the capacity of one of these models (Bidirectional Texture Functions, BTFs) to depict photo-realistic materials. On this subject, the basic foundations regarding light reflectance representations and material appearance are reviewed in the following section.

2.2 Material appearance models and reflectance acquisition

Undeniably, the understanding of materials is essential to comprehend the world surrounding us. The perception of materials depends on the way light interacts with the material surface, which in essence defines their **appearance**. In general, *'a surface may reflect a different amount of light at each position, and for each direction of incident and exitant light'* [WLL+09], depending of the characteristics of the surface itself (Figure 2.2). Hence, to completely describe the opaque reflection at a certain point of a surface, we need a function that gives the amount of light reflected per each position (2D), incident light direction (2D) and exitant light direction (2D), resulting in a 6-dimensional function. This function contains information about the nature of the surface (i.e. whether it is shiny or matte, smooth or rough, etc) and thus, allows us to describe its appear-

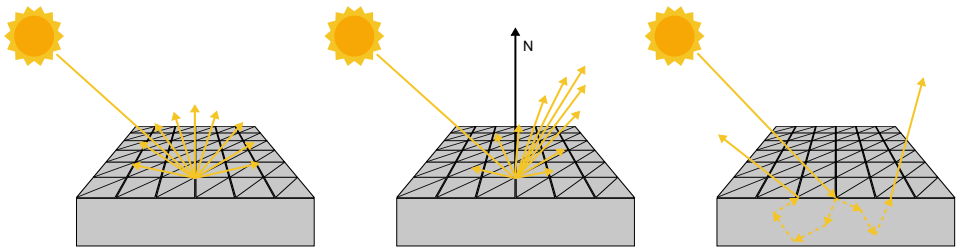


Figure 2.2: *Illustration with different types of surface reflectance, diffuse (left), specular (middle) and subsurface scattering (right).*

ance under any lighting conditions. This section details the fundamental notions of material appearance, starting from the concepts associated to the domain of radiometry, followed by the definition of the Bidirectional Reflectance Distribution Function (BRDF), which describes how light is reflected in opaque surfaces, its generalizations to more complex materials and the main notions about acquisition of appearance data. More detailed descriptions of the concepts addressed in this part are provided in the excellent reviews from Weyrich et al. [WLL+09] and Weinmann et al. [WLGK16].

2.2.1 Radiometry

Radiometry is the area of study associated with the measurement of electromagnetic radiation, including visible light, flowing in the space. According to this, there are a set of radiometric quantities and units that must be known by the reader beforehand. Although light is a form of electromagnetic energy and therefore is measured in Joules (J), we are mostly interested in the amount of energy flowing per time, known as *radiant flux* or *power*. This quantity is measured in Watts (Φ).

If we consider an ideal point light source, where the light is emitted evenly in all directions, describing its power would describe it completely. In the real world, however, this is normally not the case, thus we need to talk about the amount of power emitted in a particular direction i.e. per solid angle. The basic unit of the solid angle is the steradian, and is defined as the area of some region of the sphere divided by the square of its radius. A complete sphere has then 4π steradians. Thereby, we measure the *radiant intensity* (I), or *directional power* of a point light source in Watts per steradian (Equation 2.1). This leads us directly to the next radiometric quantity of our interest, the *irradiance* (E) or the amount of light falling

upon a surface (Equation 2.2), which is defined as the power per unit area. Finally, the most advanced radiometric concept is the *radiance* (L), which combines the two notions studied earlier. However, in practice, we consider the surface area as the projected surface instead (perpendicular to the viewing direction). Thus, the radiance describes the power emitted per unit projected area per solid angle (Equation 2.3) as illustrated in Figure 2.3.

$$I = \frac{d\Phi}{dw} \quad (2.1) \qquad E = \frac{d\Phi}{dA} \quad (2.2) \qquad L = \frac{d\Phi}{dA_{proj}dw} \quad (2.3)$$

The radiance is almost certainly the most important quantity in computer vision and graphics. It is intuitive to think that the irradiance on a camera sensor is proportional to the radiance of the surfaces being captured [WLL+09]. The equation given when radiance is integrated over all exitant angles is called *radiant exitance* (M), see Equation 2.4, where cosine term accounts for projected area. When this value is equal for all directions, this value is called *radiosity* (B).

$$M = \int_{\Omega} L(\theta, \varphi) \cos(\theta) dw \quad (2.4)$$

Lastly, the *plenoptic function*, which is a 5-dimensional positive function $L(x, y, z, \Omega, \Phi)$, allows to completely define the radiance at every light direction and every point in the 3D space [AB91]. Given that radiance is constant along rays in free space, we can neglect one of the dimensions (z), resulting into a 4-dimensional function known commonly as the *(4D) light field* [LH96].

2.2.2 Surface reflectance models

Once the basic radiometric concepts have been defined, we are ready to provide the definition of the *Bidirectional Reflectance Distribution Function (BRDF)*, first formulated by Nicodemus et al. [NRH+77], which describes the light reflection at a point in a opaque surface. In other words, it specifies the ratio between the incoming irradiance and the reflected radiance. The BRDF is generally expressed as a function of four variables, which are the polar angles of the incident and exitant light directions and is expressed in inverse steradians (sr^{-1}).

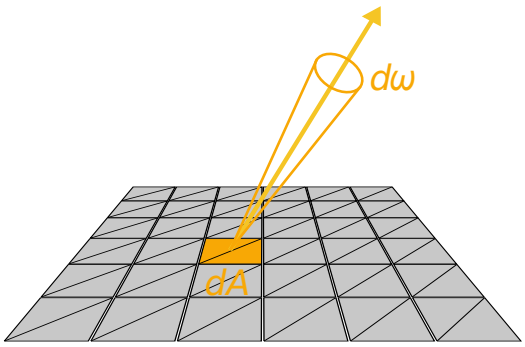


Figure 2.3: *Definition of radiance as the light emitted from a surface in a concrete direction ($d\omega$) per unit (projected) area (dA).*

$$f_r(\theta_i, \varphi_i, \theta_o, \varphi_o) = \frac{dL_o(w_o)}{dE_i(w_i)} \tag{2.5}$$

BRDF models have an important set of properties shared by most of them. The first one is the **energy conservation**: a surface cannot reflect more light that was incident on it. In other words, the light must be reflected or absorbed but cannot be created. A second characteristic is the **Helmholz reciprocity**, which states that the values of incoming and exiting light do not change if the light angles are swapped. When a BRDF satisfies the two aforementioned conditions, we consider it *physically plausible*. Another property held by some BRDF is called **isotropy**, where the BRDF value do not change if the light angles are rotated by the same amount. The opposite concept is **anisotropy**, where the BRDF value is affected by such rotations. BRDFs have more advanced characteristics representing more uncommon light effects i.e. asperity scattering or retro-reflection, that will not be covered in this section.

Over the years, several analytic BRDF models have been proposed, in which the reflectance of a material surface is given by a mathematical formula. Certainly, materials in real life present much more complex characteristics and the accuracy of this group of models is greatly limited. We introduce first the Lambertian model, in which the light reflected on a surface is a constant (Figure 2.2, left). In the Equation 2.6, parameter ρ represents the diffuse albedo (fraction of reflected vs. absorbed light). The next most elementary analytic model for material reflectance is the Phong model [Pho75], which was developed to describe the behaviour of glossy materials (Figure 2.2, center). This model is given by Equation 2.7, where V is the viewer's direction and R is the mirrored direction of the incoming light from the tangent plane (see Figure 2.4). The Blinn-Phong

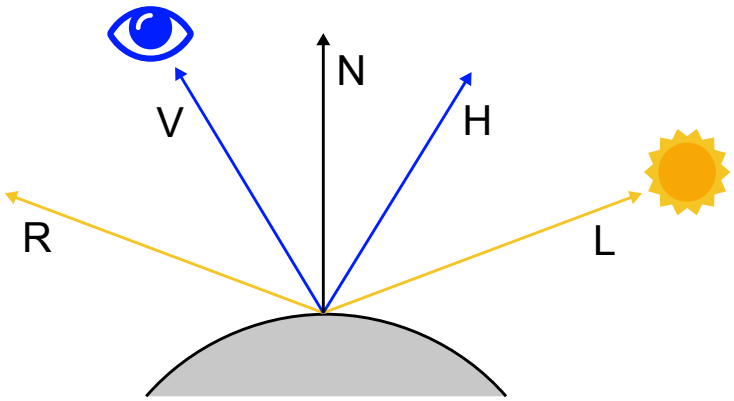


Figure 2.4: *Vectors utilized in the calculation of Phong and Blinn-Phong reflectance models.*

model [Bli77] is a physically implausible modification of the previous one that speeds up computations by replacing the dot product $R \cdot V$ by the product of the surface normal and the *halfway vector* between the viewer and light source vectors.

$$f_r = \rho/\pi \quad (2.6) \qquad f_r = k_s(R \cdot V)^N \quad (2.7) \qquad f_r = k_s(N \cdot H)^N \quad (2.8)$$

There is a considerably large collection of more advanced BRDF models explaining a greater mixture of optical phenomena i.e. the Lafortune, Torrance-Sparrow or Ashikihmin BRDFs, which are beyond the scope of this text. Instead, we are going to outline one concrete BRDF model of great relevance for the related work and some of the perceptual studies that will be presented in this manuscript, the Ward BRDF [War92] (see Equation 2.9). This model, which was developed in order to fit real-world measured surfaces, includes a specular peak shaped by a Gaussian function instead of the power-of-cosine employed by the previous equations. Furthermore, it has also the ability to model anisotropic reflections by applying different Gaussian widths (α_x and α_y) in the two perpendicular directions.

$$f_r = k_s \frac{e^{-tan^2\theta_h((cos^2\phi_h)/\alpha_x^2+(sin^2\phi_h)/\alpha_y^2)}}{4\pi\alpha_x\alpha_y\sqrt{cos\theta_i cos\theta_o}} \quad (2.9)$$

Inspired by Ward's model, other data-driven approaches which measure the reflectance of real material samples have become popular and

Figure 2.5: *Collection of BRDF materials captured by Matusik et al. [MPBM03]. As observed, all the samples represent flat, homogeneous and opaque surfaces.*

practical [MPBM03]. The set of devices and techniques for the measurement of such BRDF reflectance is convered in Section 2.2.3.

As already discussed by Nicodemus et al. [NRH+77] and also illustrated in the Figure 2.5, the range of appearances that BRDF models are able to describe is tailored to flat, homogeneous and opaque materials. However, real world objects exhibit complex behaviour that often differ among points in the same surface. In these cases, it becomes mandatory to increase the original BRDF equation with two additional spatial dimensions, leading us to the definition of *Spatially-Varying BRDF* model, also defined by Nicodemus et al. [NRH+77]:

$$f_r(x, y, \theta_i, \varphi_i, \theta_o, \varphi_o) \tag{2.10}$$

Since BRDF functions are only able to represent local phenomena, spatially-varying models cannot represent non-local light transport occurring due to e.g. self-occlusion, interreflections or subsurface scattering. Indeed, during the course of our investigations, we examined the capacity of SVBRDFs to communicate detailed characteristics of material appearance through the visual channel (Chapter 5) and observed that the abstraction from volumetric materials to flat surfaces inherent to the

model deteriorated the perception of important cues associated with the sense of touch (e.g. roughness, thickness and flexibility) among others. The evaluated SVBRDF material representations and the correspondent real-world samples employed in this study are depicted in Figure 5.3.

Still, the analysis of this function represents an important shifting of paradigm, from material appearance to object appearance, and therefore the geometry of the object in question needs to be taken into account. When this geometry is known, the spatial dimensions of the SVBRDF are paremeterized over that concrete geometry. Otherwise, the 6-dimensional function needs to be defined relative to some reference surface. In fact, when the reference geometry is planar, the concept of *Bidirectional Texture Function (BTF)* [DvG+99] arises.

BTF models are usually acquired in a data-driven fashion by gathering and combining series of photographs from a material sample from different lighting and viewing conditions and thus are able to represent the mentioned non-local effects of light reflection if sampled sufficiently densely, making them suitable to represent materials with significant mesostructure. However, the acquisition devices are limited to small sample sizes and complete measurements. Besides, it is a considerably expensive process due to the time needed for acquisition, post-processing (hours to days) and storage requirements (in the order of terabytes) [SMdB+15]. An example of the rendering quality achieved by this model in comparison to simple BRDF materials with an applied texture is shown in Figure 2.6.

Although being able to depict more intricate effects of light reflection, neither BTFs nor SVBRDFs are enough to describe all kind of material appearances. Specifically, translucency is a phenomenon in which light enters the object and emerges from a different surface point, after being reflected inside the material itself (Figure 2.2, right). The model that is able to represent such effect is known as the Bidirectional Scattering-Surface Reflectance Distribution Function (BSSRDF) [NRH+77]:

$$f_r(x_i, y_i, \theta_i, \varphi_i, x_o, y_o, \theta_o, \varphi_o) = \frac{dL(x_o, y_o, \theta_o, \varphi_o)}{d\phi(x_i, y_i, \theta_i, \varphi_i)} \tag{2.11}$$

This function adds two more parameters to the SVBRDF, which account for the location where the light leaves the surface. The BSSRDF is an 8-dimensional function defined as the input power at a single point, unlike the BRDF, which is defined over a differential area [WLL+09]. Hence, it is formulated as a fraction of incident flux instead of irradiance ($m^{-2} \cdot sr^{-1}$).

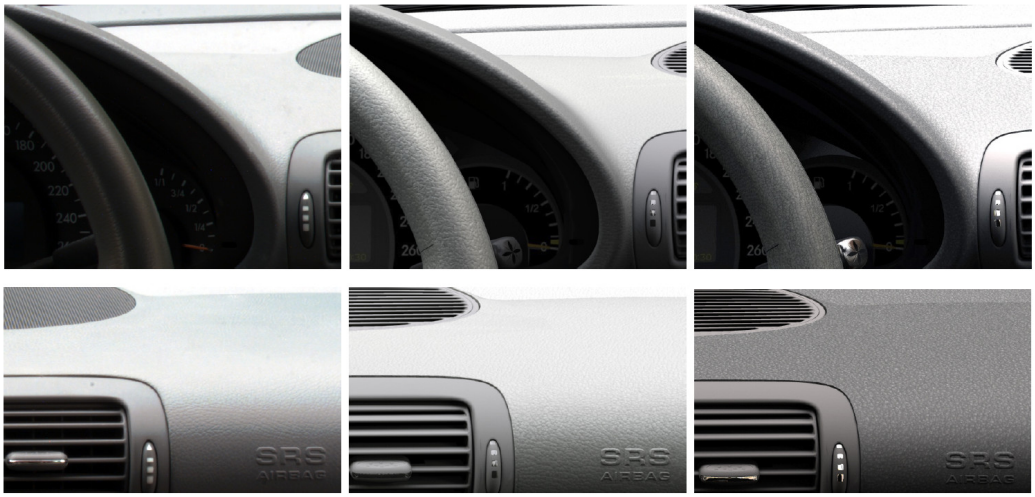


Figure 2.6: *Stimuli presented by Meseth et al. [MMK+06] in which the rendering quality of measured BTFs (middle column) is compared against the original photographs (left column) and textured BRDF materials (right column) in the context of car interior scenes.*

Even with the addition of the scattering behaviour, still not all the possible ranges of material appearances are covered. For that, one would need to consider first the previous functions as dependent from the light wavelength. Second, apart from the translucent effect, some surfaces are fluorescent, meaning that they emit light whose wavelength differ from the incoming rays. Moreover, some surfaces have properties that change over time. In summary, to completely characterize surface reflectance, two wavelengths, two time dimensions and two additional spatial dimensions should be added to the BSSRDF function, to shape a general 14-dimensional scattering function. To date, we do not have knowledge of any attempt to capture the full material appearance based on this function.

2.2.3 Reflectance acquisition

This sections provides a brief summary of the most relevant reflectance acquisition concepts and methods for the given material models. For more in-depth discussion about the topic, we direct the reader to the surveys from Weinmann et al. [WLGK16] and Weyrich et al. [WLL+09].

The most traditional and widely used device for the measurement of BRDF reflectance is the gonioreflectometer (see Figure 2.7). This apparatus typically uses servo motors to position a beam of light (w_i) and a

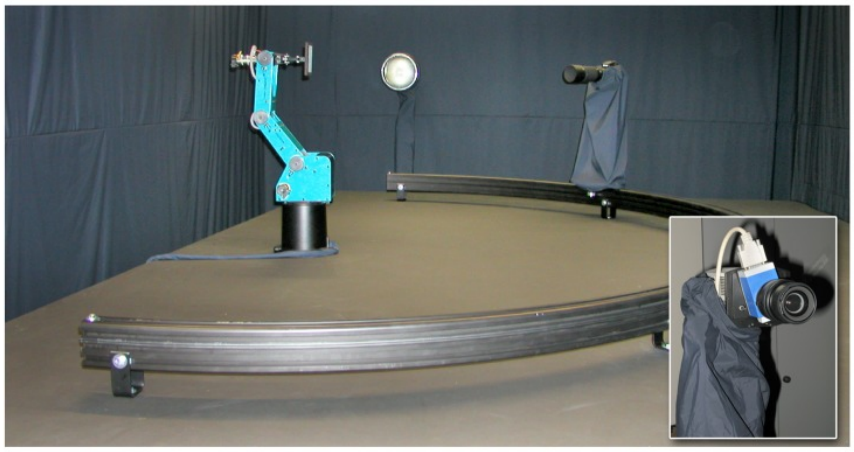


Figure 2.7: *An example gonioreflectometer from the University of Bonn. On the lower right side, the camera that is employed for hyper-spectral measurements. Image obtained from [SSW+14].*

sensor (w_o) and is designed such that the input and output directions are known relative to the local coordinate system. Furthermore, the sensor is usually linked to a spectroradiometer that allows recording spectral measurements for each source/sensor configuration. Although the measurements are reliable and repeatable, the whole process tends to be quite lengthy. However, acquisition time can be shortened by considering only isotropic BRDFs, where the model is invariant to rotations of the light angles [LFKW06]. Still, there are several physical limitations that complicate the measurement of accurate BRDFs. For example, measurements are troublesome when the incident and/or exitant light approach grazing angles due to surface foreshortening. Similarly, the measurements can become noisy with very dim or very specular surfaces.

The main difference when acquiring SVBRDFs instead of simple BRDFs is given by the spatial variation in the surface reflectance behavior that is modeled as a spatial distribution of independent BRDFs over the surface of the material sample [WLGK16]. For this reason, in order to measure higher dimensional appearance (SVBRDF or BSSRDF) more efficient and advanced techniques are required, such as arrays of light sources, digital projectors and digital cameras. On the other hand, these methods introduce several drawbacks i.e. the loss in spatial resolution, the complexity of the calibration procedure and the need to estimate the radiometric mapping between recorded intensities and scene radiance values. If the prior obstacles are properly addressed, the acquisition of these models al-

low to represent fine spatially-varying reflectance and scattering effects.

One important aspect that has to be taken into account while measuring higher dimensional reflectance properties is that the object geometry/shape must be known. One common approach to estimate the shape is the use of structured light in combination with stereo cameras [ZCS03]. This method introduces further alignment issues between the recovered shape and the images used for reflectometry. Besides, the computation of the required surface normals from the discrete shape introduces additional sources of noise. Instead, surface normals can be directly calculated from the images using photometric stereo [Woo79], whose classic approach assumes purely Lambertian material reflectance. The particular case of recovering three-dimensional shape from a single image is known as shape from shading [Hor89], and can be employed in order to model the geometry of highly specular materials. The development of methods to reconstruct geometry and surface structure from images is still an active field of research. A complete overview on the foundations of photometric stereo as well as an examination of the latest techniques in geometry reconstruction is given by Ackerman and Gosele [AG15].

There is a large amount of acquisition designs and devices, and enumerate all of them would be out of the scope of this document. In general, there are four aspects that need to be considered: acquisition time, precision, cost (computational, storage-wise, etc) and material diversity (or range of material categories that a device can measure). During certain phases of our investigation, we required the measurement of spatially-varying reflectance of a set of different materials. To that end, we employed the TAC7 Scanner [XR16], a commercial device whose detailed description is provided in Section 5.8.

2.3 Visual textures

In computer graphics we commonly use texture mapping as a standard technique to depict surface details without explicitly modeling the material's geometry or its fine reflectance properties (see Figure 2.6, right column). The image mapped onto the object/material is what we call *texture map* in the domain of computer graphics, or just *texture*, and can be used to represent numerous surface properties including color, reflection, transparency or displacement. In general, this term is used referred to an image containing arbitrary patterns, although depending on the field of knowledge it can adopt different meanings. For instance, in computer vision or image processing textures are referred as visual or tactile surfaces

composed of repeated elements while in visual perception is just defined as a property of "stuff" in the image in contrast to labeling "things" in the image [AB91].

On this subject, the visual perception of textures deals with the investigation of the underlying mechanisms of human vision. Yet, its role goes beyond that, as the appearance of textures is inherently related to the perception of particular materials, from which the texture is derived. Wile there has been tremendous efforts in the field of computer graphics to simulate real materials in virtual environments, the human perception of textural features has been also broadly investigated. However, the role of textures in the perception of virtual materials, which combines the knowledge of the two previous research directions, has received considerably less attention. A comprehensive research overview in the avenue of texture perception is provided by Landy and Graham [LG04]. This section however, is limited to a summary of the fundamental notions regarding the use of textures in computer graphics and vision. Such theory is discussed with greater detail in the review from Wei et al. [WLKT09].

2.3.1 Texture analysis and synthesis

Textures can come from many human sources, including hand-drawn pictures or photographs. While the first ones can be aesthetically pleasing, only professional texture artists can produce high-quality, photo-realistic textures, needed for many applications in computer graphics. On the contrary, scanned and photographed images usually have such required quality, but they are not directly usable for texture mapping as they tend to contain inadequate sizes, shadows, non-uniform lighting and/or undesired content. In this context, texture synthesis arises as an alternative method to computationally generate these textures. It is rather simple and general, as the user only needs to supply the input exemplar and a more-or-less large set of parameters to the algorithm. This family of methods are able to synthesize a bigger sample from an input one that, according to human observers, appears to be generated by the same latent processes. As stated by Wei et al. [WLKT09] ,texture synthesis has two main composing parts:

- **Analysis** or how to compute the latent generation process from the given example.
- **Synthesis** or how to build a generation mechanism to produce new textures from the given analysis model.

This two processes shape the fundamental differences between particular algorithms for texture synthesis.

2.3.2 Basic synthesis algorithms

A considerable amount of the methods for texture synthesis are based in *Markov Random Fields* (MRF). Also known as Markov Network, MRF is a technique of representing and visualizing a probability distribution based on the notion of *undirected graphs*. This model can compactly represent distributions, which are otherwise unable to be represented by directed models. Formally, a MRF is a probability distribution p over variables $x_1, ..., x_n$ given by an undirected graph G where the nodes correspond to the x_i variables and represent the pixels in the images. Then:

$$p(x_1, ..., x_n) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c) \tag{2.12}$$

Where $\phi(x_i, x_j)$ is a factor that assigns more weight to consistent votes among variables and Z is a normalizing constant. According to this, MRF characterize each pixel in the image by a small set of neighboring pixels and the goal of the synthesis process is to generate an output texture, such as for each output pixel its MRF neighborhood has a correspondence at the input texture. With this in mind, there are two essential techniques for texture synthesis, which at the same time rely on MRF to some extent. These techniques are the pixel-based and the patch-based synthesis.

Pixel-based synthesis The approach that bootstrapped the study of texture synthesis algorithms is the work proposed by Efros and Leung [EL99]. Its underlying idea is illustrated in Figure 2.8 and consists in the following steps:

1. Select a patch from a given input exemplar.
2. The output is initialized by copying that seed patch from the input.
3. Select an objective pixel outside the already synthesized pixels.
4. Find a set of candidate matches from the input with respect to the partial neighborhood of the objective pixel (red) and selects the center of one neighborhood randomly. The size of this neighborhood is determined by the user (3×3 in our case).
5. Repeat the previous step to grow the initial patch in an inside-out manner.

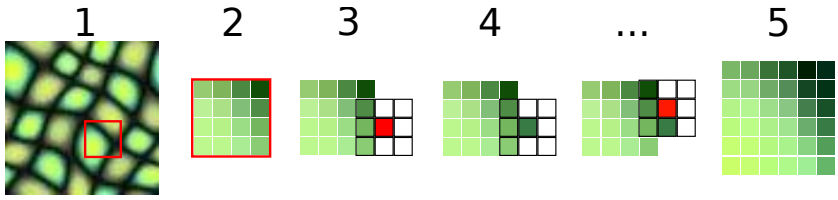


Figure 2.8: *Illustration of the basic steps for the pixel-based texture synthesis.*

This conceptually easy algorithm has as its only tunable parameter the neighborhood size, which should adequate to the size of the features in the original texture, otherwise the output might be too random. Although the method is influenced by MRFs, the process does not perform a rigorous MRF sampling and therefore is easily understandable. On the other hand, the performance of the method can be too slow and subject to non-uniform pattern distribution. These issues are addressed in the work from Wei and Levoy [WL00] by utilizing a fixed neighborhood, random initialization and synthesis in scanline order (instead of the inside-out fashion). The use of the fixed neighborhood in particular allows the application of diverse acceleration techniques (e.g. kd-trees).

Patch-based synthesis Both quality and performance of pixel-based approaches can be improved by considering the synthesis of patches instead of pixels. In order to ensure the output quality, patches are selected by considering its neighborhood, just as the pixel-based approach. The main difference between the two techniques lies in the way the input patch is copied onto the output, as patches, being larger than pixels, tend to overlap with already synthesized regions. Possible solutions are overwriting existing regions, blending, performing an optimization on the final pixels or warping the patches to ensure continuity.

In this context, Barnes et al. [BSFG09] introduced the *PatchMatch* algorithm to efficiently find nearest neighbor correspondences between image patches. The key insight is to find good candidate patches through random sampling and propagating such correspondences to surrounding areas thanks to the natural coherence of the image. Later, the authors generalized their algorithm to find k nearest neighbors instead of just one, by including rotation and scaling in the search space and by using arbitrary distance measures [BSGA10]. In our research, we will make heavy use of a synthesis algorithm that relies on PatchMatch correspondences in the context of computing perceptual similarities between textures in Chapter 6.

Additional synthesis methods There are further alternatives for the generation of synthesized textures besides the two mentioned techniques. One of them is the texture optimization, which, unlike previous approaches, considers all the pixels in the output image at the same time, and determines their values by optimizing a quadratic energy function (i.e. [KEBK05]). Such function is determined by the discrepancies between input/output neighborhoods, hence minimizing this function improves the output quality. The same principles of texture synthesis, including MRFs, can be applied also to dynamic textures i.e. textures whose appearance changes with time [WLKT09]. Although the classic example is having video data from dynamic phenomena, other approaches are also feasible, for instance materials whose appearance change over time [GTR+06]. Furthermore, it is also possible to consider static image samples and achieve dynamism by modifying the synthesis parameters over time (i.e. to generate plausible fluid simulations in 2D). Finally, procedural texturing techniques have the capacity to compute the synthesized appearance on-the-fly at any given point by training generative networks that capture the characteristics of the "Markovian patches" and generates output textures in real-time [LWM+16]. These methods, although effective, are not able to reproduce all range of texture appearances.

Part II

Perception of Digital Materials

CHAPTER 3

Multimodal Perception of Material Properties

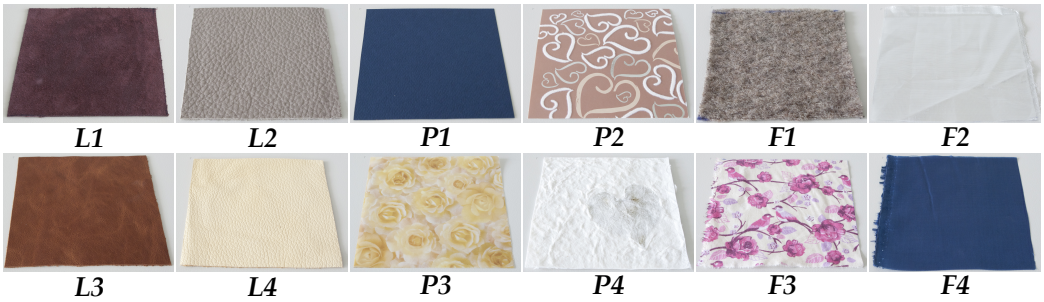


Figure 3.1: *(Teaser) Materials utilized in the experiment. Included are four leathers (L1 – L4), four papers (P1 – P4), and four fabrics (F1 – F4).*

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.

In this first approach to the perception of digital materials, we considered the question of whether prerecorded material audio cues (finger rubbing and tapping sounds) are able to facilitate the correct assessment of particular material properties or qualities. Given the insightful results obtained in an initial experiment, we went a step further and investigated if sound could bias the perception of those qualities predominantly related with the tactile feeling.

This chapter corresponds to the paper [MIWH15]: Rodrigo Martin, Julian Iseringhausen, Michael Weinmann and Matthias B. Hullin. "Multimodal Perception of Material Properties". In *ACM SIGGRAPH Symposium on Applied Perception*. September 2015.

3.1 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.

3.2 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 [Ade01], Maloney and Brainard [MB10] and Fleming [Fle14].

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 [AP96] examined the ability to judge the reflectance and the shading of the objects in three-dimensional scenes. Ho et al. [HLM06] 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 [PFG00] and together with transparency [CWFS07]. 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. [VLD07]. Bouman et al. [BXBF13] 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. [FWG13]. 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 [RL96] explored the dimensionality

[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. [WLL+09]

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. [CAKP98] 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. [KPK00] and additionally by Avanzini and Rocchesso [AR01]. Giordano and McAdams [GM06] investigated the identification of materials from impact sounds. They showed that, while listeners performed well with respect to gross material categories, their performance degraded for materials within the same gross category. Lemaitre and Heller [LH12] 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. [ESG14] 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. [GCLS02] 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. [BWG13], who looked for correspondence between visual and haptic material representations, and Hope et al. [HJZ13], 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. [BS+10] combined and analyzed levels of detail in audiovisual rendering; Fujisaki et al. [FGM+14] researched the principles that govern cross-modal integration of material information. Fujisaki et al. [FTK15] 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 respec-

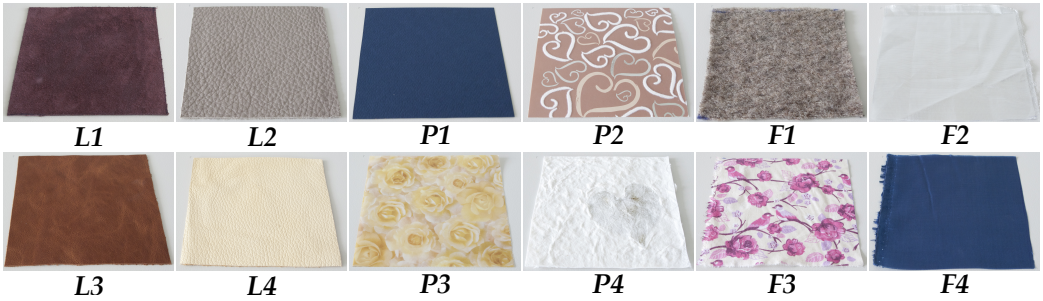


Figure 3.2: *Materials utilized in the experiment. Included are four leathers (L1 – L4), four papers (P1 – P4), and four fabrics (F1 – F4).*

tive 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.

3.3 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.

3.3.1 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.

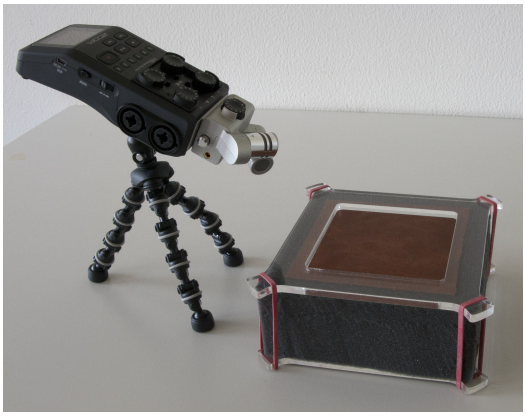


Figure 3.3: *View of the audio recording setup.*

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. 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 3.2.

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 3.3 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

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

Table 3.1: *Set of opposite property pairs utilized in Experiment 1, grouped by type.*

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 3.1). They sample the most characteristic properties from previous studies on material perception [HJZ13, BWG13, FWG13, FGM+14, FTK15]. 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.

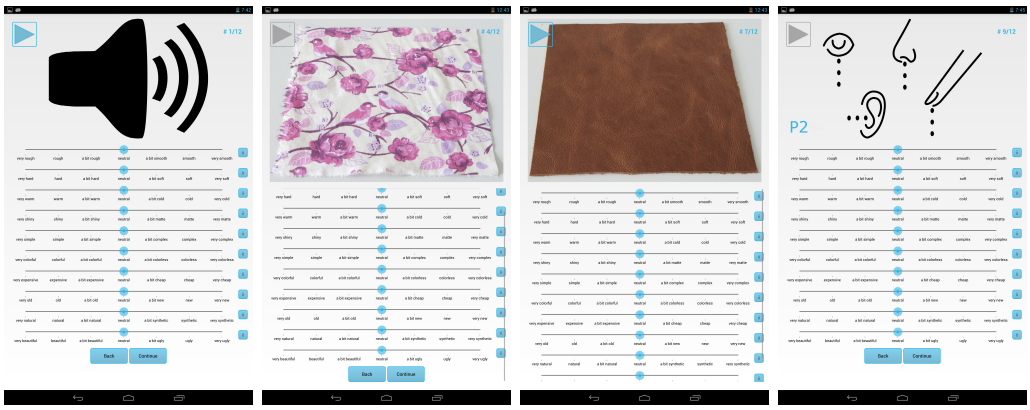
Procedure. The user study was carried out using a tablet device (Toshiba Excite Pro 10.1, 2560×1600 px resolution) running a custom Android application, shown in Figure 3.4, together with a set of headphones (Sony MDR-7506). With this kind of experiment, our setup is not only scalable to larger surveys 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 [Ste57]. 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*

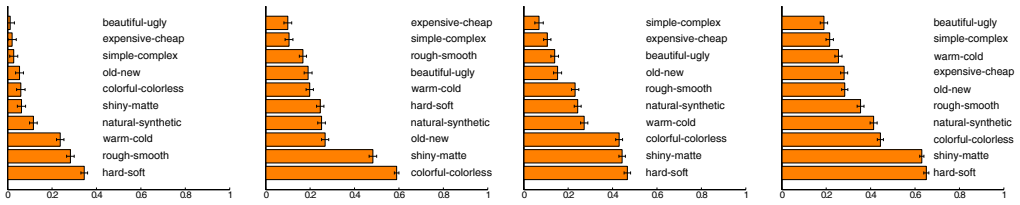
Figure 3.4: *Screenshots of the Android application. Each image corresponds to one of the four presentations composing the experiment.*

3.3.2 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 3.5 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 audi-



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

Figure 3.5: *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 overall agreement for tactile properties.*

tory 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 [Efr82, ET86]. Figure 3.6 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 ques-

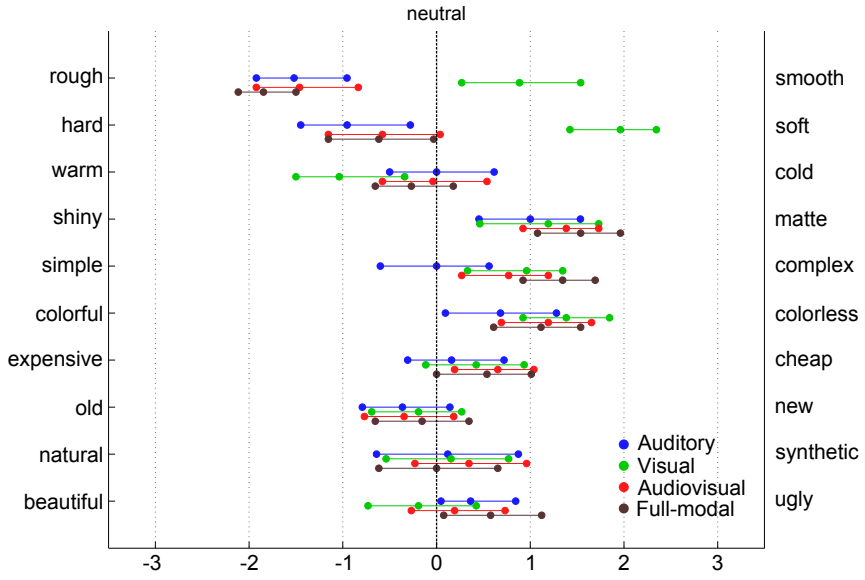


Figure 3.6: *Ratings for material P4. The central circles represent the participants' mean rating, the outer circles represent the bootstrapped 95% confidence interval for the mean. The figure is discussed in detail in Section 3.3.2.*

tion. 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 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_i be the ratings for a certain combination of material and property for a particular presentation $i, j \in \{visual, audiovisual\}$ and r_fm the ratings for the full-modal task [2]. The corresponding weights are defined as

$$w_\mathrm{i} = \begin{cases} |r_\mathrm{fm} - r_\mathrm{i}| - |r_\mathrm{fm} - r_\mathrm{j}| & \text{if } |r_\mathrm{fm} - r_\mathrm{i}| < |r_\mathrm{fm} - r_\mathrm{j}| \\ 0 & \text{else} \end{cases}. \tag{3.1}$$

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,

$$S_\mathrm{i,j} = \frac{\sum_i s_\mathrm{i,\,j}}{\sum_{i,j} s_\mathrm{i,\,j}}, \quad \text{where} \quad i, j \in \{visual, audiovisual\}. \tag{3.2}$$

The normalized scores, separated by properties, are shown in Figure 3.7. 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.

[2] Notation of the preference analysis changed for consistency with the notation in Section 5.4.1. For the original formulation we kindly refer the reader to the original publication [MIWH15].

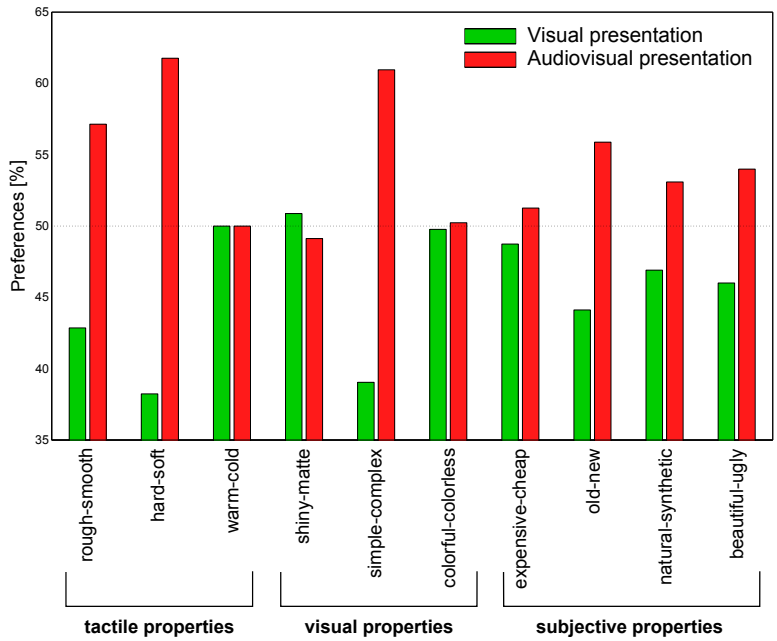
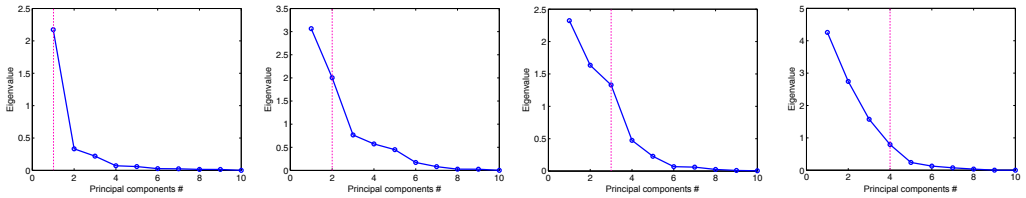


Figure 3.7: *Participants' preference for the type of presentation, according to our voting schema. A strong bias towards the audiovisual presentation can be observed especially for the tactile property pairs "rough–smooth" and "hard–soft", as well as for "simple–complex", "old–new" and "beautiful-ugly". There is no significant preference for the visual presentation in any pair, so the addition of sound does not deteriorate the perception.*

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 principal component analysis (PCA) for each type of presentation on the mean data. The factor loadings, as well as the explained variances of the first 3 principal components for each presentation are shown in Table 3.2. Furthermore, Figure 3.8 illustrates the corresponding scree plots, with the principal 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,



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

Figure 3.8: *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.*

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 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.

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

3.4 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.

3.4.1 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 3.3.2. 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.

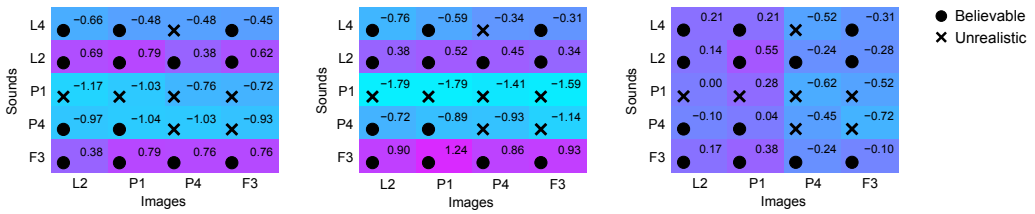
3.4.2 Results

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. Figure 3.9 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 3.9c 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 3.3. 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



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

Figure 3.9: *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.*

	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 3.3: *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.*

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".

3.5 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 prerecorded 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 [FWG13] 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.

3.6 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.

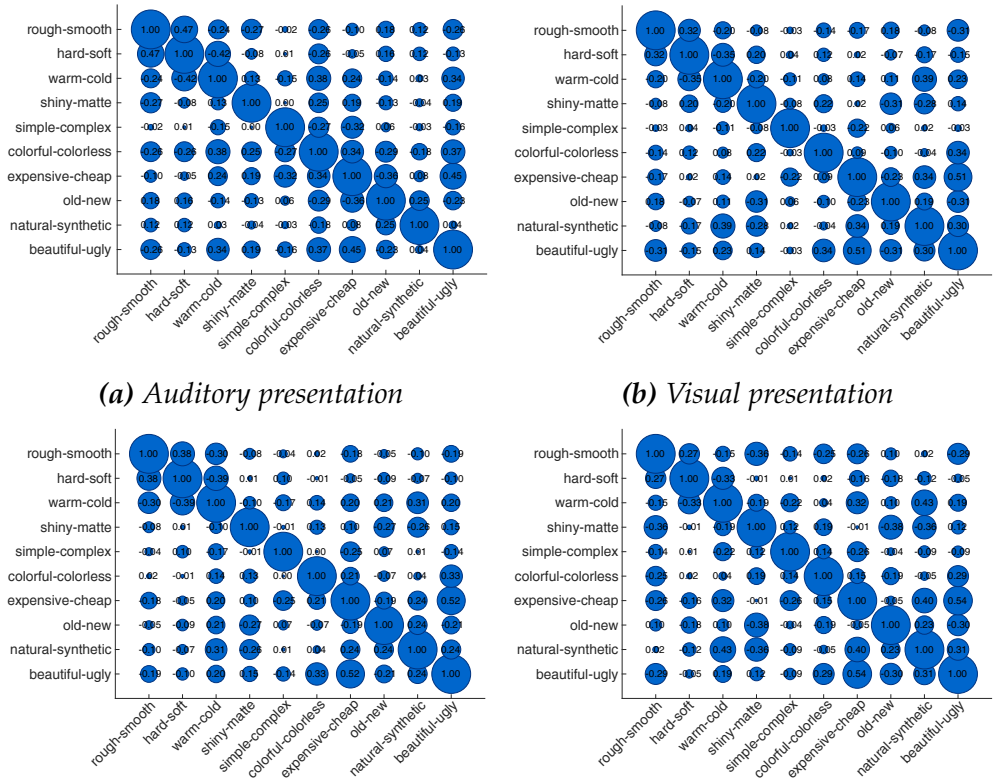
3.7 Acknowledgments

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).

3.8 Supplementary material

In this section, we introduce additional material to the paper 'Multimodal Perception of Material Properties' which is **not** included in the original text due to space constraints. First, additional research pertinent to the state of the art in tactile material perception is properly discussed. Then, we introduce the analysis of the correlation between the ratings for the studied material properties to later present a more detailed evaluation and visualization of the perceptual property space which also includes procrustes analysis. Finally, we show the means (also referred as Mean Opinion Scores, MOS) and confidence intervals (CI) for the participants' response ratings from Experiment 1 and the remaining mean responses of the manipulated audiovisual presentations from Experiment 2.

Relevant related work. A relevant study from Picard et al. [PDVG03] explored the perceptual dimensions of tactile interactions with fabrics (car seat cover materials) and the semantics associated with touch experiences. As reported by their their experimental results, haptic interactions with these materials imply a limited set of continuous perceptual dimensions (between 3 and 4), which are later interpreted by means of rating scales. According to this, seat cover materials present two orthogonal main dimensions, namely *soft/harsh* (first) and *thick/thin* (second). The third perceptual dimension (*relief*) and the fourth (*hardness*) were early understood as very closely related to the processing of the *soft/harsh* dimension.



(a) *Auditory presentation* ***(b)*** *Visual presentation*

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

Figure 3.10: *Correlation matrices between the set of properties analyzed in our experimental studies across all materials and participants. The size of the blob corresponds with the absolute value of these coefficients. Those values $r > 0.113$ are statistically significant with a 95% confidence level.*

Correlations between material properties. We would like to learn the degree of relationship between the set of material properties analyzed in our experiments. In particular, we are interested in the correlation between the adjectives belonging to the same group (tactile, visual and subjective) and how such values evolve through the different presentations. Essentially, this analysis would confirm whether the material property dimensions employed in our studies were correctly understood by the subjects, as some pairs of adjectives are intuitively expected to correlate better than others. Thus, we computed the correlation matrices per experimental presentation across all material samples and participants and present them in Figure 3.10.

Interestingly, subjective properties exhibit significant correlation values between each other, same as most of the properties categorized as

tactile. However, the correlation values between visual properties appears to be relatively low. Some of these relationships between properties are highly intuitive (e.g. smooth and soft, beautiful and expensive), while some others are not so straightforward (e.g. shiny and synthetic, natural and expensive). In general, all the considered presentations have very similar degree of correlation. Indeed, there is a very small margin between the presentation with the highest average correlation values (Auditory, $\hat{R} = 0.195$) and the one with the lowest values (Audiovisual, $\hat{R} = 0.160$).

Visualization of the perceptual space. In order to achieve a better understanding of the dimensionality of the perceptual space spanned by our set of perceptual properties (discussed in Section 3.3.2), we additionally show the projected factor loadings of the two first principal components (PCs) in Figure 3.11. Although the above-mentioned analysis demonstrated that each presentation has a varying number of underlying dimensions, here we consider only the two first ones for simplicity and clarity. The illustration shows how the subdimensional space revealed by the auditory presentation is clearly unidimensional, while the audiovisual and full-modal spaces allow a better visual clustering of the material classes.

We then compared the presentations by considering the full-modal space (FM_s) resulting from PCA as the ground-truth representation against the remaining presentations $P_s \in \{A_s, V_s, AV_s\}$ using procrustes analysis [Gow75]. This is a statistical technique that aims to map one multidimensional shape onto another by using linear transformations only. In this process, we considered all ten dimensions at our disposal and employed the minimized sum of squared errors (SSE) to measure the goodness of the mapping. The working hypothesis is that the better the mapping between P_s and FM_s is, the more accurately are real-world material interactions represented by a particular presentation (and its associated modalities). Interestingly, the best fit by far is achieved by the AV_s mapping ($SSE = 0.147$), followed by V_s ($SSE = 0.287$) and lastly A_s ($SSE = 0.504$). According to this, the addition of auditory input to the visual one is far more able to characterize real-world material information than visual or auditory input alone.

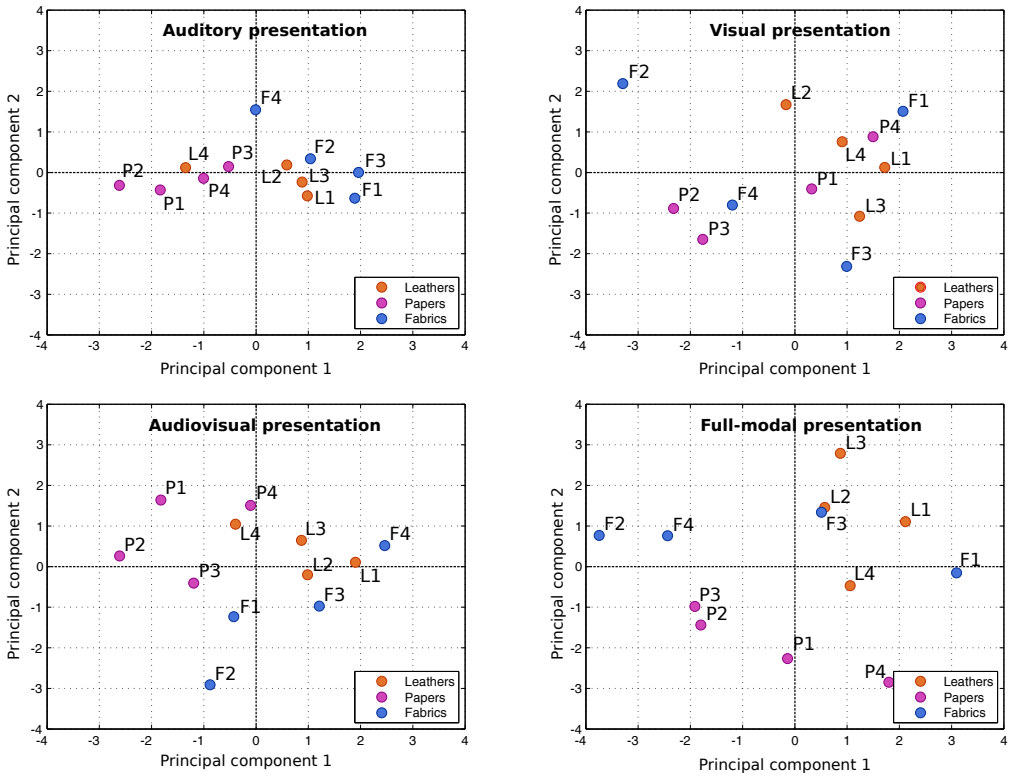


Figure 3.11: *Distribution of the samples in the first two PCs. The circles represent the projected positions of individual material samples in the subdimensional space for each presentation.*

Material ratings. The means and CI of the participants' response ratings from Experiment 1, arranged by material and category are depicted in the Figures 3.12 (for leathers), 3.13 (for fabrics) and 3.14 (for papers). The remaining mean responses for the audiovisual stimuli presented in Experiment 2, organized by property pair, are displayed in Figure 3.15.

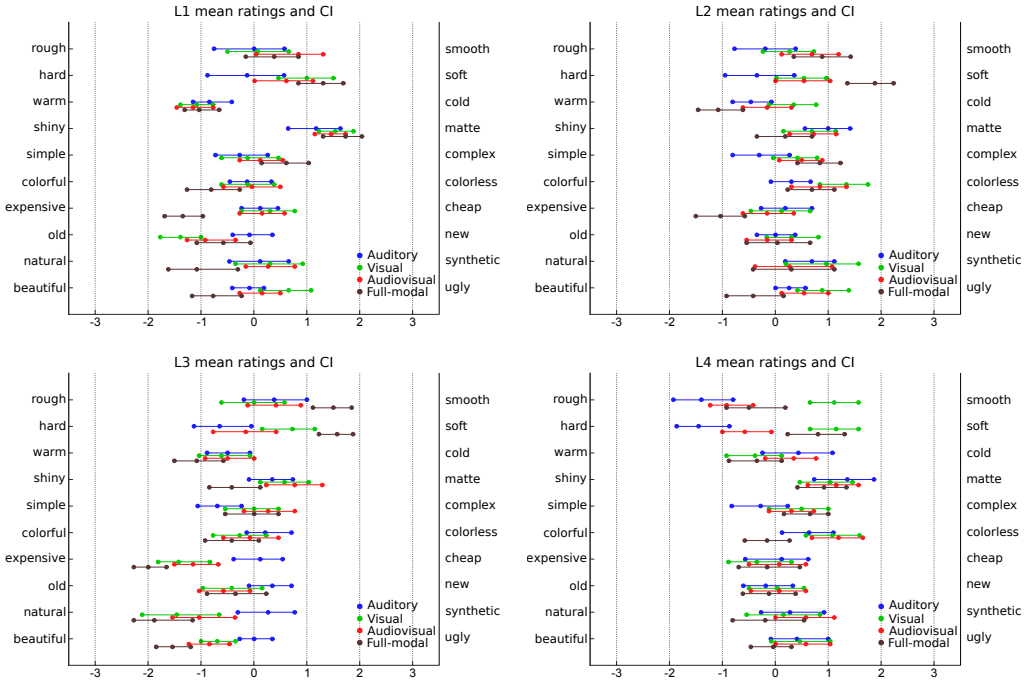


Figure 3.12: *Response ratings from Experiment 1 for each of the investigated material samples (leathers only). The central circles represent the participants' mean rating, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

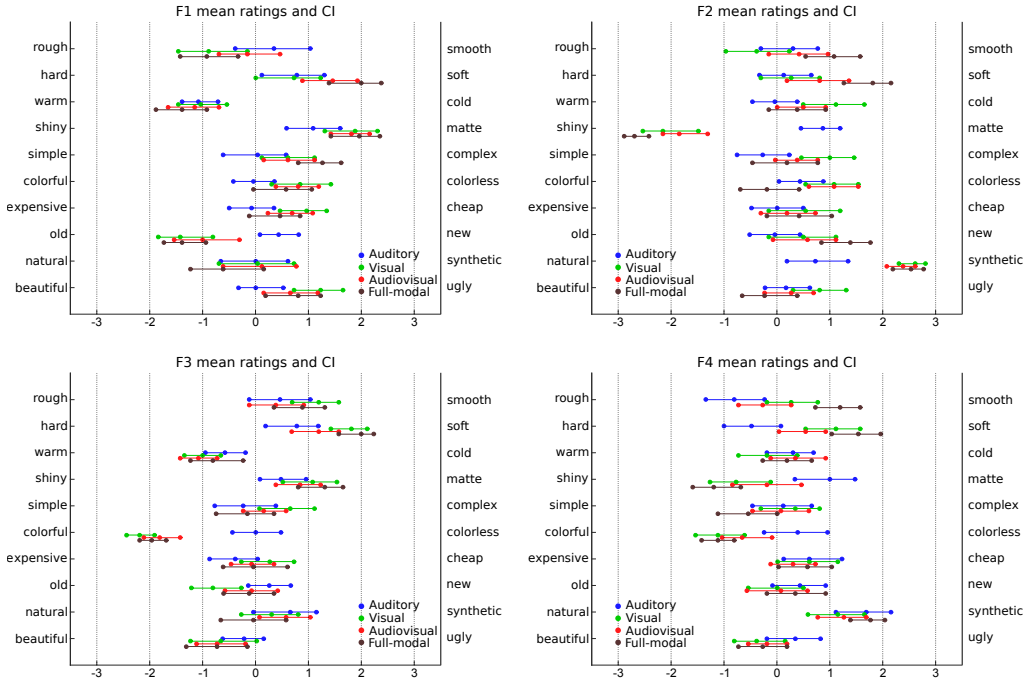


Figure 3.13: *Response ratings from Experiment 1 for each of the investigated material samples (fabrics only). The central circles represent the participants' mean rating, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

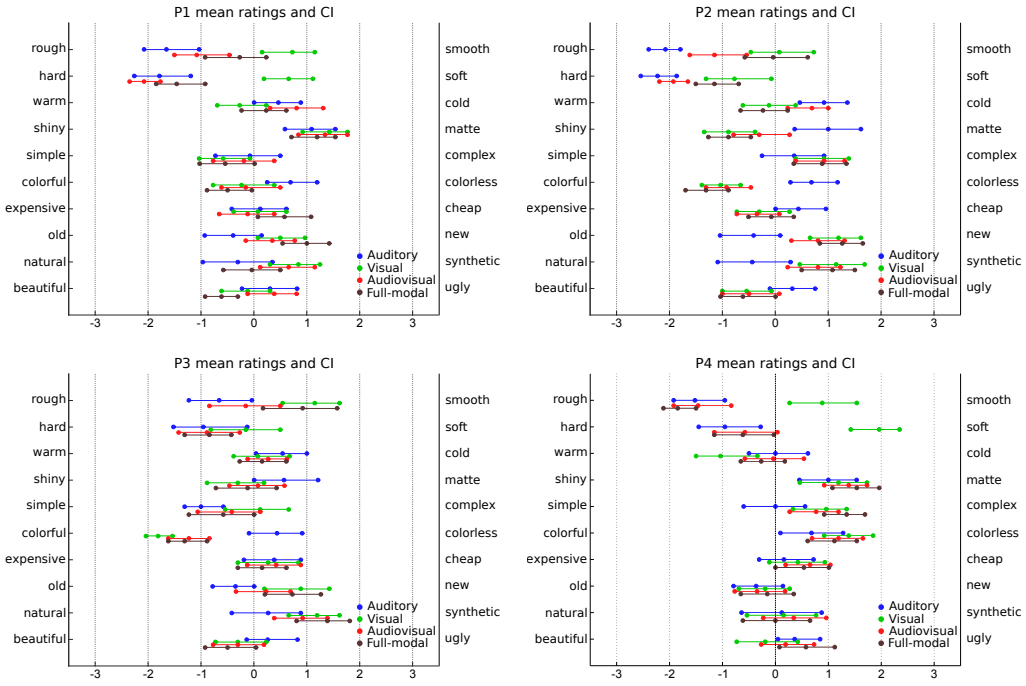
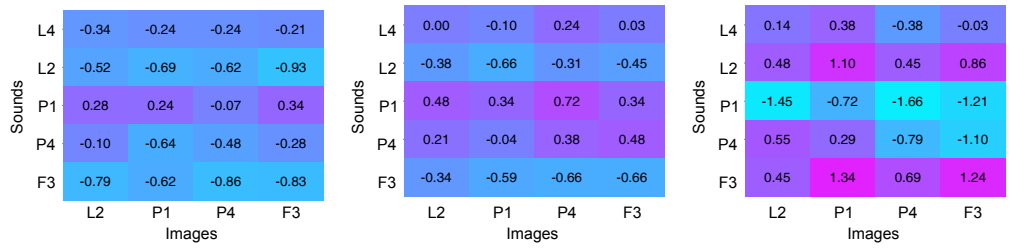


Figure 3.14: *Response ratings from Experiment 1 for each of the investigated material samples (papers only). The central circles represent the participants' mean rating, while the outer circles represent the bootstrapped 95% confidence interval for the mean. The bottom-right Figure for material P4 corresponds with Figure 3.6.*



(a) *Warm–cold* ***(b)*** *Beautiful–ugly* ***(c)*** *Unrealistic–believable*

Figure 3.15: *Mean ratings for the manipulated audiovisual presentations from Experiment 2 that were not presented in Figure 3.9. Rows indicate the corresponding auditory stimulus employed while columns indicate the visual stimulus. The mean ratings are color-coded and displayed in each cell.*

CHAPTER 4

Evaluating the Effects of Material Sonification in Tactile Devices

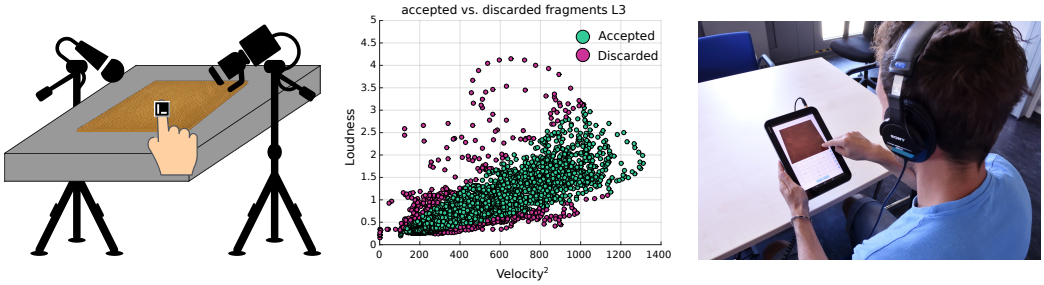


Figure 4.1: *(Teaser) From left to right, an illustration of the audio acquisition setup, the processing of the sound grains and the interface of the sonification software employed during the user experiments.*

Abstract. Since the integration of internet of things technologies in our daily lives, there has been an increasing demand for new ways to interact with commodities in the domain of e-commerce, that go beyond passive product visualization. When considering materials from retailing stores, the utilization of audio cues has proven to be a simple but effective way to enhance not only the estimation of well-established physical qualities (e.g. roughness, flexibility), but also affective properties (e.g. pleasantness, value), which have an important leverage in the user decision for or against a product. In this paper, we propose to investigate augmenting visual representations of leather and fabric materials with touch-related audio feedback generated when rubbing the fingertip against the samples. For this purpose, we developed an interactive material sonification

system for tactile devices that allows evaluating the impact of such audio cues on the human perception of materials by means of a psychophysical study based on rating scales for a set of relevant physical and affective material qualities. Our experimental results indicate that the evaluated touch-related audio cues do not significantly contribute to the perception of these attributes for the considered materials. In light of these findings, we suggest complementary directions of research in interactive material sonification, which may lead to profitable results.

Building on the outcome from the preceding chapter, we hypothesized that the presence of continuous, touch-related material sound, played back upon user interaction with a tactile device, would not only aid the estimation of tactile qualities, but also influence the overall affective experience. Thus we developed an interactive sonification algorithm for tactile devices and evaluated it from a perceptual point of view. One difference that turned out to be decisive w.r.t. the previous chapter is that, in this investigation, we only considered rubbing sounds from a more limited spectrum of materials, in order to narrow down the range of sounds that the synthesis algorithm is required to reproduce.

The contents of this chapter were published after the submission of this thesis [MWH18b]: Rodrigo Martín, Michael Weinmann and Matthias B. Hullin.* "A Study of Material Sonification in Touchscreen Devices". In *ACM International Conference on Interactive Surfaces and Spaces (ISS '18)* July 2018. Whose extended version has been also made available online [MWH18a].

4.1 Introduction

The shopping experience in our everyday life is determined by various types of interaction with commodities. A consumer's decision-making process in a store, for instance, is based on a multitude of inner rating processes that do not only involve the perception of physical properties through different senses such as sight, hearing or touch, but also an emotional or affective experience. In the context of online shopping and materials in particular, the sensory and emotional bandwidth of interaction with the respective commodity is greatly reduced and mostly limited to a passive visual representation and, occasionally, a textual description. Previous investigations have shown that the lack of a multimodal experience in general, and tactile input in particular, significantly affects the user's capability of assessing physical material properties (e.g. softness, flexibility) and developing affective emotions (e.g. pleasantness, value) evoked by the product [CJSC03]. As a logical consequence is desirable to enhance the digital material with additional cues on top of the purely visual user experience. In this regard, auditory cues and sonification techniques have demonstrated to compensate the absence of tactile interaction for digital material samples to some degree [HJK+13, MIWH15], without deteriorating its general impression. Instead of simply triggering a prerecorded audio sample of the interplay with the material, directly allowing customers to interact with the digital material, where audio information corresponding to this interaction is is automatically synthesized in real time represents an interesting challenge for the digital commerce.

The goal of this work is the analysis of the effects in the perception of physical and affective material qualities when visual material representations (photographs) are augmented with interactive audio feedback generated as the response to a single finger rubbing motion. Thereby, we employed a granular synthesis approach to build a sonification system that allows to enrich the user interaction with digital material samples through tactile devices. We then conducted a user study in which participants rated a set of relevant material qualities across a purely visual condition, two audiovisual conditions (including static pre-recorded sound and the interactive sonification system) and a full-modal condition, in which they were able to interact with the actual specimen. The experimental results were examined by means of the degree of correlation between participants, the analysis of the perceptual space spanned by the material qualities and the performance of each condition in a material classification task.

The key finding from this set of analyses is that the addition of rubbing material sounds as such does not seem to significantly improve the perception of material properties, although the overall material experience is not compromised by the presence of auditory cues. Considering these insights, we believe that future investigations in the domain of material sonification should contemplate additional material classes and audio cues from more user-centered material interactions. In accordance to this, alternative synthesis techniques able to deliver the sound corresponding to these interactions would have to be considered.

4.2 Related work

This section provides a summary of relevant investigations in the areas of multimodal perception of materials as well as sonification and sound synthesis of material interactions.

4.2.1 Multimodal perception of materials.

The majority of the research in product perception has been focused on the visual modality. Nevertheless, auditory cues have demonstrated to have a significant influence in the perception of product quality/efficiency (including electric toothbrushes, cars or foodstuff), and to be able to provide semantic signatures to a certain brand (e.g. the breaking sound of a 'Magnum' or the opening of a 'Schweppes' bottle) [SZ06]. Like any other product, the perception of materials is inherently multimodal and, with it, several strands of research have been conducted to investigate how the interplay between different senses shapes the perception of textures, materials and objects. An extensive review of the perception of textures regarding touch, vision and hearing is provided by Klatzky and Lederman [KL10], in which texture is understood as a perceptual property that characterizes the structural details of a surface.

There are not many approaches that focus on the investigation of purely acoustic material perception. Klatzky et al. [KPK00] analyzed the relationship between material perception and variables that govern the synthesis of impact sounds. Their results indicate the importance of a shape-invariant decay parameter in the perception of the material of which an object is made, while the frequency content plays also an important part. In a related fashion, Giordano and McAdams [GM06] studied the human performance when identifying materials from impact sounds. Interestingly, they concluded that listeners performed well with respect

to the gross categories, but their performance degraded for materials belonging to the same category.

Beyond the human performance in classifying materials, also the ability to infer concrete material qualities has received particular attention. Fleming et al. [FWG13] conducted a set of experiments to investigate the interactions between material classification and quality judgments. A high degree of consistency between these two assignments was detected, indicating that they facilitate one another by accessing the same perceptual information. The multisensory nature of the communication of material qualities has been further explored by Martín et al. [MIWH15], where the authors employed contact and stroking material sounds to complement the visual stimuli. Their results demonstrate the strong linkage between the auditory channel and the haptic perception to a point in which sound is capable of biasing the visual judgment of concrete qualities. In addition, Fujisaki et al. [FTK15] examined how a set of physical and affective qualities of wood are evaluated in three different modalities of vision, audition and touch, and observed that all three senses yield somewhat similar representations. Lastly, The qualities related to aesthetic perception of materials also play an important role in the decision-making process. In fact, strong connections have been observed, for instance, between the assessed smoothness of tactile textures and their perceived pleasantness [ESG14].

4.2.2 Sonification and synthesis of material sounds.

The synthesis of a multitude of sounds, from artificial to natural and pure musical ones, has important applications in movie sound effects, video games, virtual reality, multimedia or in art installations. An exhaustive survey on the predominant digital audio analysis and synthesis methods has been presented by Misra and Cook [MC09]. The former includes a taxonomy which introduces the most suitable synthesis approaches for each sort of sound and addresses the value and adaptability of the family of granular synthesis methods in the generation of audio textures. In fact, although granular methods have been mainly used in the creation of soundscapes, their possible range of applications include the synthesis of acoustic instruments, pitched sounds, speech, singing voice, and contact sounds from virtual surfaces when bouncing, being broken or scraped [BA02]. Belonging to the same family of techniques as granular synthesis, concatenative synthesis has been utilized in the context of simulating the particular sounds that certain materials produce. An et al. [AJM12] developed a motion-driven algorithm that is able to synthesize cloth sounds for

a wide range of animation scenarios. Their technique avoids expensive physics-based synthesis but still produces plausible results. However, it requires a certain amount of manual intervention and does not achieve interactivity.

In the context of visualization, information about the scene is represented in terms of shapes of varying sizes with attached color information and used to create pictures we can look at. Sonification is the equivalent concept translated to the sense of hearing, that is, the synthesis of non-speech audio to convey certain information. In this regard, the use of synthesized material sound has a large range of applications in sonification systems, which may be used to overcome limitations in representing tactile properties of digital objects and materials, among other purposes. An early investigation from Guest et al. [GCLS02] evaluated how tactile textures are perceived under real-time manipulation of touch related sounds. Their study indicates that the frequency content of textural sounds represents the dominant factor for such sort of interactions as e.g. attenuating high frequencies caused the textures to be perceived smoother. Later, Tajadura-Jiménez et al. [TJL+14] explored the ability of a sound-based interaction technique to alter the perceived material of which a touched surface is made. Their granular sonification algorithm reproduces samples (grains) with three different frequency levels where the grain selection is guided by the finger pressure on a wooden surface. With their results, the authors determined that increasing sound frequency alters either the surface perception (colder material) and the emotional response (increased pressure and touch speed). Finally, by using textural sounds in the context of a retail clothing application, Ho et al. [HJK+13] demonstrated that the simple addition of realistic auditory feedback to the unimodal visual experience favors the feeling of immersion, which becomes evident in longer interaction times with the product and also willingness to pay a higher price for it.

The present investigation establishes, to our knowledge, a novel and interactive approach to material sonification with consumer hardware and the first assessment of its effects on the perception of physical and affective material qualities. We hence arranged a user experiment to evaluate such effects in comparison to additional stimuli, including the actual material samples. The description of our experimental setup is introduced in the following section.

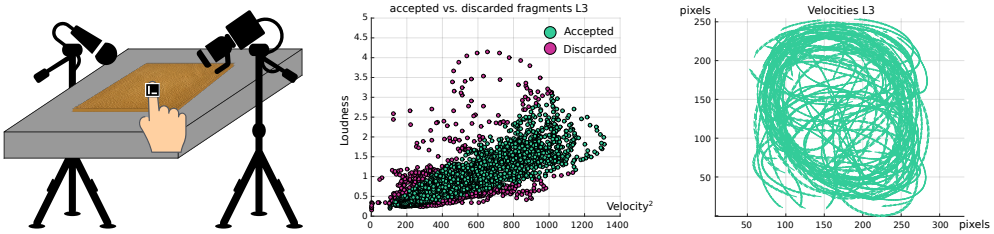
Figure 4.2: *Pictures of the leathers materials (upper row L1–L5) and fabrics (lower row F1–F5) as displayed in our study.*

4.3 Experimental design

The key elements of our experimental design are given by the considered visual stimuli, the auditory stimuli, the actual material samples and the description of the user study. In the following, we provide details regarding each of these components.

4.3.1 Visual stimuli

In the scope of this research, we explore the perception of physical and affective material qualities for two semantic classes (leathers and fabrics), which are commonly available in retailing websites. For this purpose, we have chosen ten material samples, each of them with an approximate size of 120×120 mm^2, with nearly flat geometry to avoid possible sources of visual variability and sound artifacts. Restricting our selection to these two concrete, well-known classes allows us to keep the study and its conclusions manageable. We then situated each specimen on a bright background under natural illumination and took a picture using a digital camera (Nikon1 J5, resolution 5568×3712 pixels) located at approximately 200 mm from the sample under a slight angle. The resulting images were corrected regarding white-balance and scaled to match the resolution of the final device (see Section 4.3.5). The characteristic borders of each specimen were additionally cropped, since they have been demonstrated to provide supplementary information to the material texture that could bias the visual stimuli [MWH17]. The resulting photographs are displayed in Figure 4.2.



(a) *Acquisition setup.* **(b)** *Fragment loudness vs. velocity*^{2}. **(c)** *Velocities of accepted fragments.*

(d) *Participant during the study.* **(e)** *Elements of the user interface.*

Figure 4.3: *Illustration of the relevant steps from the synthesis process and the experimental study, including the acquisition of touch-related material sound (a), pre-processing phase for a concrete material L5 (b) and (c), a picture of a participant during the study (d) and the elements composing the user interface (e).*

4.3.2 Auditory stimuli: acquisition of material sound.

In order to record the touch-related sounds arising from the interaction with the selected materials (i.e. brushing them with the fingertip), we assembled a setup composed by a piece of polyurethane foam with the size of 400×400 mm^2, on which the sample was placed. The actual recording step was carried out in an acoustically isolated chamber using a directional microphone (Beyerdynamic MCE 86 II) located about 20 cm away from the sample and facing towards it. The contact sounds were generated by gently rubbing the material's surface creating random trajectories while increasing the velocity of the movement for roughly a minute. Other than trimming, no further post-processing was applied to the audio. Such recordings were employed as the static audio in one of our experimental conditions (see Section 4.3.5).

To later guide our sound synthesis method, we annotated the resulting signal with the position of the finger during the interaction. We achieved this by attaching a fiducial marker [GJMS+14] to the nail of the interacting finger, which we tracked using a machine vision camera (Point Grey GS3-U3-23S6M-C Grasshopper). An illustration of the complete setup is

depicted in Figure 4.3a. After basic analysis of the video data (marker tracking, trajectory smoothing, numerical differentiation), we thus obtained 2D finger velocity data $\vec{v}_i$ at 100 samples per second (see Figure 4.3c) along with the 48 kHz stereo audio clip $s(t)$.

4.3.3 Auditory stimuli: synthesis of material sounds.

To deliver material contact sounds in response to the user's input in real-time, we developed a sonification algorithm based on granular synthesis. This family of techniques has been broadly used in many applications due to their flexibility [MC09], including the synthesis of material contact sounds [BA02]. The moderate computational load, in comparison to physically-based approaches, represents a decisive aspect as it allows to perform the synthesis on consumer hardware (such as tablet computers) at interactive rates, thus facilitating the present study. To this end, we employed the open-source audio processing language `Pure Data` (Pd)[1] as an embeddable library (`Libpd`) under Android.

We divided the recorded sound clip $s(t)$ into *fragments* with a length of 480 samples, corresponding to the spacing of velocity samples (10 ms). Each such fragment was annotated by velocity values $\vec{v}_j$ (see above) and its root-mean-square (RMS) loudness a_j, noting that the loudness of a fragment roughly scales with the square of the corresponding velocity. To avoid artifacts when re-synthesizing fragments into a new audio stream, we discarded those fragments that were unusually soft or unusually loud for the given velocity. We identified such outliers by computing the ratio $\alpha_j = a_j / |v_j|^2$ for each fragment, and then removed those fragments whose ratio was below the 5^{th} percentile or beyond the 95^{th} percentile (see Figure 4.3b). The remaining set of annotated fragments constitute the input to the sonification system.

During user interaction with the tactile device, the touch interface measures the user's finger velocity $\vec{v}_{\text{in}}$ on the screen. The granular synthesis uses this value to retrieve suitable sound fragments according to a distance metric that considers the velocity and loudness of the j^{th} fragment:

$$d_j = \sqrt{\left\|\vec{v}_{\text{in}} - \vec{v}_j\right\|_2^2 + (|v|_{\text{in}}^2 - a_j/\hat{\alpha})}, \tag{4.1}$$

where $\hat{\alpha}$ is the mean of the ratio α_j across all fragments. To ensure variation, we follow a standard practice in granular synthesis by retrieving not

[1] `Pure Data` (Pd) is an open source visual programming language for multimedia. For more information and resources we refer to the corresponding webpage: http://libpd.cc

Tactile	Visual	Affective
rough–smooth	shiny–matte	expensive–cheap
hard–soft	bright–dark	natural–synthetic
thick–thin	transparent–opaque	beautiful–ugly
stiff–flexible	homogeneous–heterogeneous	unpleasant–pleasant

Table 4.1: *Opposite-meaning quality pairs, grouped by category.*

only the single closest hit for the given query velocity $\vec{v}_{\text{in}}$, but the $k = 25$ closest fragments instead. With the goal of real-time operation in mind, this k-nearest-neighbor search is implemented using a balanced binary space partitioning (BSP) tree [FKN80]. The synthesis algorithm randomly selects one of these fragments and "freezes" it for the upcoming few iterations to avoid repetition artifacts. The fragment is then extended into a longer *grain* by incorporating its $n = 28$ neighbor fragments in the input sound clip. Finally, the grain is concatenated and blended (cross-faded) with the previous grain into a continuous audio output.

Overall, this rather simple system is capable of producing a smooth and interactive stream of contact sound that is free of disturbing artifacts (transitions, repetition) that are otherwise typical for granular synthesis. A major drawback of the mobile platform remains the somewhat long system latency of approximately 500 ms, inherent to the utilization of the `Libpd` library under Android.

4.3.4 Real materials

During the progress of the experiments, participants were also asked to evaluate the actual materials samples (full-modal interaction). Instead of using the same specimens utilized during the audiovisual stimuli acquisition, smaller portions of the same samples (approximately 70×70 mm^2) were handed to the users. With this, we avoided damaging the originals during the interactions and facilitated the scalability of the experiment.

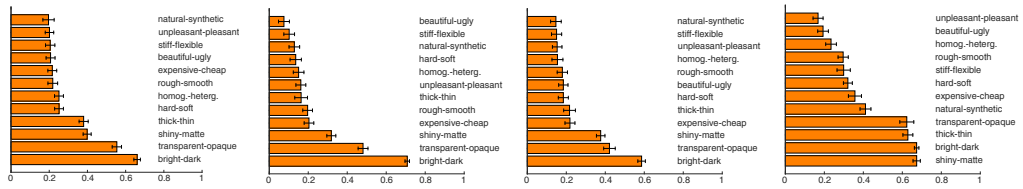
4.3.5 Task and procedure

Inspired by previous investigations [FWG13, MIWH15, MWH17] we gathered a collection of 24 adjectives describing material appearance. At the same time, these adjectives were organized into 12 opponent pairs which were assigned to either the tactile, visual or affective category, depending on the nature of the physical or emotional interaction that best reveals

them (see Table 4.1). In order to rate this set of qualities across our multimodal stimuli, we made use of single stimulus ratings in which the participants assessed each quality pair under study on a 7-point Likert scale, represented with a slider with values ranging from -3 to 3. The values along the scale were consistently labeled with a term indicating the intensity of the stimuli (e.g., very hard, hard, a bit hard, neutral, a bit soft, soft and very soft). The user study was conducted using tablet computers (Toshiba Excite Pro 10.1, resolution 2560×1600 pixels) and a set of headphones (Sony MDR-7506) running a custom Android application which connects with the Pd module. The complete experimental setup can be seen in Figure 4.3d and the user interface is depicted in Figure 4.3e with greater detail. The procedure itself consisted of four presentations or conditions, in which material stimuli were presented in random order to the participants along with the 12 slider widgets. The four conditions that compose the study are the following:

- *Visual condition (VI)*, where the stimuli are photos taken from real materials.
- *Static Audiovisual condition (SA)*, where the photos were complemented with prerecorded audio from the material.
- *Dynamic Audiovisual condition (DA)*, where the photos were complemented with interactive sound generated by our sonification system. This means that real-time contact sound is played back upon tactile interaction with the images on the device.
- *Full-modal condition (FM)*, consisting of physical material samples that were given to the participants so that they could interact with them.

Since the interaction with the real samples could bias the realization of the visual and audiovisual conditions, the full-modal presentation was constrained to be the fourth and final one, while the order of the remaining conditions was randomized. 19 participants took part in the experiment (12 females, mean age 27.08; 7 males, mean age 28.57). All the participants were naïve to the goals of the experiment, provided informed consent, reported normal or corrected-to normal visual and hearing acuity and were compensated economically for their cooperation. From this experiment, a total of $19 \times 10 \times 12 \times 4 = 9120$ responses were collected and evaluated.



(a) *Visual* ***(b)*** *Av. Static* ***(c)*** *Av. Dynamic* ***(d)*** *Full-modal*

Figure 4.4: *Average inter-participant correlation per property, grouped by condition and sorted in ascending order w.r.t. the correlation. Note that the differences between the visual and the two audiovisual conditions are relatively small and how the full-modal condition presents significantly higher correlation values.*

4.4 Results

In order to investigate the effects of our sonification system on the perception of material qualities, we evaluate the correlation between the participants' ratings, the dimensionality of the spanned perceptual space per experimental condition and the performance in a classification task based on the material quality ratings.

4.4.1 Inter-participant correlation

Due to the diverse collection of materials and qualities considered in this investigation, we first provide an analysis of the level of agreement between the participants' ratings for the given stimuli. To that end, we computed the inter-participant correlation coefficients for each condition and quality pair, over all materials. The hypothesis assumption is that the higher the correlation coefficient, the better a specific quality would be represented by the condition at hand. Contrarily, if such quality is not well depicted, the users would have to infer it using their imagination, resulting in a lower degree of agreement. Figure 4.4 illustrates the resulting correlations in ascending order, separated by experimental condition.

The largest coefficients presented by the visual condition are those corresponding to the pairs "bright–dark", "transparent–opaque", "shiny–matte" and "thick–thin", which are properties mostly categorized as visual. Likewise, both audiovisual condition exhibit the largest correlation values for the pairs "bright–dark", "transparent–opaque" and "shiny–matte". However, the "thick–thin" dimension shows a much lower value for the DA and particularly the SA condition in comparison to the VI presentation. Allegedly, the proposed rubbing/stroking sounds employed

are less suitable for communicating this particular dimension and seem to mislead users' judgments of the material thickness. This is further implied by the correlation values for the full-modal condition, where this pair shows again a significant level of agreement. Another interesting observation is that the user agreement for the tactile qualities as well as the pair "beautiful–ugly" is slightly higher in the DA condition when compared to the static audio (SA). Albeit being a promising trend, the effect is not significant enough to draw categorical conclusions. In general, the correlation values (R) and ordering are quite similar for the three digital conditions VI ($\hat{R}_{VI} = 0.32$), SA ($\hat{R}_{SA} = 0.27$) and DA ($\hat{R}_{DA} = 0.28$), and follow a comparable ordering as the full-modal condition ($\hat{R}_{FM} = 0.46$).

Although in principle this analysis is analogous to the inter-participant correlation from Martín et at. [MIWH15], the results are not directly comparable, as neither the stimuli employed nor the quality set are entirely identical in both experiments. Specifically, the pictures from the visual presentation in the former investigation display the distinctive borders of materials, which are known to be a powerful discriminator [MWH17]. Furthermore, the authors included tapping impact sounds in their audio-visual condition, which possibly allowed the inference of additional material information. Taking this into consideration, the larger discrepancy between both studies concerns to the resulting correlation coefficients for the "hard–soft" dimension, where the present experiment exhibits considerably lower values for the SA, DA and FM conditions. We conclude that tapping sounds provided decisive cues to assess the hardness of the material. Moreover, the presence of relatively hard paper materials in [MIWH15] probably established an upper bound for this concrete quality, which is not present when solely considering leathers and fabrics.

4.4.2 Dimensionality of the perceptual space

In the previous section, we examined the four different conditions through the correlation between participants, observing little effects between the conditions VI, SA and DA. To further explore this insight, we analyze the dimensionality of the perceptual space spanned by the perceptual qualities. For this purpose, we averaged the ratings over all participants and performed principal component analysis (PCA) for each experimental condition on the mean data. The resulting factor loadings of the first three principal components as well as the explained and accumulated variance are shown in Table 4.2, separated by condition.

	Visual			Static audiovisual			Dynamic audiovisual			Full-modal		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
rough–smooth	−0.235	**−0.351**	0.177	−0.008	**0.483**	0.161	0.001	**0.445**	0.191	−0.163	**0.403**	0.132
hard–soft	−0.233	−0.063	−0.106	0.115	0.258	−0.173	0.001	0.254	**0.535**	−0.125	−0.165	0.306
thick–thin	**−0.390**	0.065	**0.359**	−0.265	0.182	0.344	−0.329	0.241	−0.033	**−0.466**	0.179	**0.372**
stiff–flexible	−0.301	−0.112	0.052	−0.036	0.255	−0.028	−0.116	0.297	**0.383**	−0.192	−0.085	0.154
shiny–matte	**0.404**	−0.090	0.065	**0.395**	−0.188	−0.037	**0.417**	−0.052	−0.157	**0.418**	**−0.387**	**0.372**
bright–dark	**0.478**	−0.134	**0.753**	**0.599**	−0.178	**0.657**	**0.584**	0.139	**0.351**	0.213	−0.073	**0.664**
transparent–opaque	**0.361**	**0.364**	−0.265	**0.521**	0.038	−0.330	**0.522**	0.034	−0.005	**0.508**	0.131	−0.242
homog.–heterog.	0.262	0.041	**−0.397**	0.131	−0.080	**−0.466**	0.209	−0.184	0.059	0.106	**−0.390**	−0.006
expensive–cheap	0.011	**0.446**	−0.044	−0.233	**−0.390**	−0.043	−0.105	**−0.360**	0.337	−0.274	**−0.381**	−0.027
natural–synthetic	−0.033	**0.451**	0.151	−0.220	−0.279	0.272	0.178	−0.283	**0.431**	**−0.356**	−0.310	−0.159
beautiful–ugly	0.141	**0.403**	−0.036	−0.070	**−0.351**	−0.072	−0.020	**−0.437**	0.277	−0.088	**−0.349**	−0.120
pleasant-unpleasant	−0.189	**−0.361**	0.024	0.063	**0.418**	0.050	−0.061	**0.364**	−0.022	0.065	0.290	0.221
Explained variance [%]	41.02	28.75	12.70	41.35	25.94	12.46	38.35	28.64	13.13	60.97	17.53	8.69
Cumulative variance [%]	41.02	69.78	82.49	41.35	67.29	79.76	38.35	67.01	80.15	60.97	78.51	87.20

Table 4.2: *Factor loadings, explained variance and cumulative variance of the first three principal components for each condition. Bold, underlined values represent the strongest factors (greater than* 0.35*) for each principal component.*

A detailed inspection of the coefficients exposes that the first principal components (PC1) in all three digital conditions (VI, SA, DA) are dominated by the visual qualities ("shiny–matte", "bright–dark" and "transparent–opaque"), which account for most of the variation in the users' ratings. Additionally, VI exhibits somewhat large values in the tactile dimensions, especially for the "thick–thin" pair, which are not so evident in SA or DA. This is in accordance with the correlation values reported earlier. Furthermore, the second PC of all three conditions is commonly determined by the "rough–smooth" quality and the affective properties, while PC3 has diverse values for each condition. In contrast, the first PC of the full-modal condition is driven by a mixture of qualities ("thick–thin", "shiny–matte", "transparent–opaque" and "natural–synthetic"), while the second PC explains much less variance and is dominated by the roughness, shininess, heterogeneity and the affective qualities.

When considering the cumulative variance, two dimensions are able to explain 69.78%, 67.29%, 67.01% and 78.51% of the variance for the VI, SA, DA and FM conditions respectively. Therefore, projecting the factor loadings into a 2-dimensional space seems to be a plausible and easy-to-visualize option to analyze the distribution of the user data (see Figure 4.5). By inspecting the arrangement of materials in the subdimensional space, we observe that the sample distributions presented by VI, SA and DA are quite similar (PC2 in VI is upside-down). Meanwhile, the variance in FM is primarily accumulated in the first PC, which allows a smooth clustering of the two material classes.

Lastly, we applied procrustes analysis [Gow75] to compare our experimental conditions $C_s \in \{VI_s, AS_s, AD_s\}$ against the full-modal space (FM_s) resulting from PCA, which is taken as a reference. We took the 12-dimensional space spanned by the qualities into account and used the minimized sum of square errors (SSE) to measure the goodness of the mapping. The fitting values for all three C_s are, again, extremely similar with a relatively low error, where VI_s achieves the best result ($SSE = 0.214$) closely followed by SA_s ($SSE = 0.240$) and DA_s ($SSE = 0.242$). From this analysis we conclude that the considered visual (VI) and audiovisual (SA, DA) stimuli are capable of effectively transmitting information about our set of materials and qualities. However, the addition of these specific audio cues, no matter whether in terms of their static form or the sonification system, does not contribute with significant additional information to simple photographs.

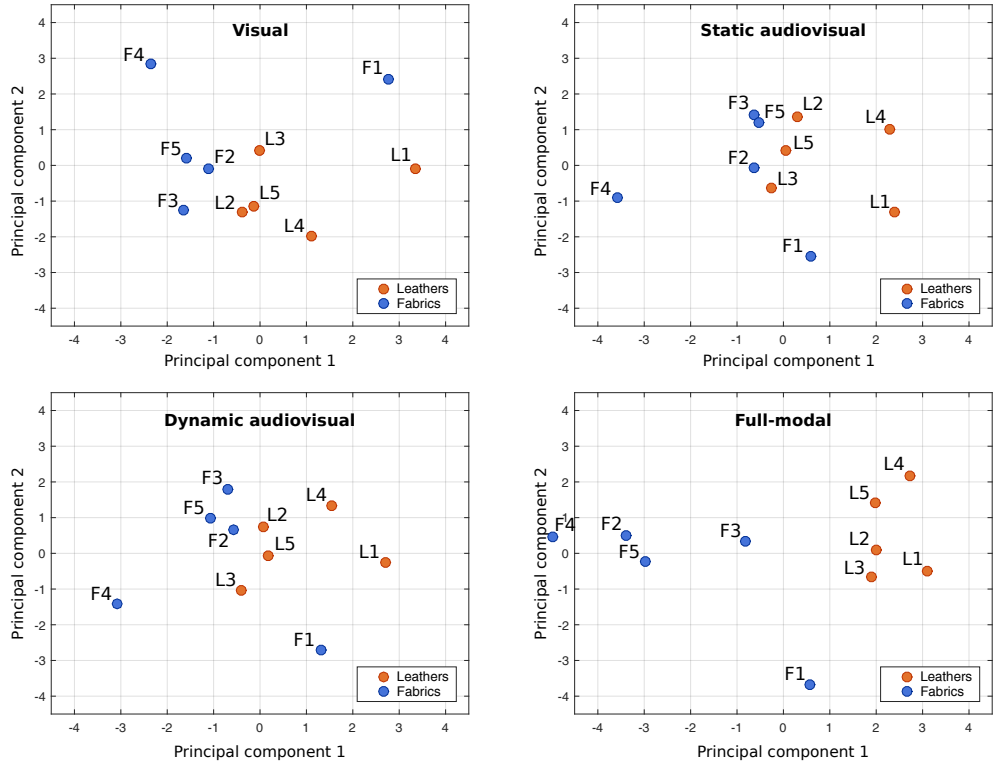


Figure 4.5: *Distribution of the samples in the first two PCs. Circles represent the projected positions of individual material samples (10 in total) in the subdimensional space for each condition. The sample distribution presented by VI, SA and DA are rather identical, while the variance in FM corresponds mostly to the first PC.*

4.4.3 Material classification

The previous analysis facilitated the understanding of the ability of the considered stimuli to depict a set of relevant material qualities through the agreement level between subjects and the subdimensional space that they span. Previous studies have demonstrated that humans access the same perceptual information about materials while performing both material categorization and quality rating tasks [FWG13]. Keeping this in mind and considering that our stimuli consist of two classes of materials, this section attempts to clarify to what extent such classes can be by predicted based on the participants' ratings. Concretely, we aim at answering the following questions:

1) Which is the classification performance of the experimental condi-

	Set of Predictors						
[%]	**All**	**(T)actile**	**(V)isual**	**(A)ffective**	**T+V**	**T+A**	**V+A**
VI	75.8%	63.2%	67.9%	65.3%	70.5%	71.6%	73.7%
SA	72%	56.1%	66.1%	65.1%	68.8%	61.4%	69.8%
DA	66.3%	60.5%	61.6%	62.6%	66.3%	69.0%	69.5%
FM	89.5%	78.4%	82.6%	72.6%	89.5%	84.2%	89.5%

Table 4.3: *Accuracy [%] of a SVM material classifier based on the perceptual qualities. Each row represents the accuracy for the considered experimental condition, while the columns describe the set of qualities used as predictors.*

tions (VI, SA, DA) in comparison to the FM condition?

2) Do any of the utilized sound cues facilitate the discrimination between leathers and fabrics?

3) Which set of considered qualities allows a better material classification?

For this purpose, we trained a binary Support Vector Machine (SVM) classifier to obtain a model which employs the user ratings of the twelve perceptual features to predict the material class to which each sample belongs. We then conducted leave-one-out cross validation tests per user and material sample. Additionally, we performed the same analysis using the ratings from each perceptual category individually as predictors (tactile, visual and affective) as well as all three combinations of them. The accuracy across each condition and group of qualities is provided in Table 4.3.

Regarding the first inquiry, performing the classification task on the ratings from the FM condition results into considerably higher accuracies (at least above 72%) in comparison to the rest of the conditions. This outcome is to be expected, as judging the real material samples will always allow a more confident quality assessment as images or sounds. With respect to question number two, the classification results of the SA and DA conditions exhibit lower values as the visual condition for all the set of predictors considered, by a slight margin. In light of this results, we assume that the addition of rubbing sounds does not help in distinguishing leathers and fabrics, and that the proposed sonification system has additional value over static sounds only when tactile-related predictors are included. Concerning the third question, using all twelve perceptual

qualities as predictors yields by far higher accuracies. The table shows also how the visual predictors, alone or in combination with other features, have more discriminating power than tactile or affective qualities. More interestingly and less anticipated is the fact that the use of affective predictors lead to higher accuracies than tactile ones, when the conditions VI, SA and DA are considered. However, tactile features provide better discrimination in the full-modal case, since the participants were able to actually touch the specimens.

4.5 Discussion and future work

After a comprehensive analysis of the collected user data, the primary finding of the present study is that the auditory cues employed in our studies do not contribute with additional value to the perception of material qualities. All of the conducted evaluations indicate that the three digital conditions evaluated (VI, SA and DA) have a fairly equal ability transmit material information without weakening the overall experience. The most plausible explanation for this outcome may be that the utilization of contact sounds from rubbing interactions exclusively did not yield enough information to discriminate between the two different material classes or to characterize specimens within the same class. Indeed, previous investigations concerning the perception of textiles have asessed other gestures like two-finger pinching, stroking or sample scrunching as the most repeated interactions when evaluating real fabric samples [AOP+13]. Another reason that may have influenced our results is the fact that the employed material classes (leathers and fabrics) do not differ significantly when considering their characteristic sounds. Although our research aimed at examining the perceived intra-class differences between materials, it has been documented that not even striking sounds provide sufficient cues to differentiate samples within the same class [GM06].

Another interesting finding concerns the ability to discriminate between leathers and fabrics through the experimental conditions. Both the PCA analysis and the SVM classification indicate that it is possible to discern between these two classes based on the ratings for the selected set of attributes, when the real materials are provided. However, this capacity is not translated well to the conditions where only images and sound from the material samples are provided. As regards to which types of qualities allow better material discrimination, considering the visual quality features alone provide the highest accuracies. Interestingly, the presence affective features has certain influence in the digital conditions (VI, SA

and DA) in comparison to tactile qualities, which are more salient when the real materials are provided. This supports our intuition that affective properties have a meaningful role in the perception of digital products.

During informal interviews after the realization of the experimental task, subjects reported to have enjoyed the utilization of the tactile sonification interface. This reported amusement translated into significantly longer interaction times with the system during the DA condition, 55.5 seconds of average interaction per material, in contrast to the 32.3 and 28.6 seconds on average for the VI and SA conditions respectively. This is in accordance with the experimental results from Ho et al. [HJK+13], where the addition of realistic auditory feedback led to considerably longer (30%) interplays with their AR system. Moreover, the similarity of the ratings between all these three conditions dismisses the possibility that such effect is due to the complexity of the task. Nonetheless, some of them declared to be a bit puzzled by the latency of the sounds (about half a second), intrinsic to the usage of `Libpd` under Android. Indeed, the investigation of how the presence of interactive sounds affects the level of engagement when exploring digital materials remains a promising avenue for future research.

Given the previous analysis and considerations, we conclude that the main shortcoming of our sonification method is not the audio synthesis itself, but the utilized rubbing/scrapping material sounds, which do not bring in supplementary information for the selected set of materials. Future investigations may turn over to other kinds of touch-related material sounds, more consistent with the actual human behavior, for which alternative synthesis approaches could be more suited. For instance, physically-based synthesis methods have been able to generate the distinctive crumpling sound of materials [CLG+16] at, however, unfeasible computation times. Deep learning techniques could also leverage the synthesis of contact sounds [OIM+16], provided that a sufficiently rich database of sounds is given for the training of the model.

4.6 Conclusion

The main purpose of this investigation is to determine the impact of an interactive material sonification system in the perception of physical and affective material qualities. For the development of this sonification algorithm, we relied on granular synthesis to interactively reproduce characteristic contact sounds generated when rubbing leather and fabric materials with the fingertip. This method, which has been specifically devel-

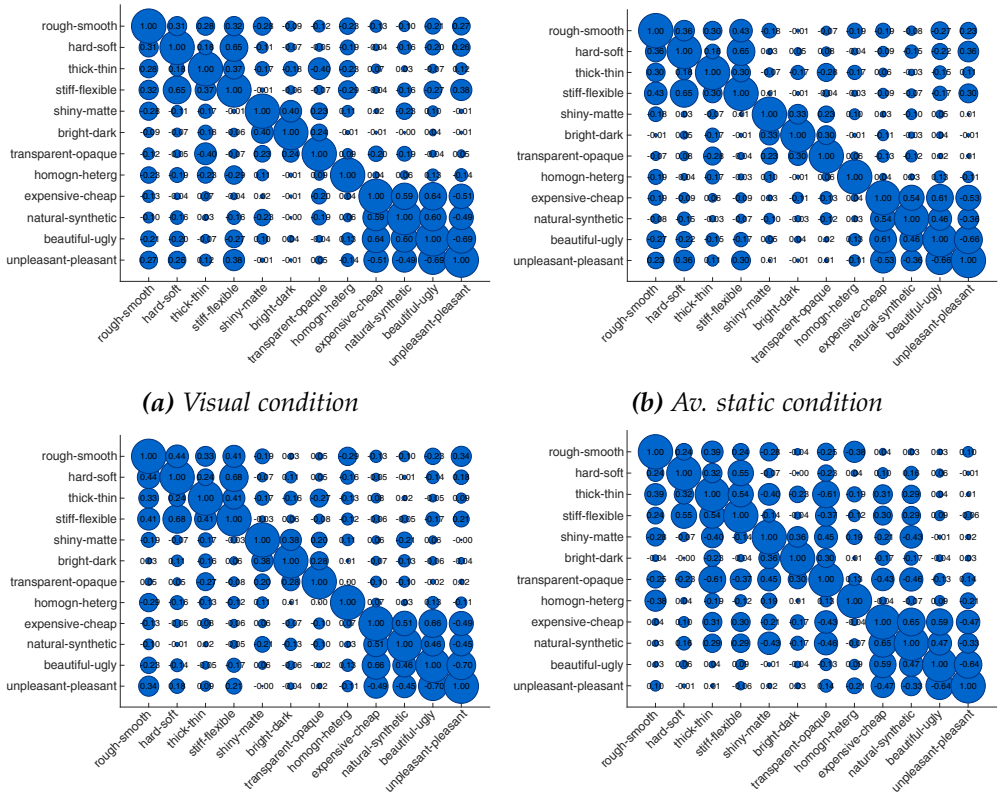
oped for tactile devices, plays back chunks of sound (grains) upon tactile interaction with the material images on the screen. Its performance has been then evaluated by examining its ability to describe concrete material qualities in contrast to additional visual, audiovisual and full-modal conditions, via a psychophysical study. We discovered that the contact sounds employed in our experiment do not contribute with additional information to the perception of material qualities, since all the considered digital condition exhibit almost similar performance. In light of these findings, we provide several potential lines of future research regarding sonification systems for digital materials, which may include additional material categories, more suitable types of interaction and alternative synthesis techniques.

4.7 Supplementary material

In this section, we introduce additional material to the paper 'Evaluating the Effects of Material Sonification in Tactile Devices' which is **not** included in the original text. First, we introduce the analysis of the correlation between the ratings for each of the respective qualities. Then, we present the mean opinion scores (MOS) and confidence intervals (CI) of the participants' response ratings.

Correlations between material properties. The reason for this analysis is to comprehend the degree of relationship between the different material qualities analyzed in this study. In particular, we would like to analyze the correlation between the adjectives belonging to the same group (tactile, visual and subjective) and how the resulting coefficients vary through the different conditions. Essentially, this examination would confirm whether the explored qualities were well understood by the participants, as some pairs of adjectives are intuitively expected to correlate better than others. In this manner, we computed the correlation matrices per experimental condition across all materials and participants, which are displayed in Figure 4.6.

According to the figures, the evaluated affective qualities are highly correlated between each other. This is to be expected since it is reasonable to regard an expensive material as beautiful, pleasant and natural as well. We can also observe an important relationship among tactile qualities themselves and other intuitive associations (e.g. transparent with thin or homogeneous with smooth). Moreover, the coefficients yield some additional relationships that are not so obvious a priori (e.g. expensive and



(a) *Visual condition* ***(b)*** *Av. static condition*

(c) *Av. dynamic condition* ***(d)*** *Full-modal condition*

Figure 4.6: *Correlation matrices between the set of qualities analyzed in our experimental study across all materials and participants. The size of the blob represents the absolute value of the coefficient r. Those values $r > 0.159$ are statistically significant with a 95% confidence level.*

opaque, thin and shiny). In general, the FM condition presents the highest correlation values on average ($\hat{R}_{FM} = 0.225$) followed by VI ($\hat{R}_{VI} = 0.197$), DA ($\hat{R}_{DA} = 0.180$) and finally SA ($\hat{R}_{SA} = 0.173$).

Material ratings. The MOS and CI of the participants' response ratings for each experimental condition arranged by material and category are depicted in Figure 4.7 (for leathers) and 4.8 (for fabrics). The accuracy of this interval depends on the normality of the data. Therefore, in order to obtain confidence intervals that do not rely on the normality assumption, we bootstrapped the confidence intervals of the mean. Interestingly, we observe that all leathers under examination present a signature (values for the means and confidence intervals) that resembles one another,

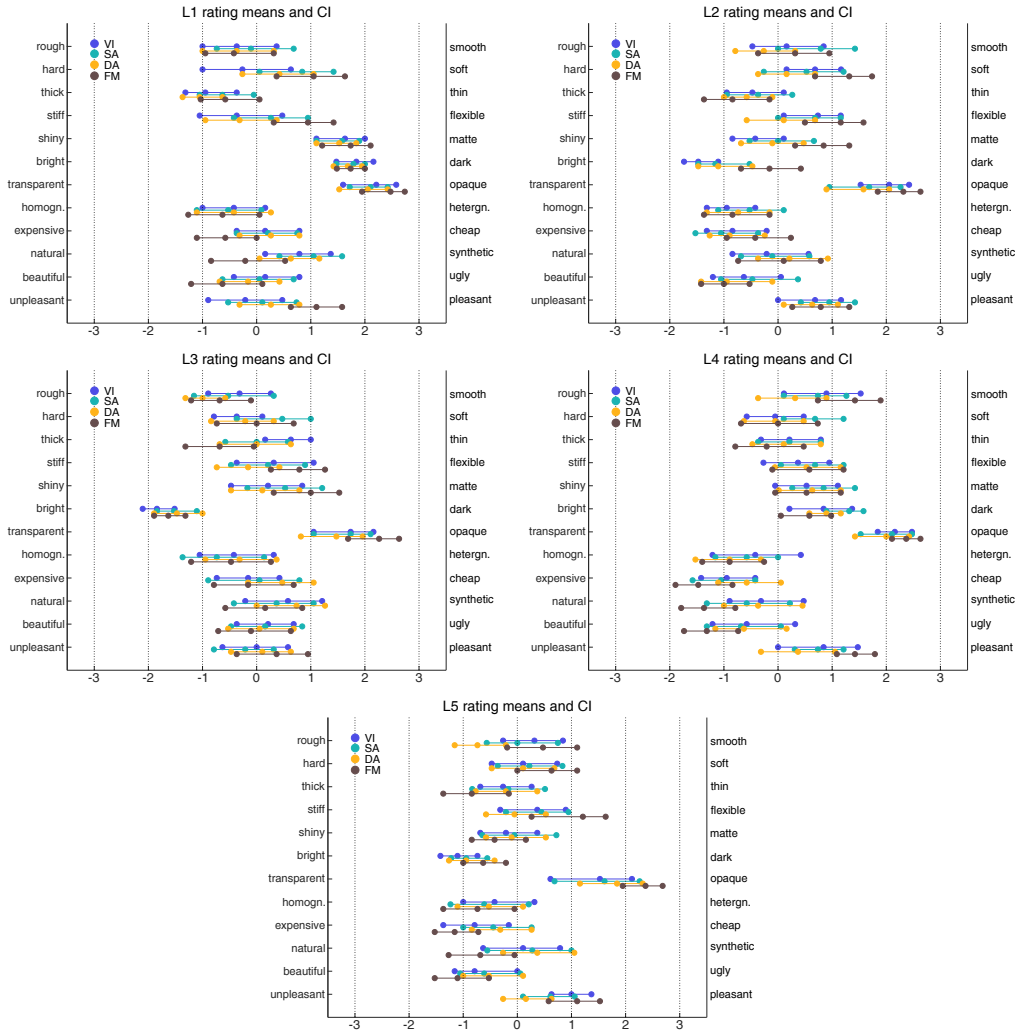


Figure 4.7: *Response ratings for each of the investigated material samples (leathers only). The central circles represent the participants' mean opinion scores, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

being the most evident variations located in the rough–hard and shiny–bright dimensions while the signatures for fabric materials appear to be more distinct. The latter suggests that, despite their visual differences, the rubbing sounds generated by leathers are not particularly diverse across samples according to human perception, as it is the case for fabric materials.

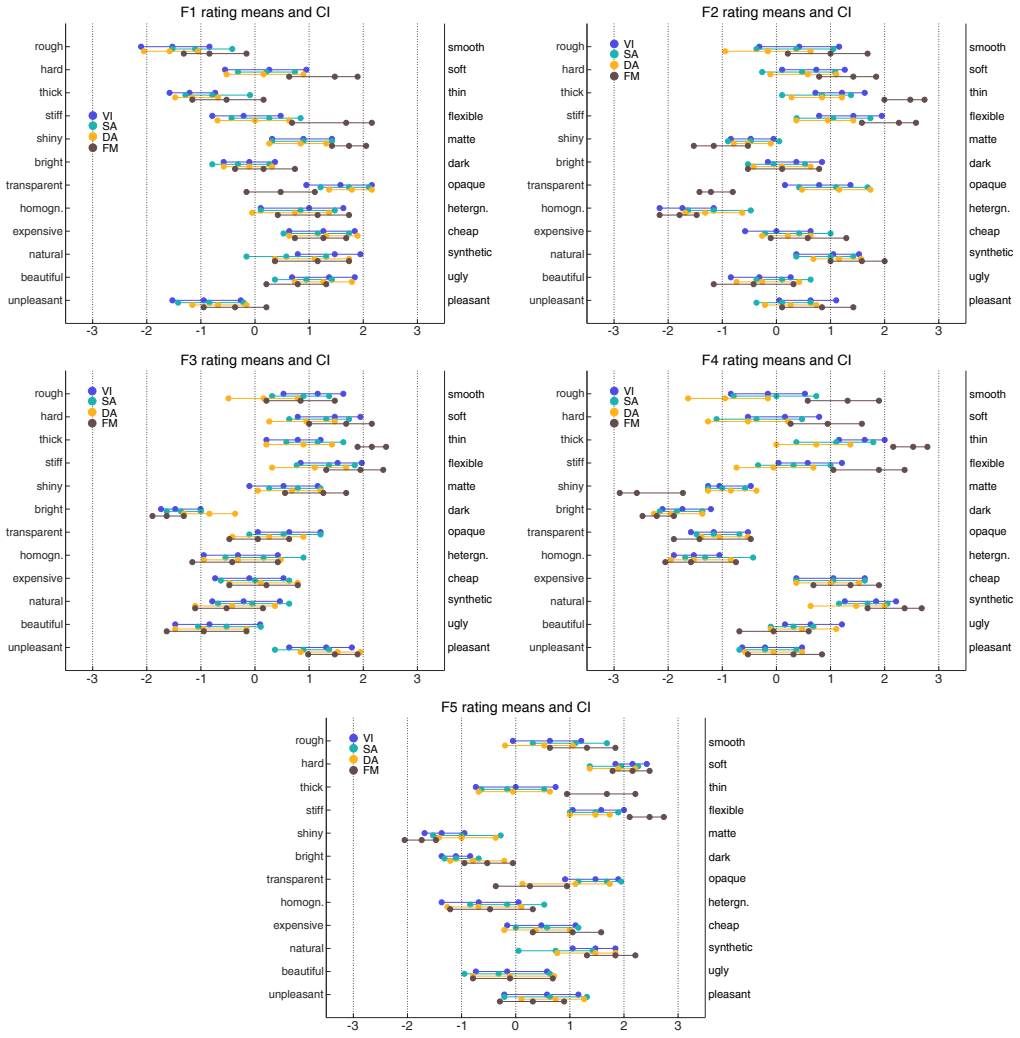


Figure 4.8: *Response ratings for each of the investigated material samples (fabrics only). The central circles represent the participants' mean opinion scores, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

CHAPTER 5

Digital Transmission of Subjective Material Appearance

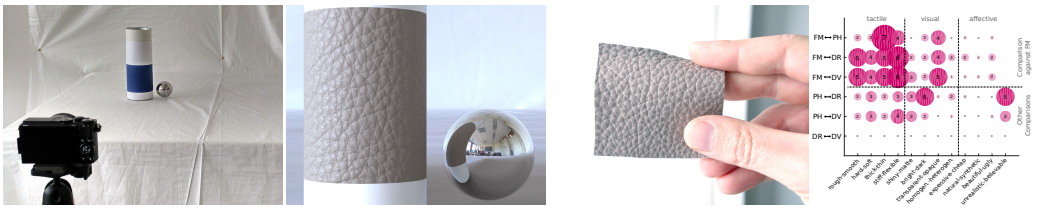


Figure 5.1: *(Teaser) From left to right, an illustration of the photo setup, one of the digitized material utilized in our its experiments, its real-world counterpart and the results of applying non-parametric tests between experimental conditions.*

Abstract. The digital recreation of real-world materials has a substantial role in applications such as product design, on-line shopping or video games. Since decisions in design or shopping are often driven by qualities like "softness" or "beautifulness" of a material (rather than its photo-accurate visual depiction), a digital material should not only closely capture the texture and reflectance of the physical sample, but also its subjective feel. Computer graphics research constantly struggles to trade physical accuracy against computational efficiency. However, the connection between measurable properties of a material and its perceived quality is subtle and hard to quantify. Here, we analyze the capability of a state-of-the-art model for digital material appearance (the spatially-varying BRDF) to transport certain subjective qualities through the visual channel. In a psychophysical study, we presented users with measured

material SVBRDFs in the form of rendered still images and animations, as well as photographs and physical samples of the original materials. The main insight from this experiment is that photographs reproduce better those qualities associated with the sense of touch, particularly for textile materials. We hypothesized that the abstraction of volumetric materials as opaque and flat textures destroys important visual cues especially in border regions, where fluff and protruding fibers are most prominent. We therefore performed a follow-up experiment where the border regions have been removed from the photographs. The fact that this step greatly reduced the capability of photos to transport important qualities suggests strong directions of future research in applied perception and computer graphics.

In the former chapters we investigated the multisensory perception of material qualities, using as visual stimuli pictures from real samples. The main reason for the election of such stimuli is that the capability of digital material models in representing detailed real material characteristics is yet to be investigated. In this regard, we considered the Spatially-Varying BRDF model as representative of the state-of-the-art in computer graphics, and employed it to depict a collection of materials (leathers and fabrics). We then compared the ability of such digitized samples to transmit subjective material appearance information (represented by a set of material qualities) against simple photographs and real materials. Thereby, we arranged an experimental study inspired by those from the previous chapters and analyzed the resulting outcome using non-parametric tests.

This chapter corresponds to the paper [MWH17]: Rodrigo Martin, Michael Weinmann and Matthias B. Hullin. “Digital Transmission of Subjective Material Appearance”. In *Journal of WSCG*. Jun 2017.

5.1 Introduction

The recent progress in the photo-realistic depiction of digitized materials has led to a paradigm change in important applications where the conventional approach of communicating objects in terms of photos taken by experts is more and more replaced by virtual surrogates. This methodology allows new possibilities such as cooperative product design, product advertising in prototype phase, exhibition of furniture or wearables in specific environments or visualization of cultural heritage objects. The entertainment industry has also drawn a major benefit from advanced digital material models as they allow a more realistic experience of virtual scenarios. While a remarkable reproduction quality has been achieved for virtual/digitized materials, there is still a gap in appearance between them and their physical counterparts which, in application, may distort the perception of the product. In particular, the accurate reproduction of surface reflectance behavior under varying illuminations and viewing conditions still remains a challenging task. For this reason, many material catalogs [HJ10] still opt for using pictures or even physical material samples to illustrate their collections instead of digitized models, despite the potential benefits that they entail.

In this paper, we aim at investigating this breach in appearance between digitized materials and their physical counterparts by analyzing how perceptual material information is transmitted through different stimuli. For this purpose, we consider the perception of materials by assessing a set of subjective qualities that can be assigned to either the tactile, visual or affective category, depending on the nature of the interaction that best reveals them. We then conducted a psychophysical study to compare the communication of these attributes based on different representations given by real material samples, photographs of these samples as well as static and animated renderings of the digitized materials represented by the spatially-varying BRDFs (SVBRDF) model [NRH+77], which is deemed to be a standard representation in research and industry [EL12]). The materials evaluated belonged to two semantic categories, leathers and fabrics. A key observation obtained from this experiment is that the SVBRDF model is not capable of preserving important qualities of material appearance, especially the tactile ones. Even a dynamic change of viewpoint does not seem to improve the perception of materials. Thus, the loss of information is presumably not caused by the limited resolution of the digitized samples but due to the abstractions intrinsic to the model. Upon closer inspection of the stimuli we observed that

the differences between photos and virtual materials are most prominent at grazing angles, where the SVBRDF model fails to capture the volumetric material structure, intricate light scattering effects and the partial transparency of protruding fibers. Consequently, the perception of material properties such as softness, stiffness or transparency is not accurately recreated in the digitized representation, deviating from the correspondent photos and physical samples. With that in mind, we designed a follow-up study in which the respective border regions were digitally removed from the photos for a subset of relevant materials. Indeed, this step led to a significant deterioration in the transmission of tactile and affective properties, confirming our initial suppositions.

Our main findings are:

- Digitized materials (SVBRDF model) are not capable of adequately transmitting certain perceptual material information, being outperformed by simple photos of material samples. However, there is also an gap between photos and real materials.
- There are no significant differences in material-quality perception between static and animated digitized representations.
- The depictions of digitized materials suffer from a significant loss of information at grazing angles, where the SVBRDF model cannot represent appearance accurately.

To the best of our knowledge, this is the first perceptually-motivated work in evaluating how subjective material appearance is transmitted through digitized models (SVBRDF) in comparison to photos and real material samples. Conclusions from this set of studies are restricted to the given stimuli, but can provide useful insights for future research in developing realistic material representations.

5.2 Related work

In this section, we provide a condensed synopsis of research on the perception of subjective material qualities and the evaluation of digital material appearance models for graphics applications.

5.2.1 Perception of materials and their qualities.

The interest in unveiling the principles and reasons that determine the visual appearance of materials as well as in how humans visually perceive materials and their properties has received an increasing attention over the last years. Respective surveys [Ade01, And11] provide a discussion

of the main problems and challenges in this area of research including the perception of material surface and properties. A further examination of the challenges in material perception is provided by [Fle14], where the author outlines a new theory of material perception based on 'statistical appearance models'. Among the phenomena that contribute to material appearance, glossiness has received a considerable amount of attention. Several approaches aimed at finding perceptually meaningful reparameterizations of material gloss by exploring the relationships between physical parameters and the perceptual dimensions of glossy appearance [PFG00, WAKB09]. The human capability of perceiving material gloss (gloss constancy) under varying motion, disparity and color conditions was investigated by Wendt et al. [WFEM10]. In addition to gloss, Ho et al. [HLM06] researched the visual estimation of surface roughness, discovering that its perception is strongly influenced by the illuminant angle.

Motion is another aspect that has an important impact on the appearance of materials' surface. By analyzing the optical flow, Doerschner et al. [DFY+11] identified three motion cues, in which the brain could rely in order to identify material shininess. Our investigation further evaluates which additional subjective information (if any) is revealed by motion when compared to still renderings of digital materials. Other than motion, shape and geometry have proven to be critical aspects in the perception of materials [VLD07]. The importance of the shape for material categorization is well-known [Ade01] and also can be used as an additional cue for material recognition [DGFH16]. In this regard, one of the conclusions of our research is the emphasis on geometry and appearance under grazing angles, as a decisive feature to accurately assess material qualities. Indeed, the tasks of material categorization and material property judgment are closely related as demonstrated by Fleming et al. [FWG13]. Their studies revealed a high degree of consistency between these two assignments, implying that humans access similar information about materials when performing both tasks.

Although the experimental procedure initially involves purely visual stimuli, the participants also rated the same attributes for the real material samples in a sort of interaction that makes use of all senses (multimodal or full-modal interaction). The described approach relate to previous studies in multimodal material perception [FTK15, MIWH15], which highlight the importance of the tactile and auditory channels in the perception of material information. Their work also relies on ratings not only for surface material properties, but also a set of affective attributes.

5.2.2 Perceptual evaluation of material appearance models.

Material appearance acquisition and modeling have been deeply researched [HF13, WLGK16]. Widely used digital representations of materials exhibiting a spatially varying appearance include Spatially-Varying BRDFs (SVBRDFs) [NRH+77] and Bidirectional Texture Functions (BTFs) [DvG+99]. Both representations model material appearance depending on the spatial position $\mathbf{x}$ on the object surface, the light direction ω_i and the view direction ω_o. SVBRDFs allow a more compact modeling of surface reflectance behavior than BTFs at the cost of neglecting effects of light exchange at subtle surface structures. As SVBRDFs have become a standard in industry [EL12], we use this representation to analyze differences in the human perception of real and digitized materials.

Regarding the perceptual evaluation of appearance models, several investigations focused on analyzing the level of realism achieved by a concrete model. In this context, Meseth et al. [MMK+06] verified the ability of BTF models to achieve photo-realism in comparison to standard representations (BRDFs) and photographs, at a coarse and fine scale. It was demonstrated that BTF materials entail a significant increase of realism over BRDFs at both scales, albeit being still inferior to the scene photographs. A study from Filip et al. [FVHK16] determined and predicted the critical viewing distances at which a certain BTF can be replaced by the correspondent BRDF representation without decreasing the overall visual impression. Additionally, Jarabo et al. [JWD+14] examined the effects of approximate filtering on the appearance of BTFs in different domains (spatial, angular and temporal). The authors identified interesting correlations between high-level descriptors and perceptually equivalent levels of filtering as well as with low-level BTF statistics.

5.3 Experiment 1: methods

The proposed experiment investigates the performance of a well-known appearance model when transmitting subjective material qualities in comparison to equivalent photographs from real materials. In addition, the exercise examines whether the consideration of a higher spatial resolution through motion in digital scenes provides additional cues in the aforementioned task. Throughout this section the stimuli acquisition, selection of material qualities and experimental procedure will be detailed.

5.3.1 Stimuli

Selection of materials In the scope of this research, we explore the perception of physical and affective material qualities for two semantic classes (leathers and fabrics). Restricting our selection to these two concrete, well-known categories allows us to keep the study and its conclusions manageable. Next, we have chosen ten material samples pertaining to these classes, each of them with an approximate size of 120×120 mm^2, with nearly flat geometry to match the requirements of the acquisition device, which is described in the following paragraphs. With this fine selection, we intended to maximize the relative intra-class heterogeneity not only in terms of the physical properties but also the aesthetic characteristics.

Photographs of materials. In order to make the real and the virtual materials as comparable as possible, both the real and the virtual scene should share comparable geometry and illumination conditions. With that intention, our real scene was composed by a cardboard cylinder (80 mm diameter) to which the sample was attached. Cylindrical geometries have been frequently used in previous perceptual studies [FVHK16] because of its well-defined texture mapping and for being one of the most discriminating shapes [VLD07]. We covered the uppermost and lower part of the sample with white pieces of cardboard, which gently fixed the material to the cylinder. The height of the visible part of the sample along the vertical axis was approximately 90 mm. A reflecting sphere with a diameter of 50 mm was situated 10 mm right from the cylinder, and the whole setup was placed under natural illumination using a white, uniform piece of cloth as background. This arrangement is not arbitrary, given that during our internal tests we learned that subjects are more adept at this kind of subjective exercises when some context regarding the scene is provided. The complete setup can be observed in Figure 5.2, where the digital camera (Nikon 1 J5, resolution of 5568×3712 pixels) was situated at a distance of 280 mm in front of the material sample. We took a picture for each specimen while keeping the light and viewing conditions constant. The images were then corrected regarding white-balance, cropped and scaled to match the resolution of the final device (see Section 5.3.3). Moreover, during the photo session we used a remote-controlled 360° spherical panoramic camera (Ricoh Theta S) to probe the scene illumination. The resulting high-dynamic-range environment map was utilized to illuminate our virtual scenes.

Figure 5.2: *View of the photo setup*

Digitized materials. The digitization of the material samples was carried out using a commercial scanning device [XR16] that allows the measurement of (flat) material samples. After taking images of the material sample from different viewpoints and under different illumination conditions, a surface normal map is obtained and the reflectance behavior is stored in terms of a Ward-SVBRDF. We refer to the supplementary material for more details on the material digitization process. The output format (AxF) is supported natively by several rendering applications such as Autodesk VRED, which was employed to generate the renderings used in this study. We approximated the geometry of the described photographic setup in a virtual scene and used VRED's Full Global Illumination algorithm to render it, lighting the scene with the previously calculated environment map. For the animated scene, we rotated the camera 60° back and forth around the cylinder in the Y-axis, and rendered the scene at 60 frames per second to get a clip with a duration of 4 seconds. The resulting photos and renderings are shown in Figure 5.3.

Real materials. During the course of the experiment, we handed samples from the actual materials to the participants, hence, allowing a full-modal experience of the individual material qualities. Instead of the samples that were used for the acquisition, we used smaller portions of the same sample (approximately 70×70 mm^2) to avoid damaging the originals due to the interaction and in favor of the scalability of the process.

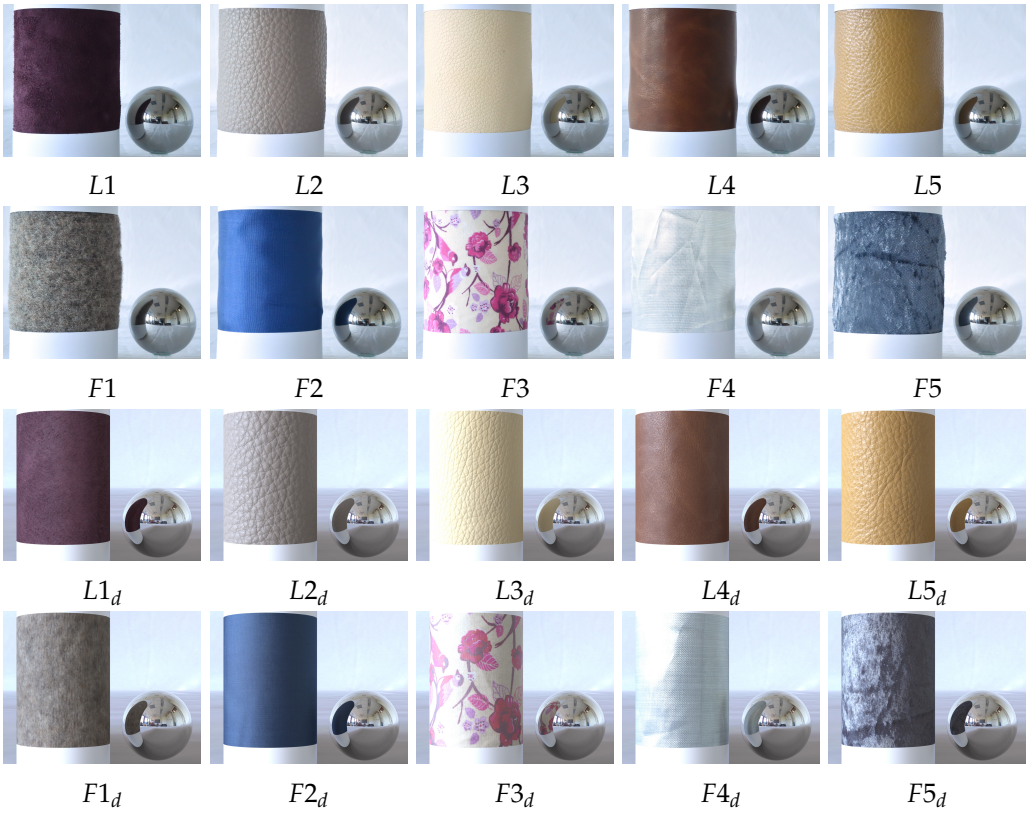


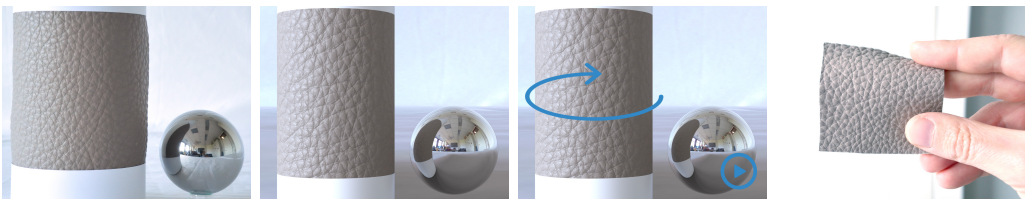
Figure 5.3: *In the two upper rows, the pictures from the samples utilized in the study. In the two lower rows, the correspondent digitized material renderings.*

Tactile	**Visual**	**Affective**
rough–smooth	shiny–matte	expensive–cheap
hard–soft	bright–dark	natural–synthetic
thick–thin	transparent–opaque	beautiful–ugly
stiff–flexible	homogeneous–heterogeneous	unrealistic–believable

Table 5.1: *Opposite-meaning quality pairs.*

5.3.2 Selection of material qualities

In an initial step, we focused on finding a meaningful subspace of subjective adjectives that characterizes our selection of materials. The importance of this task was first addressed by Rao and Lohse [RL96] for the concrete case of visual textures. We collected a list of 42 subjective material qualities organized in 21 opposite-meaning adjectives, which were



PH *DR* *DV* *FM*

Figure 5.4: *Stimuli presented in the psychophysical experiment corresponding to the four different conditions for an example material L2. From left to right, photograph (PH), digitized render (DR), digitized video (DV) and the physical sample (FM).*

observed to be the most recurrent ones in related literature regarding material perception [FWG13, FTK15, JWD+14, MIWH15]. Such qualities were conceptually separated in three different groups with respect to their tactile, visual or affective nature. In pursuance of getting a smaller subspace of qualities that maximizes the transported information about our particular material collection, we conducted a pilot experiment in which we handed out the 10 original material samples to 7 participants along with a list of the 42 individual adjectives. The subjects were asked to mark the adjectives that better describe each sample. There was no restriction regarding the number of adjectives to choose. From the results, we selected the most voted attribute pairs in each of the three groups for our experiments, leading to a final assortment of 11 adjectives (see Table 5.1). Although it was not in our original list, we additionally included the pair 'unrealistic-believable', which provides information about the level of realism portrayed by the virtual materials.

5.3.3 Experimental procedure

The user study was conducted using tablet computers (Toshiba Excite Pro 10.1, resolution of 2560×1600 pixels) running a custom Android application. This experimental setup makes our study scalable to larger surveys in addition to representative of contemporary consumer hardware. The procedure was carried out in a quiet, well-illuminated room and organized in sessions with a maximum of 7 participants. An introductory presentation was provided before performing the exercise to explain the procedure and clarify inquiries. Participants were instructed to infer the qualities which were not evidently revealed in a particular representation.

Different techniques were considered to perform perceptual quality

ratings across our stimuli. Although double stimulus ratings or forced-choice pairwise comparisons may lead to smallest measurement variance, they would also increase the number of trials and, thus, make the whole study more difficult to accomplish. Therefore we decided to employ single stimulus ratings in which, for each stimulus, the subjects had to rate the selected qualities on a 7-point Likert scale characterized by a slider with values ranging from -3 to 3 (see supplementary material). Each of the values in the slider was consistently labeled with a term indicating the intensity of the stimuli in both axes (e.g., very bright, bright, a bit bright, neutral, a bit dark, dark, very dark). The actual procedure consisted of four different presentations or conditions, in which different material images were presented to the participants in randomized order along with the rating questionnaire. In addition, the participants had the chance to examine the real samples, serving the respective ratings as ground truth. The conditions that compose the experiment are illustrated in Figure 5.4 and listed below:

- *Photographs (PH)* taken from the real materials.
- *Digitized static renderings (DR)* from materials using the SVBRDF reflectance model.
- *Digitized video renderings (DV)* using the same reflectance model, where the camera rotates around the sample in the Y-axis.
- *Full-modal condition (FM)*. Physical material samples were given to the participants so that they could interact with them.

As the interaction with the real samples may bias the rest of the task, the condition FM was constrained to be the final one, while the order of the remaining conditions was randomized. Additionally, the application was instrumented to identify incorrect realizations of the assignment (e.g. skipping a material), in order to make the data more reliable. A total of 20 subjects (13 females, mean age 27.69; 7 males, mean age 27.00) participated voluntarily in the experiment. They were all naïve to the purpose of the experiment and reported normal or corrected-to normal visual acuity. They also provided informed consent and were compensated economically for their participation. From all the combinations of the conditions, materials, qualities and subjects, we obtained $4 \times 10 \times 12 \times 20 = 9600$ rating responses that are analyzed in the next section.

5.4 Experiment 1: results

To evaluate how the subjective attributes were perceived in the aforementioned material presentations we performed a conjoint analysis of the participants' preferences and non-parametric tests. The participants' rating responses were deemed reliable (Cronbach's $\alpha = .93$). In addition, participants' mean ratings and confidence intervals for each material and quality are included in the supplementary material.

5.4.1 Conjoint analysis

With the purpose of gaining a general understanding regarding how material perception differs between the individual conditions, we made use of conjoint analysis techniques [GS90] on the subjects' ratings. This method has been used extensively in market research to measure the preferences of the customers among multi-attributed products and services. In our experiment, we analyzed the three visual conditions C_i with $i \in \{\mathrm{PH}, \mathrm{DR}, \mathrm{DV}\}$ in the conjoint analysis and considered the following question: 'To what degree does condition C_i transmit the quality $q_k \in Q$ in comparison to the other conditions?'. Due to our experimental procedure based on single ratings, we cannot directly compare two conditions. Instead, we can evaluate them with respect to the full-modal representation (FM). From this, we can infer that a certain condition C_i is more suitable to represent an individual property than another C_j if the participants' ratings better agree to the ones obtained for the full-modal condition (FM). In contrast, if the ratings are distant, the depiction is less realistic and, consequently, less suitable.

In order to carry out this comparisons, we use the weighed voting schema described in Martín et al. [MIWH15] to compute the 'utility scores' (or 'part-worth utilities') s_i. For a certain combination of material, quality and subject, r_i and r_j denote the ratings for two particular conditions (C_i and C_j with $i, j \in \{\mathrm{PH}, \mathrm{DR}, \mathrm{DV}\}$) and r_{FM} denotes the ratings for the full-modal task which serves as ground truth. The calculated intermediate utility scores ($s_{i,j}$) are defined according to

$$s_{i,j} = \begin{cases} |r_{\mathrm{FM}} - r_i| - |r_{\mathrm{FM}} - r_j| & \text{if } |r_{\mathrm{FM}} - r_i| > |r_{\mathrm{FM}} - r_j| \\ 0 & \text{else} \end{cases}. \tag{5.1}$$

To compute the final utility scores s_i, and the normalized 'importance scores' $T = (t_i)$ for each condition, we consider the matrix composed of

the calculated intermediate scores $S = (s_{i,j})$ with $S \in \mathbb{N}^{N \times N}$ and $s_{i,i} = 0$. Note that, in general, the matrix S is not symmetric. Then T is given by

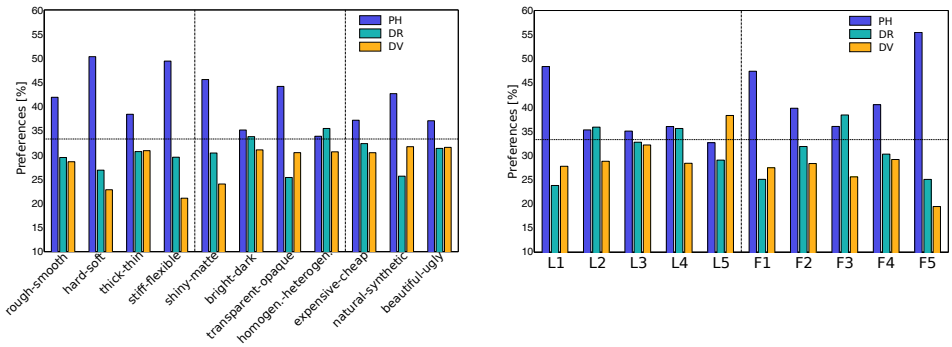
$$T = \frac{\sum\limits_{i} s_{i,j}}{\sum\limits_{i,j} s_{i,j}} \quad \text{where} \quad i, j \in \{\text{PH}, \text{DR}, \text{DV}\}. \tag{5.2}$$

The resulting scores, separated by property and material, can be seen in Figure 5.5. A clear evidence regarding the preference of a certain condition with respect to the other ones would imply that the respective condition depicts the reality more accurately, for the corresponding material or quality. Indeed, the obtained results indicate a clear predilection towards PH for almost all qualities and materials. This preference is particularly noticeable for the tactile adjective pairs (e.g. 'hard-soft', 'stiff-flexible') but can also be observed for visual properties (e.g. 'shiny-matte', 'transparent-opaque') and affective properties (e.g. 'natural-synthetic'). Considering the preferences organized per material, the condition PH is especially favored for fabric specimens. In fact, if applying conjoint analysis between the two material classes, the scores obtained for digitized leathers ($t_{\text{PH}} = 38.09\%$, $t_{\text{DR}} = 31.15\%$ and $t_{\text{DV}} = 30.76\%$) are higher than the ones for digitized fabrics ($t_{\text{PH}} = 45.26\%$, $t_{\text{DR}} = 29.23\%$ and $t_{\text{DV}} = 25.51\%$). Another interesting finding shows up when comparing the importance scores among static and dynamic renderings (DR and DV). Initially, the video presentation only performs better when transmitting transparency and naturalness. Applying conjoint analysis between DR and DV exclusively led to a more balanced overall preference of $t_{\text{DR}} = 52.60\%$ and $t_{\text{DV}} = 47.40\%$. The pair 'unrealistic-believable' does not apply to the real material stimuli and hence, it was not considered during the analysis.

This way, conjoint analysis provides insights regarding how well individual qualities are transmitted by the different conditions and, hence, which of the corresponding representations is most suitable. In the next section, we intend to additionally discover if and where significant differences among the ratings of the conditions are manifested.

5.4.2 Non-parametric tests

In addition to compare each condition against the ground truth (FM), we would also like to detect whether, and if so also where, meaningful discrepancies between the individual conditions occur. This may help us to understand how differently these representations transmit material



(a) *Preferences arranged by quality* ***(b)*** *Preferences arranged by material*

Figure 5.5: *Summary of the conjoint analysis revealing participants' preferences for each condition according to our voting schema. The preferences are separated by quality (left figure) and by material (right figure). The PH condition is preferred for almost every quality and performs particularly better for fabric materials.*

qualities. A preliminary Shapiro-Wilk normality test determined that, for certain combinations of material and quality, our data do not come from a normally distributed population. This fact, together with the ordinal nature of the Likert scales, discredit analyses based on group means. Thus, we applied non-parametric tests (Friedman and Wilcoxon) in order to detect significant differences between the ratings (dependent variable) of the four conditions (independent variable).

Given our experimental design, we will be able to draw valid conclusions only for a single material-quality pair at a time ($p_\mathrm{i} = \{m_\mathrm{j}, q_\mathrm{k}\}$, given a material $m_\mathrm{j} \in M$ and quality $q_\mathrm{k} \in Q$), across all subjects. For better understanding, we first consider the pair p_i given by the combination of material $L1$ and the quality 'rough-smooth'. Applying Friedman's test revealed that the effect of the different conditions on the subjects' judgments is significant ($\chi^2(3) = 19.86, p < .05, r = .00$). The post-hoc analysis with the Wilcoxon signed-rank procedure resulted into rejecting the null hypotheses for the comparisons FM $\leftrightarrow$ DR, FM $\leftrightarrow$ DV and DV $\leftrightarrow$ PH, i.e. these representations have a significantly different effect on the participants' ratings. In contrast, the comparisons FM $\leftrightarrow$ PH, DV $\leftrightarrow$ DR and DR $\leftrightarrow$ PH showed no interaction effect on the ratings. In order to extend our findings to the complete collection of M materials and Q qualities, we performed the same analysis for each possible combination of material and quality $p_i \in P$. Then, we summed up the number of occurrences in which, for a particular p_i, we rejected the null hypothe-

ses and, therefore, the ratings among conditions were determined to be significantly different (at least $p < .05$ for all the cases). We refer to this sum hereafter as the "dissimilarity score". Here, the presence of a high dissimilarity score between FM and another condition would outline how good or bad the respective depiction transmits real world information. In addition, by means of the same evidence for the rest of the scores, we may learn how differently photographs and digitized materials illustrate the individual qualities and if there is any significant impact in the ratings coming from motion. The outcome for all ten materials is separated by quality and shown in Figure 5.6.

As can be observed, the largest scores are mainly concentrated when the conditions FM $\leftrightarrow$ DR and FM $\leftrightarrow$ DV are compared, and this is especially appreciable for qualities categorized as tactile (upper-left quadrant). Besides, we can observe high scores between FM and the rest of the conditions for the adjective pairs 'thick-thin', 'stiff-flexible' and 'transparent-opaque'. This fact indicates that none of our representations is able to fully communicate these concrete qualities. Furthermore, the small scores found between the conditions FM $\leftrightarrow$ PH in the remaining adjective pairs suggests that photos transmit most of the qualities good enough. In general, these results correlate well with the findings from the conjoint analysis, as they also tend to indicate the predominance of photographs over our digitized materials, especially in the tactile domain. In fact, the differences in the perceived realism ('unrealistic-believable' dimension) between PH and virtual materials confirm this trend. Finally, no significant dissimilarities were discovered in the comparison DR $\leftrightarrow$ DV, i.e. the overall perception of material qualities is not affected by motion.

5.5 Experiment 2

During the course of the previous experiment, we observed that the samples with padded and fluffy appearance do not transmit appropriately material appearance in the digitized conditions and, hence, were deemed to be more unrealistic (see supplementary material). These features are more salient in the distinctive border regions, which possibly behaved as one of the main sources of information in favor of photographs. Due to the limited resolution of the reconstructed surface geometry, these structures are not accurately captured and the SVBRDF model is not capable of reproducing surface effects like self-occlusions, interreflections or transparency. To better understand how this matter influences the transmission of material appearance and which subjective attributes are most

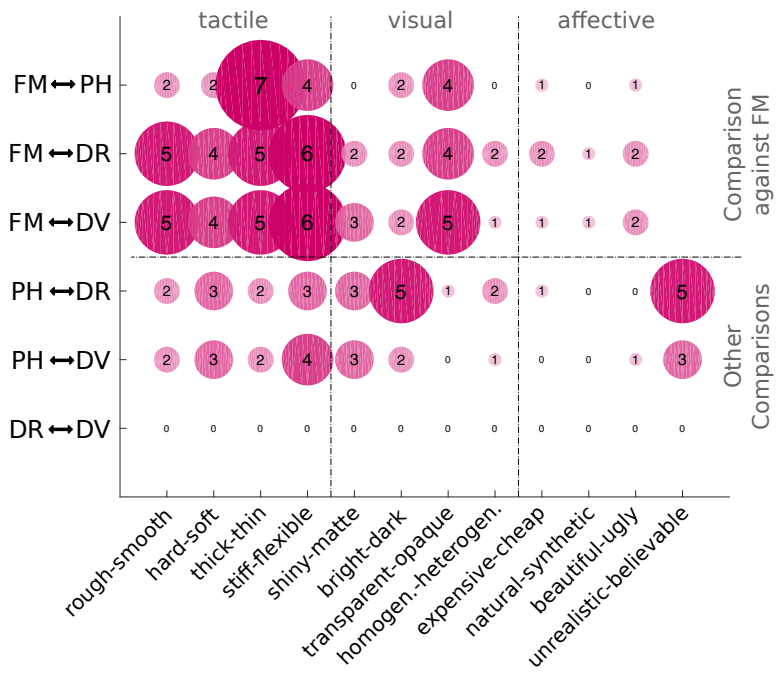


Figure 5.6: *Number of tests with a significant effect on subjects' ratings (at least $p < .05$), summed along all ten materials. Each row compares two conditions while each column represents a quality pair. The largest scores are located in the second row (FM ↔ DR) and third row (FM ↔ DV), especially for tactile qualities. In contrast, there are no significant effects in the lower row (DR ↔ DV).*

affected, we designed a follow-up study in which the perception of digitized materials was compared to the perception of real materials within photos where the border features have been digitally removed. The description of the experimental procedure and results are provided in the following sections. The supplementary material additionally provides the mean response ratings and confidence intervals for each material and quality.

5.5.1 Methods

From the materials selected for the previous experiment, we chose a subset of samples whose digitized stimuli were perceived to be particularly different from their correspondent photos in the experimental analysis, failing to transmit many of the considered attributes. According to this, we selected the set $M_2 = \{L1, L5, F1, F5\}$, where material $L5$ was only

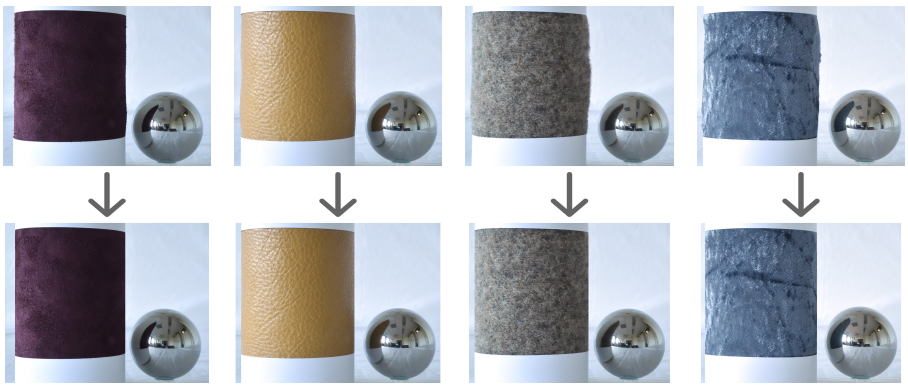


Figure 5.7: *Pictures from the two leathers (L1, L5) and two fabrics (F1, F5) selected for the follow-up experiment, before and after crop operation.*

included to have an equal number of leathers and fabrics in the scope of this study. From the original photographs we removed the visible material borders from the cylindrical geometry to which the sample was attached using Adobe Photoshop, resulting into a flat silhouette shape as shown in Figure 5.7. Accordingly, we rendered again the digitized materials to match the new resolution from the cropped photographs. Other than that, we also aimed at comparing our visual stimuli against the real materials and we considered the same assortment of perceptual qualities as in the previous experiment. Nevertheless, in this experiment no motion was included, i.e. the considered conditions are:

- Cropped photographs (PH_c) taken from the real materials, where the borders have been removed.
- Digitized static renderings (DR) from materials using the SVBRDF reflectance model.
- Full-modal condition (FM). Physical material samples were given to the participants so that they could interact with them.

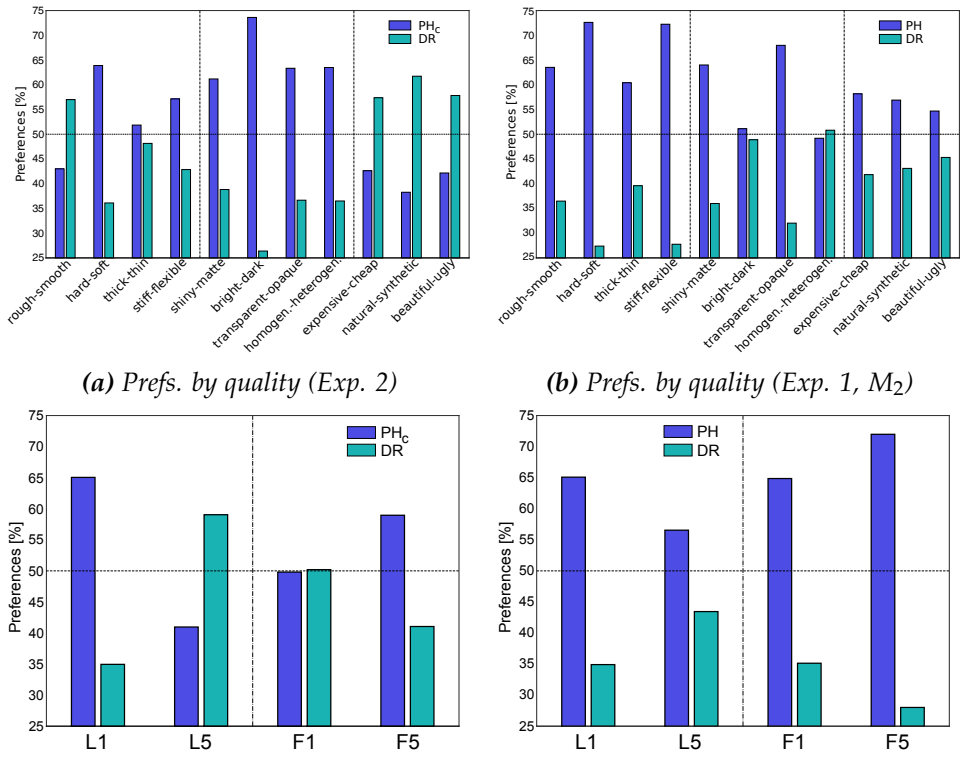
Again, the order of the materials and conditions was randomized except for FM, which was constrained to be the last one. 19 subjects (12 females mean age 27.08; 7 males, mean age 28.57) took part in the experiment under the same conditions as the previous one. The resulting $3 \times 4 \times 12 \times 19 = 2736$ rating responses are evaluated in the next section.

5.5.2 Results

In the following, we show the outcome of performing conjoint analysis and non-parametric tests on the subjects' ratings and compare them w.r.t. the results of the previous experiment. A Cronbach's alpha value of $\alpha = .89$ confirms the reliability of the ratings.

Conjoint analysis Similar to Section 5.4.1 we performed a conjoint analysis in order to reply the question: 'To what degree do the conditions PH_c and DR transmit the quality $q_k \in Q$ in comparison to FM?'. Figure 5.8a illustrates the importance scores per quality for the current experiment, in which the borders were removed, while Figure 5.8b shows the scores obtained for Experiment 1, if only the data from the subset M_2 of materials were taken into account. Direct comparison between the scores corresponding to the conditions PH_c (Experiment 2) and PH (Experiment 1) reveals how the preferences for the cropped photographs become significantly smaller for all the tactile and affective attributes so that, for certain cases, these are surpassed by the DR scores. Certainly, the score difference between conditions PH and PH_c in both experiments should be a good indicator regarding which perceptual attributes were most damaged with the border-feature removal. According to this, the most deteriorated pair was 'rough-smooth' (-20.52%), followed by 'natural-synthetic' (-18.63%), 'expensive-cheap' (-15.55%) and 'stiff-flexible' (-15.10%). Contrarily, the pair 'bright-dark' ($+22.50\%$) and also 'homogeneous-heterogeneous' ($+14.30\%$), were surprisingly better communicated without the borders. In this case, the silhouette information present in the photos from Experiment 1 could have acted as a misleading cue to judge homogeneity and brightness. Finally, the importance scores separated by material are shown in Figure 5.8c. When compared to the scores obtained in Experiment 1 (Figure 5.8d), we notice substantial changes as the preferences for PH_c diminish in favor of the ones for the DR condition except for the material $L1$, whose scores remain relatively constant.

Non-parametric tests As in our previous study, we perform non-parametric tests (Friedman and Wilcoxon) to detect meaningful differences among the respective ratings (dependent variable) for the three conditions PH_c, DR and FM (independent variable). Anew, we carried out multiple comparison tests between the conditions (applying Wilcoxon signed-rank procedure) and generalized our findings by summing the resulting occurrences of rejected null hypotheses for each material-quality



(a) *Prefs. by quality (Exp. 2)*

(b) *Prefs. by quality (Exp. 1, M_2)*

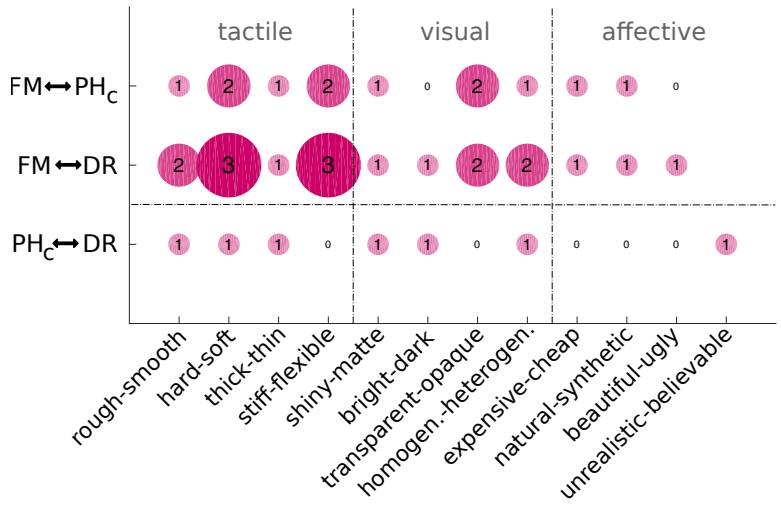
(c) *Prefs. by material (Exp. 2)*

(d) *Prefs. by material (Exp. 1, M_2)*

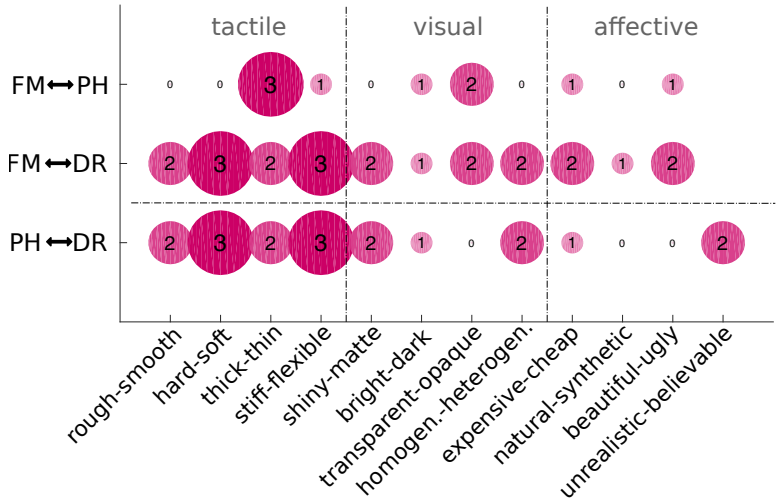
Figure 5.8: *Summary of the conjoint analysis showing the participants' preferences for each condition in Experiment 2, where the material borders were removed from the photos, in contrast to Experiment 1, separated by quality (left figures) and material (right figures). The lower scores for the condition PH_c in comparison to PH show how the transmission of tactile and affective qualities as well as fabric samples deteriorates when the borders are removed.*

pair $p_i \in P$. The resulting dissimilarity scores are displayed in Figure 5.9a together with the scores resulting when applying the same test in Experiment 1 (Figure 5.9b), for the material subset M_2. the DV condition was ignored as the video stimuli were not used in the follow-up study.

From the results depicted in the figures, we can outline three main observations. First, the large dissimilarity scores found in the bottom comparison PH $\leftrightarrow$ DR for Experiment 1 have disappeared when moving to PH_c $\leftrightarrow$ DR in Experiment 2, which suggests that both representations lie much closer in the follow-up study. Second, the middle row comparing FM $\leftrightarrow$ DR only contains subtle changes in the scores obtained for both experiments. This fact is coherent with the stimuli as these conditions have not changed between experiments. Third, the top row comparing



(a) *Dissimilarity score (Exp. 2)*



(b) *Dissimilarity score (Exp. 1)*

Figure 5.9: *N° of tests with a significant effect on subjects' ratings (at least $p < .05$), summed along the subset M_2 of four materials. On the left, the dissimilarity scores for Experiment 2, where the borders were removed from the photos. On the right, the respective scores for Experiment 1. Note the high scores in the comparison PH ↔ DR for Experiment 1, whereas they partially move to the first row FM ↔ PH_c in Experiment 2. Meanwhile the middle row presents little variation.*

FM $\leftrightarrow$ PH_c in Experiment 2 presents, for most of the considered qualities, higher scores as in the original study. This fact suggests that the perception of photographs and real materials differs more significantly when the silhouette-border information is not present. However, the pair 'thick-thin' displays an unexpected opposite tendency. Again, borders may have acted as a misleading cue to judge thickness on these concrete samples.

5.6 Conclusions

In the scope of this investigation, we have studied the perceptual differences between stimuli based on standard digital material appearance models in terms of Spatially Varying BRDFs, photos of real materials (leathers and fabrics) and the actual material samples on the task of transmitting a rigorously selected group of subjective qualities. Additionally, we explored the effect of motion on the perception of the stimuli based on digitized material representations. Because of the observation that the appearance of photographed materials and their digitized counterparts differ particularly at the material borders, a second experiment was designed to explore to what degree the appearance of materials under flat viewing angles could cause the loss of information between photographs and renderings.

One of the main findings of our investigations is that the considered digitized models are not able to fully transmit basic subjective properties according to the reality. Most of the analyzed perceptual qualities were better perceived in photos of real materials in comparison to renderings, but there is also a perceptual gap between photos and physical materials. This effect has proven to be true especially, but not exclusively, for tactile attributes and the fabric samples. Furthermore, motion information did not affect the perception of digitized materials significantly. The latter is especially relevant for the 'shiny-matte' dimension as they may contradict the documented fact that motion cues can override static ones while judging shininess [DFY+11, WFEM10]. Nevertheless, their experiments are based in much simpler appearance models (isotropic Ward model and grayscale Phong model respectively) which probably led to a better shininess isolation and recognition. Finally, our investigations indicate that more attention has to be paid to the accurate reconstruction of the distinctive material geometry as well as the acquisition of material appearance under grazing angles. In our measurements, the lowest camera was mounted with a zenith angles of 67.5° and, hence, these particular

appearance effects cannot be recovered.

Although our studies provide interesting evidences, they cannot be extrapolated to other material categories (e.g. paper, stone, wood, etc.) for which additional experiments would have to be performed. We also acknowledge certain aspects that could have limited the expressiveness of the digitized materials used in our experiments, including:

- The generation of the virtual scene is approximate, i.e. the virtual camera position and the scene geometry slightly deviate from their physical counterparts.
- Scale differences between virtual and real material sample. Due to restrictions of the acquisition process, the digitized material represents a slightly smaller patch from the original one.
- The environmental light varied during the photo session due to the movement of sun and clouds.
- A color shift between the real and the virtual materials which also comes from the acquisition process.
- Not all the materials presented in this study were suitable to be represented by the SVBRDF appearance model, since it does not account for important surface effects. Consequently, some digitized representations were visibly defective.

By and large, we consider the results presented in this investigation an important step in the immense task of unveiling the perception of digital environments to improve the overall experience. To conclude, we point out the necessity of research in several directions such as the application of more appropriate, material-specific appearance representations. In this regard, BTFs models might help regarding the reproduction of fine effects of light exchange within the digital material representation at the cost of rather long acquisition times. Another interesting avenue of research could be to explore the linkage between perceptual qualities and physical measurable material properties (i.e. stiffness or roughness). Finally, the transmission of material qualities could benefit from a multisensory approach. In particular, the use of sound has proven to be beneficial for the assessment of tactile qualities [MIWH15], which were not successfully transmitted using purely visual models.

5.7 Acknowledgments

This work was developed in the X-Rite Graduate School on Digital Material Appearance at the University of Bonn. We greatly thank Max Hermann and Christopher Schwartz for their assistance during the material acquisition process.

5.8 Supplementary material

This section presents additional material to the paper 'Digital Transmission of Subjective Material Appearance' which is **not** present in the original publication due to space constraints. This includes additional related work relevant to the state of the art, further information regarding the device employed for the acquisition of material samples as well as the screenshots depicting the user interface from the custom android application. Furthermore, we provide the mean opinion scores (MOS) and confidence intervals (CI) of the participants' response ratings.

Relevant related work. Beyond the use of motion as a feature to correctly assess material properties, some physical qualities would be better revealed after certain interactions with the specimen (e.g. stretching the sample to communicate stiffness, letting it fall to better assess hardness). This direction of research is explored by Bouman et al. [BXBF13], where the authors collected a database of videos from moving fabrics and trained a regression model to predict two properties of the fabric, stiffness and area weight. The values predicted are well correlated with the ground-truth measurements, but not that much with human perception.

Moreover, an investigation from Xiao et al. [XBJ+16] addresses the importance of material wrinkles and folding information when discriminating fabric materials, using tactile stimuli as the ground truth. Their results show that the presence 3D shape cues and color information in the stimuli images allows a better accuracy than flattened or grayscale materials.

On the digitization of materials. In order to digitize the selected material samples, we made use of a commercial scanning device [XR16]. The apparatus includes a rotatable sampleholder on which the specimen is placed below a hemispherical gantry with 20 attached LED light sources, four illuminations with 10-band spectral filter wheels, four monochro-

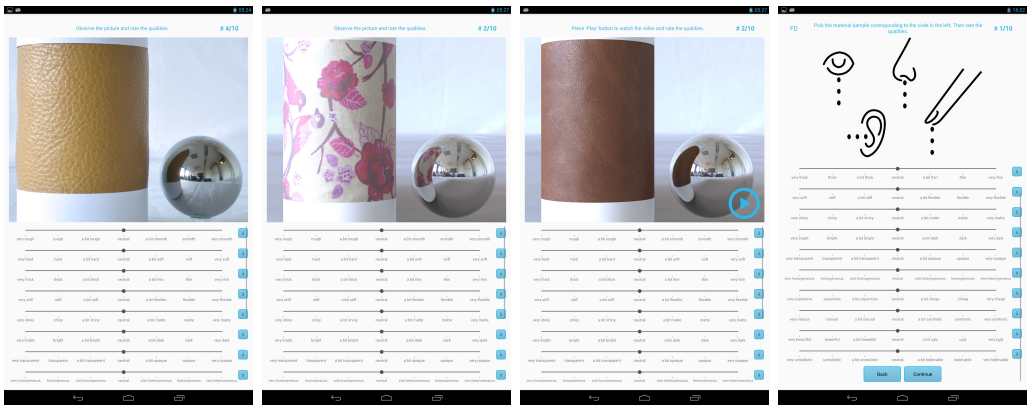


Figure 5.10: *Screenshots from the Android application employed to carry out the studies, where each image corresponds to one of the four conditions composing Experiment 1. From left to right: photograph (PH), digitized rendering (DR), digitized video (DV) and full-modal (FM) conditions.*

matic cameras at zenith angles 5.0°, 22.5°, 45.0° and 67.5° and an additional motorized linear light scanner. After taking images of the material sample from different viewpoints and under different illumination conditions, a surface normal map is obtained and the reflectance behavior is stored in terms of an Spatially-Varying BRDF representation. The latter is based on a diffuse Lambert model and a specular anisotropic Ward model [War92] with bounded albedo [GMD10] and modulated with a simple Fresnel term [Sch94]. The output Appearance Exchange Format (AxF) is natively supported across several CAD and rendering applications.

User interface. The screenshots of the Android application used to conduct the experiments are displayed in Figure 5.10.

Material ratings. The MOS and CI of the participants' response ratings for each experimental condition arranged by material and category are depicted in Figures 5.11, 5.12 and 5.13.

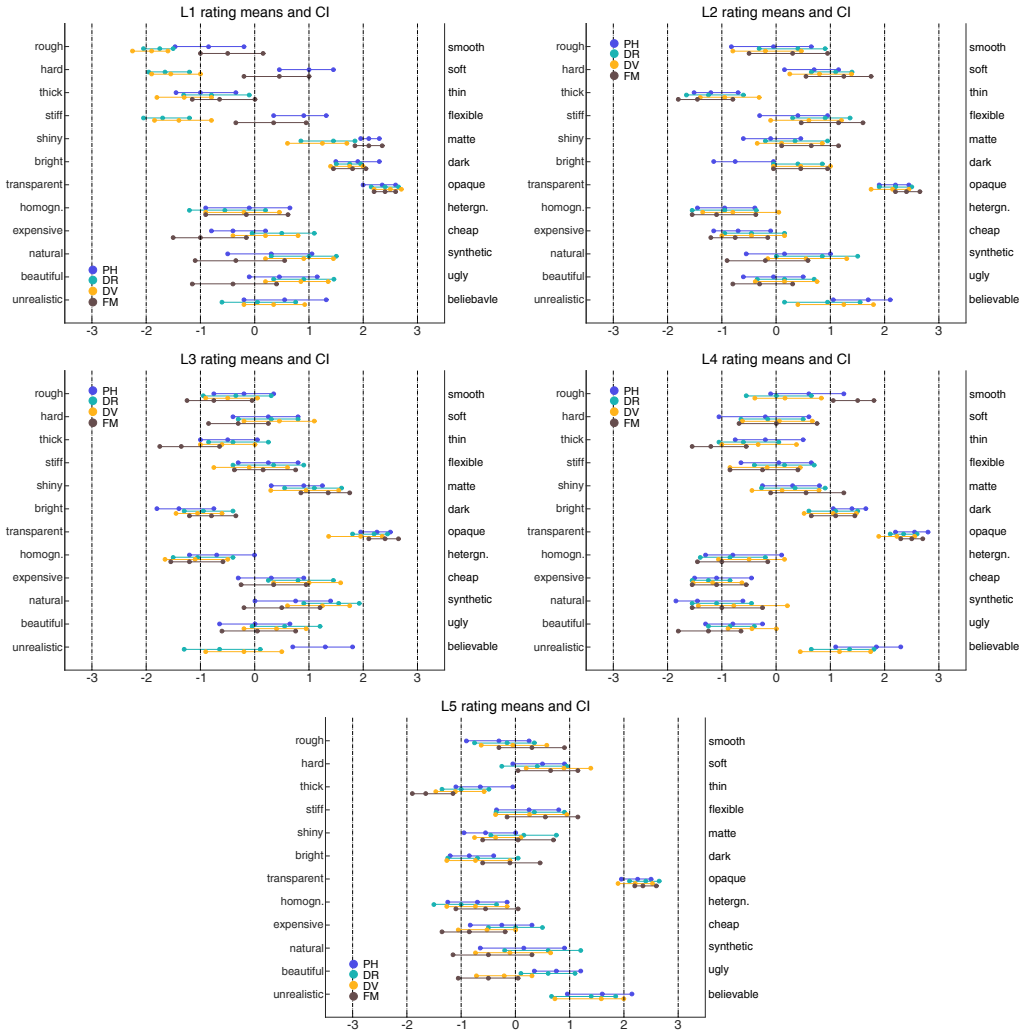


Figure 5.11: *Response ratings from Experiment 1 for each of the investigated material samples (leathers only). The central circles represent the participants' mean rating, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

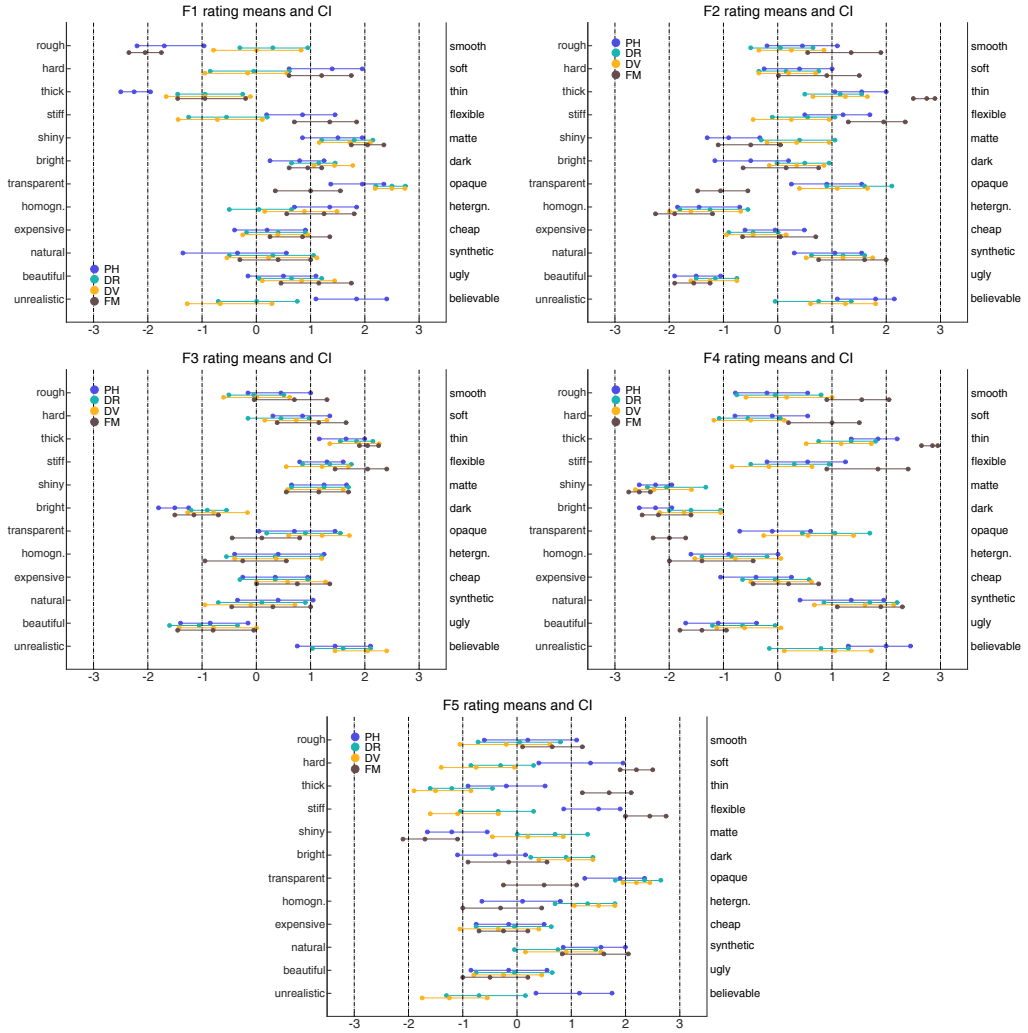


Figure 5.12: *Response ratings from Experiment 1 for each of the investigated material samples (fabrics only). The central circles represent the participants' mean rating, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

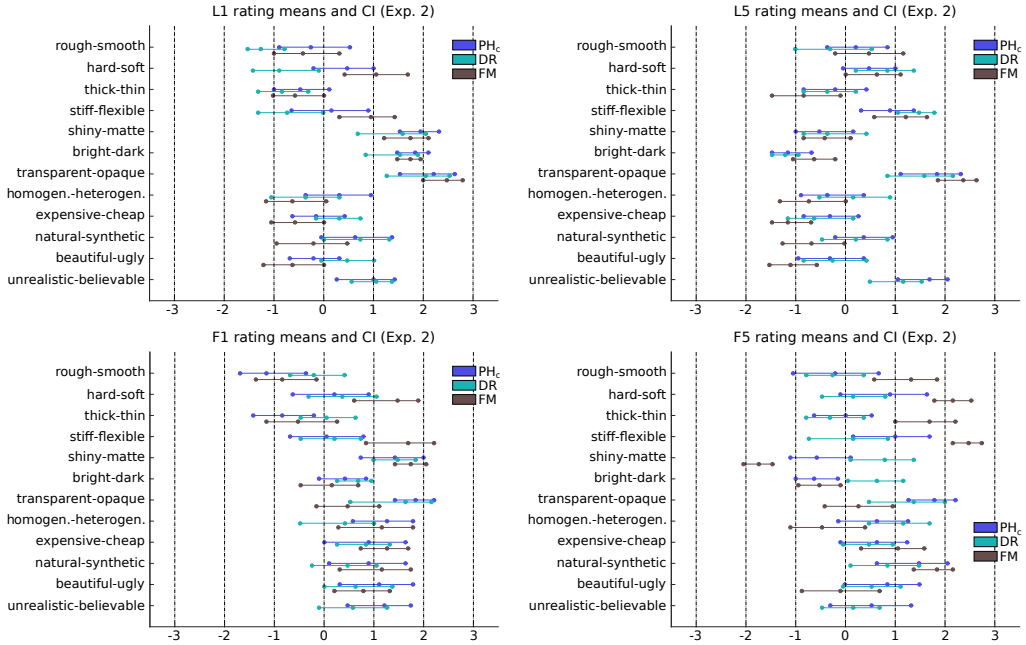


Figure 5.13: *Response ratings from the investigated material samples in Experiment 2. The central circles represent the participants' mean rating, while the outer circles represent the bootstrapped 95% confidence interval for the mean.*

Part III

Perceptual Similarity Metrics

CHAPTER 6

Measuring Perceptual Similarity via Texture Interpolation

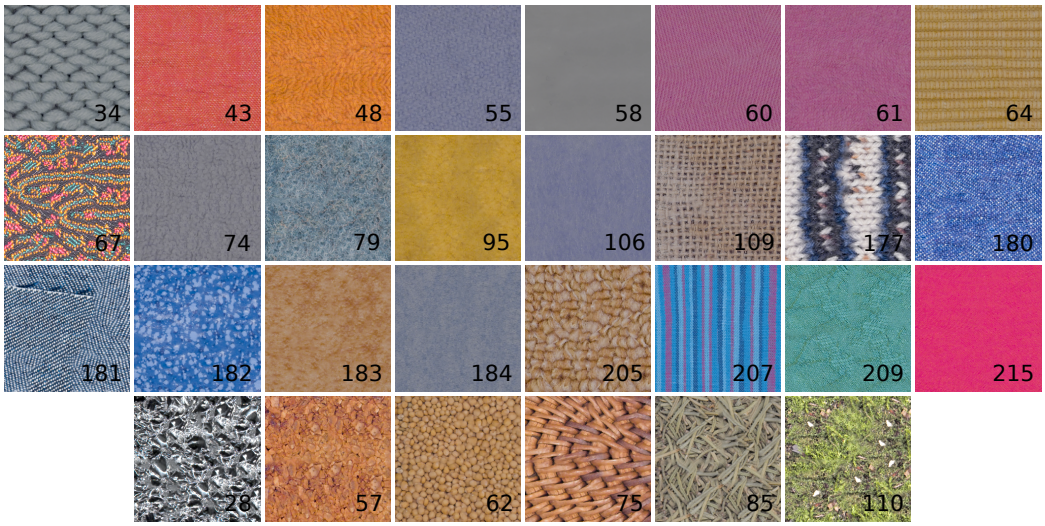


Figure 6.1: *(Teaser) The collection of textures from the ALOT database utilized in our experiments, after several processing steps and before removing the color. The superimposed numbers correspond to indexes given to the materials in the original dataset.*

Abstract. The perceptual similarity or dissimilarity of textures has been subject to considerable research in the imaging and vision communities. Here, we focus on the challenging task of estimating the mutual perceptual similarity between two textures on a consistent scale. Unlike previous

studies that more or less directly queried pairwise similarity from human subjects, we propose an indirect approach that is inspired by the notion of just-noticeable differences (JND). Similar metrics are common in imaging and color science, but have so far not been directly transferred to textures, since they require the generation of intermediate stimuli. Using patch-based statistical texture synthesis, we generate continuous transitions between pairs of materials. In a user experiment, participants are then asked to locate an interpolated specimen in the linear continuum. Our intuition is that the JND, defined as the uncertainty with which participants perform this task, is closely related with the perceived pairwise texture similarity. Using a database of fabric textures, we show that this metric is particularly suitable to address fine-grained similarities, produces approximately interval-scale measurements and is additionally convenient for crowdsourcing. We further assess the validity of the obtained perceptual distances by producing a perceptual texture space that correlates well with previous investigations.

During the preceding investigations, we examined the perceived differences between visual and multisensory material perception as well as between real and digitized materials. The latter was achieved by analyzing how relevant material qualities were assessed by a group of subjects across different stimuli. One of the reasons for such methodology is the lack of effective and suitable metrics for material appearance. This chapter addresses this problem by considering the most common fashion of representing materials in many online applications, textures from real samples, and proposes a novel methodology to compute the perceived similarity between textures which exploits the noise resulting from a texture localization task.

The contents of this chapter were published after the submission of this thesis [MXK+19]: Rodrigo Martín, Xue Min, Reinhard Klein, Matthias B. Hullin and Michael Weinmann. "Using patch-based image synthesis to measure perceptual texture similarity". In *Computers and Graphics*. Jun 2019.

6.1 Introduction

The perception of textures is an essential field of research that has a broad range of applications in numerous vision problems. While textures play an important role in the visual appearance of materials, they also serve as a cue in many understanding tasks such as judging material attributes (e.g. [FWG13]), material recognition and classification (e.g. [SN15, SLRA13, WGK14]) and object detection (e.g. [SWRC06]). The main challenge when dealing with textures is that their effect on human perception is very complex and highly subjective. In this paper, we focus on the problem of perceived texture similarity. Questions like *"are textures A and B more similar than textures C and D"?* arise in many applications, for instance when reasoning whether objects are made of the same or a similar material or when searching materials that are *perceptually in-between* certain other given samples. Reliable texture similarity metrics could greatly facilitate the development of efficient user interfaces for retailing web stores that are consistent with human perception, or be used to guide design processes, find suitable replacements for unavailable materials, and so on.

In color science, distance and similarity metrics have reached a mature state and enabled highly efficient digital solutions for applications like gamut mapping for display and print. Textures introduce additional spatial structure and possibly semantic context, which makes any measure of similarity much less well-posed and subjective. Pioneering research in perceptual texture similarity has aimed at identifying a meaningful, low-dimensional and perceptually uniform space for textures [RL96, RFS+98], where the axes represent meaningful quantities and distances between points in this space are proportional to variations in the human perception. These studies and those who have built upon them rely mainly on *free-grouping* experiments to collect perceptual input. This data are later shaped into a similarity/distance matrix from which a perceptual texture space (PTS) is constructed by means of dimensionality reduction techniques. In general, this approach has proven valid in the task of constructing different PTSs, although the procedure may become impractical for large datasets due to the substantial duration of the grouping task, which may lead to fatigue and boredom issues. In fact, arranging e.g. a large collection of printed images into clusters has been shown to be a time-consuming experiment [CHN+11] taking 2 to 4 hours, where only six out of thirty participants were able to complete the whole assignment. Moreover, this metric, by definition, does not allow the addition of un-

seen points and, given the sparsity of the resulting matrix, it is not clear whether the derived PTS accurately represents the perceived distances for an assortment of samples coming from the same material category (e.g. leathers, fabrics, etc), exhibiting rather similar appearances. Several alternatives have been proposed to collect similarity data while partially overcoming the previous issues e.g. *progressive grouping*, *pairwise comparisons* or *forced-choice combinations*. Even so, they depend on direct user queries for pairwise similarity between textures, a concept that is ill-defined, depends on the individual subject experience and therefore is subject to bias.

In this paper, we attempt to mitigate some of the obstacles from the aforementioned approaches by introducing a novel strategy for the measurement of mutual perceptual similarity between two textures. Our approach relies on a localization task inspired by the notion of *just-noticeable differences (JND)*, and applies state-of-the-art patch-based texture synthesis [DSB+12] to produce a gradual transition, or gradient, between two materials A and B. An additional intermediate texture exemplar C, generated using the same technique, is then given to the user, along with the task to locate it on the continuum from A to B. The working hypothesis for our metric is that the more similar the two materials at the end of the gradient are, the less certain the user will be about the placement of C. For multiple users tasked with the same question, this will result in a variance which we expect to correlate with the similarity of textures. In contrast, a good agreement between users suggests a high dissimilarity or perceptual distance between the two textures. As the texture interpolation itself is not driven by perceptual metrics, the synthesized transition between the textures may not represent a smooth, linear transition between materials. We therefore propose a preliminary study that allows the generation of perceptually linear interpolations. Our approach is easily parallelizable among multiple subjects, extensible to additional samples and produces approximately interval-scale measurements. Thanks to these characteristics, our study is suited for implementation on human intelligence platforms like Amazon Mechanical Turk and results in a space of textures that complements previous investigations.

Altogether, the main contributions of this paper are:

- We introduce a novel formulation for the perceptual similarity between textures that exploits the noise inherent to a JND-like task. To generate the stimuli used for the experiments, we exploit state-of-the-art texture interpolation.
- To derive a perceptually linear interpolation between textures that is so far not guaranteed by current techniques, we present a novel

linearization method based on user judgments.
- From the derived distance matrix, a meaningful, low-dimensional and perceptually uniform perceptual space for material textures is computed.

6.2 Related work

Establishing a texture similarity measure belongs to the essential aspects that need to be considered for understanding the appearance of materials. Therefore, we review major research in the field of material appearance perception as well as related work regarding the similarity of textures and the practical acquisition of psychophysical data. As our approach involves example-based techniques for texture synthesis to generate the experimental stimuli, we also report the main developments in this regard.

6.2.1 Perception of material appearance.

Undoubtedly, the way we perceive our surroundings guides our interaction with objects. Even by solely considering the materials objects are made of, we can infer characteristics regarding fragility, weight, value, etc. As a consequence, there has been a considerable amount of work studying the fundamentals of the perception of material surfaces and their properties [Ade01]. Further research from Fleming et al. [FWG13] relates two of the main tasks encompassed in the field: the inference of material qualities and the classification of materials according to given semantic concepts. The perceived dimensions of sensory material perception has also been a matter of research. In this regard, Okamoto et al. [ONY13] extensively review and complement previous studies on the dimensionality of tactile perception. Sound has received some attention as well in the context of material perception, e.g. by studying its influence in the obtained dimensionality of the perceptual space spanned by a set of material qualities, as reported by Martín et al. [MIWH15].

Beyond investigations on material appearance based on real samples or images as stimuli, research has also focused on the perception of digital material representations. Widely applied material appearance models include bidirectional reflectance distribution functions (BRDFs) that model material appearance depending on the viewing and illumination conditions, as well as spatially varying BRDFs [NRH+77] and bidirectional texture functions (BTFs) [DvG+99] that both additionally capture

spatial variations of the material across the surface. Whereas SVBRDFs allow a compact representation of the reflectance behavior in terms of a parametric function, the computational burden of BTFs is balanced by their capability to also capture effects of light exchange at fine structures such as interreflections, self-occlusions and local subsurface scattering. For a comprehensive discussion on the technical aspects regarding material acquisition and representation, we refer to the survey provided by Weinmann et al. [WLGK16]. The previous models have been subject of perceptual studies in order to discover control spaces for concrete dimensions. Examples include the exploration of a perceptual space for material gloss [WAKB09] and predictable editing of captured BRDF data [SGM+16]. Moreover, a database of digitized materials has been successfully used for classification under complex real-world scenarios [WGK14].

Despite the increasing reproduction quality of digital material representations, they are still not capable of accurately preserving all the fine details that contribute to material appearance. Besides, the resulting gap between virtual materials and their corresponding physical samples may result in the miscommunication of certain qualities, especially those which exhibit spatially-varying characteristics [MWH17]. For this reason, many material catalogs and online fashion [HJ10] illustrate their collections using pictures or even physical samples instead of their virtual equivalents, regardless of their potential benefits.

6.2.2 Perceptual texture similarity.

The visual perception of textures has been intensively studied in the past [LG04]. Measuring texture similarity is an essential aspect of texture perception which has gained particular attention in recent years due to its relevance for several potential applications. The importance of image similarity metrics in different applications has been addressed by Zujovic et al. [ZPN+15], where the authors state the differences between quantitative and qualitative similarity. Quantitative metrics focus on the monotonic relationship between measured and perceived similarity and have implications mainly in the domain of image compression. In contrast, a valid qualitative metric should distinguish between similar and dissimilar images as required by wide-spread applications such as content-based image retrieval and image understanding. The technique proposed in this paper belongs to the second category of metrics and attempts to address the fine-grained distance computation between similar textures as well as to identify those which are not alike.

One of the first efforts in this direction has been presented by Rao

and Lohse [RL96], who identified some relevant perceptual dimensions for textures. Their research establishes an experimental methodology for the acquisition of image similarity data from human annotations based on a free-grouping task, which has been followed to a greater or lesser extent in later studies [CHN+11, Hal12, LDC+15, RFS+98]. These investigations focus on the exploration of a low-dimensional PTS and its relevant dimensions, while using different analysis techniques and texture databases composed either by natural images, procedural textures or rendered heightmaps. In an investigation by Balas [Bal08], the same experimental principle has been applied to reformulate texture similarity as a categorization task in order to compare the efficiency of several parametric texture models. Alternatively, Gurnsey and Fleet [GF01] computed a perceptual space for noisy artificial textures using an experimental design based on triads.

Certainly, the evaluation of texture similarity would profit from more computational approaches. A comprehensive review of the existing structural texture similarity metrics that integrate insights from human perception in order to grant point-by-point differences in textured regions is provided by Pappas et al. [PN+13]. In contrast, many different computational texture features for image classification have demonstrated good performance on texture databases (e.g. Gabor Filters [MM96], Local Binary Patterns [OPM02] and Filter Banks [VZ05]), although they do not always correlate well with observations made by humans [CHN+11], especially for image pairs that appear dissimilar to the observers. Furthermore, Dong et al. [DMC14] investigate the capability of a broad set of computational features to estimate texture similarity in comparison to perceptual rankings in a unified framework. Despite obtaining interesting insights on which features correlate better with perceptual data, none of the employed traits was able to ideally replicate human performance. In addition, hand-crafted features and deep image features extracted with convolutional neural networks (CNN) were used to train regression models against perceptual data from free-grouping experiments [LQD+16]. Their results indicate that a combination of deep features casts results that correlate slightly better with user ratings. Motivated by the intuition that the traits that influence human perception of material similarity correspond to visual ones, Schwartz and Nishino [SN15] address material recognition by exploiting human perception of visual similarity. In their research, ground-truth perceptual information is gathered via crowdsourcing where small image patches from natural images are shown to users who rate whether they look similar or not. The obtained ratings are used for the representation of material distances in terms of

average probabilities of similarity to the different categories. While the pairwise similarities can be obtained in a confident manner, they do not represent a direct measurement of distances on a continuous scale. With our approach, we avoid directly querying annotators for texture similarity and allow them to perform the rating on a consistent scale.

6.2.3 Practical acquisition of psychophysical data.

As indicated earlier, the assessment of texture similarity or distance generally involves a methodology which comprises the computation of a *perceptual similarity matrix* (PSM) from a free-grouping experiment. However, this procedure and the resulting perceptual similarities present several points in question. As the dataset grows larger, the user task may become rather impractical, since a considerable amount of textures has to be compared simultaneously. Indeed, it has been demonstrated that the election of a data collection method is critical in data positioning studies. Concretely, methods which entail larger completion times significantly affect the fatigue, boredom, missing values and the resulting perceptual map [BW95]. Besides, this setup does not support collecting similarities for unseen points and the derived PSM is very sparse, due to the shrinking possibility of two samples being grouped together for increasing sizes of the dataset.

Alternatively, Rogowitz et al. [RFS+98] propose two psychophysical tasks. In the first, participants arrange the images on a table and measure the physical distances between samples to obtain a dense similarity matrix. In the second, a forced-choice task is presented to the users to select the most similar image to a reference in random batches of eight samples shown at a time, resulting in a rather sparse matrix. Although the obtained PTSs resemble one another, the one from the forced-choice experiment presents a less-informative circular structure. Liu et al. [LDC+15] address the sparsity problem by dividing the database into smaller subsets to be grouped individually and later merged into larger clusters, until all groups are merged together. This procedure allows to assign an extra confidence score to each action based on the visual similarity between groups of textures. An additional progressive grouping assignment has been proposed by Zujovic et al. [ZPN+15]. This method allows the parallelization of the data collection process across different subjects and minimizes the completion time, but still lacks a solution regarding the sparsity of the similarity matrix and its extension to new points.

Approach	Task complexity	Parallelizable	Human factors	Extensible	Matrix sparsity	Data quality
Free-grouping [Bal08, BW95, CHN+11, Hal12, RL96, RFS+98]	$O(n\log(n))$	**no**	✗✗	**no**	✗	✓
Progressive grouping [LDC+15, ZPN+15]	$O(n\log(n) - nlog(m))$	**yes**	✓	**no**	✗	✓
Conditional ranking [BW95]	$O(n^2)$	**yes**	✗	**no**	✓	✓
Pairwise comparisons [BW95]	$O(n^2)$	**yes**	✓✓	**yes**	✓	✗
Triadic combinations [BW95, GF01]	$O(n^3)$	**yes**	✓✓	**yes**	✓	-
Multiple forced-choice [RFS+98, SN15]	$O(kn)$	**yes**	✓	**no**	✗	✗
Our method	$O(n^2)$	**yes**	✓✓	**yes**	✓	✓

✓ Better performance ✗ Worse performance - Not enough evidence m = size of the subset k = stimuli presented at a time

Table 6.1: *Benefits and drawbacks of different data collection methods. This comparison is conducted considering the complexity of evaluating the entire database, their ability to be parallelized across users or to be extended to unseen points, the influence of fatigue and boredom (human factors) in potentially large databases, the sparsity of the derived matrix and the general quality of the collected data based on the resulting PTS.*

Our proposed setup relies on pairwise texture similarities obtained from micro-tasks performed by participants on a crowdsourcing system, thus overcoming the previous issues regarding the parallelizability of the task, the sparsity of the resulting matrix and the extension of the database to unseen textures. Other data collection methods, like pairwise comparisons, may be similarly used for crowdsourcing. However, this technique depends on subjective magnitude judgments which are highly prone to bias [RFS+98]. Instead, we propose to explore the noise inherent to a JND task as a mean to gather high-quality similarity assessments. A comparison between the predominant methods to collect similarity data is provided in Table 6.1. The latter is given in terms of the complexity of evaluating the entire database, their capacity to be parallelized across users or extended to new points, the influence of human factors (fatigue and boredom) in experiments with potentially large databases, the sparsity of the computed similarity matrix and the overall quality of the collected data based on the derived PTS.

6.2.4 Example-based texture synthesis.

Beyond the analysis of texture perception and the features that contribute to this task, several investigations approach the generation of synthetic textures while trying to preserve the characteristics of given real-world samples. This allows producing texture variations that depict the same material as well as creating larger exemplars. While a comprehensive survey on this research domain is provided by Wei et al. [WLKT09], which covers a broad set of synthesis techniques from simple pixel-based to the more complex patch-based synthesis [EL99, PS00], the recently emerging developments in deep learning can also be utilized for texture synthesis [GEB15].

Other than producing different variations or bigger exemplars from example images, texture synthesis can be also applied to generate a transition region between two source images, such that inconsistencies in color, texture and structure change gradually from one source to the other. In this regard, Ruiters et al. [RSK10] combined patch-based and statistical approaches to produce high-quality interpolated texture patches between two sources, however, at rather long computation times of several hours. Furthermore, their method requires a certain degree of user input, as the synthesis is guided by manually-generated feature maps. Later, Darabi et al. [DSB+12] developed a technique named *Image Melding*, which combines the potential of patch-based and gradient-domain approaches. This procedure builds on a modified version of the *Patch-*

Match algorithm [BSGA10] in order to compute nearest neighbor patch correspondences, including additional key optimizations to successfully interpolate between textures with rather different structures, without sacrificing texture sharpness during the process. The procedure is fully automatic and produces convincing sequences with an acceptable time performance (in the order of minutes). More recently, deep learning has been applied for texture interpolation [GEB16, UGP+17]. However, the interpolations are performed only for relatively small image patches and the results are noisy due to artifacts produced by the neural networks. In order to generate the stimuli for our work, we make use of the Image Melding method [DSB+12] (IM) mainly due to its excellent trade-off between time performance and visual quality. Nevertheless, our methodology for computing texture similarities is not constrained to the utilization of one particular synthesis algorithm.

6.3 Stimuli

Before providing detailed discussion regarding our novel approach for the formulation of a perceptual similarity measure in the next section, we introduce the involved image stimuli obtained from selecting a suitable texture dataset and a few pre-processing operations. The diagram for the generation of the stimuli and its embedding into the overall methodology is illustrated in Figure 6.2.

6.3.1 Selection of texture database

For the present study, we focused on fabric materials, since they exhibit a rich and diverse range of appearances, have been largely used in previous investigations and are of great importance in industrial applications. Besides, the boom of the e-commerce has been particularly significant in the world of fashion, making these textile materials a perfect research candidate with direct applications of great relevance. In order to gather such a collection of respective fabric textures, we inspected existing material datasets, which were extensively reviewed in the survey provided by Hossain and Serikawa [HS13]. We then discarded collections of textures captured under uncontrolled lighting or viewing conditions (e.g. Brodatz [Bro66], MeasTex [Mea97], UMD database [XJF09]), as it has been demonstrated that changes in illumination drastically affect the appearance of textures and materials [Cha94], and would significantly bias the users' judgments. Likewise, databases in which samples were not annotated or

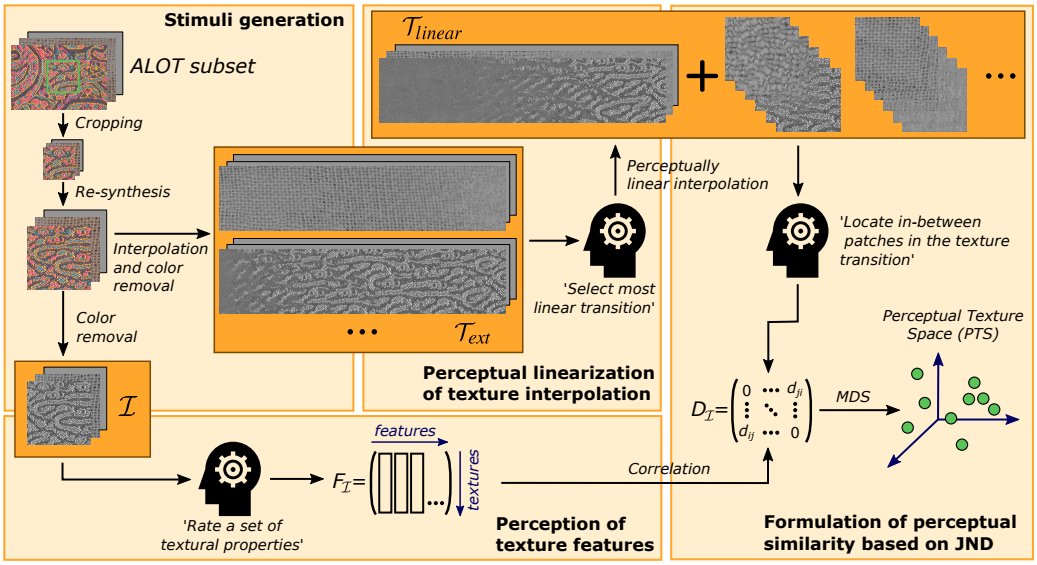


Figure 6.2: *Overview of our approach for the derivation of a perceptual texture similarity metric: In an initial step, stimuli are generated by cropping and resynthesizing the images and removing the color information in order to focus on the textural characteristics. In addition to obtaining a set of single gray-scale images ($\mathcal{I}$), continuous interpolation strips between pairs of textures are synthesized. As the computed sequences are only linear w.r.t. computational features, we perform an additional perceptual linearization by asking subjects to select the most linear interpolation among a set obtained via different power functions. From the resulting linearized strips ($\mathcal{T}_{linear}$), interpolants are synthesized by sampling at pre-defined locations along the sequence. By letting people locate the generated interpolants on the correspondent sequence and analyzing the variance in the obtained ratings, we derive a distance matrix $D_{\mathcal{I}}$ containing the perceptual distances between pairs of textures (I_i, I_j). Multidimensional scaling (MDS) was then used to compute a perceptual space of textures (PTS). Finally, to evaluate the correspondence between the dimensions of our PTS and the textural features, we involved user ratings regarding textural properties for $\mathcal{I}$ such as roughness, contrast or randomness.*

which contained too few fabric-like images (e.g. CUReT [DvG+99], PhoTex [Pho18], PerTex [Hal12]) were equally disregarded.

As a result, in the scope of our investigation we make use of a subset of textures from the *Amsterdam Library of Textures (ALOT)* database [UA09], which fulfills the mentioned prerequisites. This collection consists of 250 annotated materials that have been captured under six different illumination conditions, four viewpoints and four camera orientations plus one

additional image for each camera at a reddish spectrum. Therefore, a total number of 100 images is captured per material sample. From the original annotations, we identified 24 materials coming from fabric/textile samples and selected one single color image per material, taken with a camera oriented perpendicular to the sample surface (condition c1 in ALOT) under illumination by all hemi-spherically distributed light sources involved in the setup (condition l8 in ALOT). Furthermore, we expanded this selection with 6 supplementary samples from diverse material categories (paper, man-made, foliage, food), in order to study their behavior in the derived space of textures. This subset, at the lowest resolution provided by the authors (384 $\times$ 256 pixels) for presentation convenience, served as the initial input for our experiments.

6.3.2 Stimuli generation

Our experimental studies rely on measuring the accuracy with which users are able to locate intermediate patches in the interpolated sequence between two samples. For the purpose of generating these sequences, we exploited the Image Melding (IM) algorithm [DSB+12], which is a standard texture interpolation technique for the generation of interpolated sequences between two given textures[1]. It is to be expected that a higher diversity in spatial structure between the images would lead to a wider search space for the PatchMatch method integrated in IM and, consequently, the interpolation process would produce less deterministic and more inconsistent results. For this reason, we reduced the structural variability of the initial images by using IM to re-synthesize a cropped, normalized patch (128 $\times$ 128 pixels) from the original image into a larger version with a resolution of 256 $\times$ 256 pixels. Additionally, some of the cropped patches were manually rotated to align dominant structures in the respective textures accordingly, in an initial step to later facilitate the interpolation process. The steps for the re-synthesis process are illustrated in Figure 6.2. Like most of the synthesis algorithms, the computational performance and visual quality of IM depends on a plethora of involved parameters that have to be tuned up manually beforehand, to achieve the best possible results. To allow a fair comparison of the resulting images, we include the most important IM parameter values employed to generate our stimuli in the middle column from Table 6.2. For further information on this set of parameters, we refer the reader to

[1]Source code has been published by the authors at http://www.ece.ucsb.edu/~psen/melding

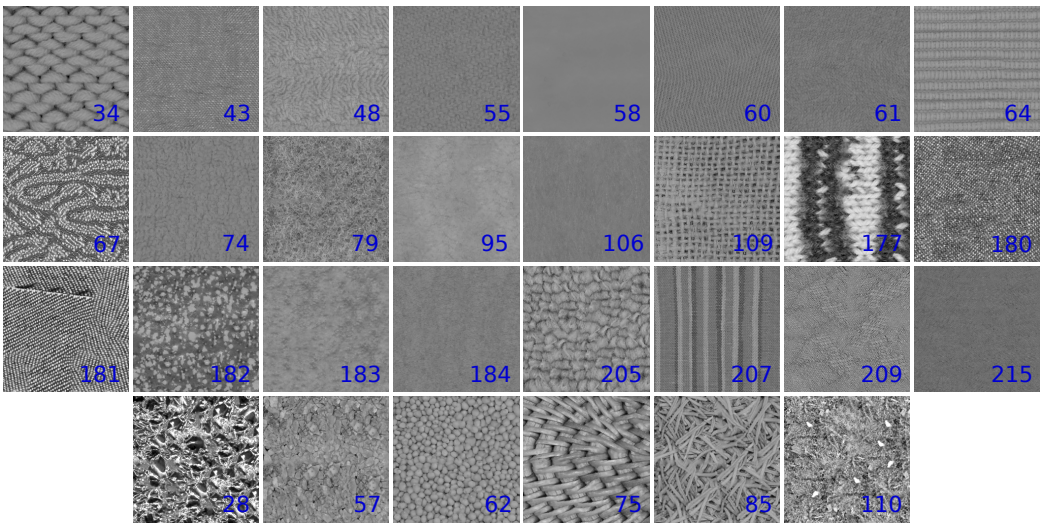


Figure 6.3: *The collection of textures utilized in our experiments ($\mathcal{I}$), after several processing steps (normalization, cropping, re-synthesis and color-removal). Non-fabric materials gathered to diversify the dataset are presented in the lower row. The superimposed numbers correspond to indexes assigned originally to the materials in the ALOT database [UA09].*

the original publication [DSB+12].

Color has been documented to be one of the most salient visual cues when judging image similarity [RFS+98]. In general, separate subjects tend to differently balance color composition and structure when assessing texture similarity and, in consequence, more effective metrics can be developed when color and structure are decoupled [ZPN+15]. Given that this study particularly focuses on the structural similarities between textures, we followed the previous insight and performed the entirety of our experiments in luminance-only images, in accordance with related investigations [Bal08, CHN+11, LDC+15, RL96, ZPN+15]. Thereby, we ensure that the obtained perceptual distances are driven by textural features instead of merely color. However, for the stimuli generation step itself, we used full-color images as the input to the synthesis algorithm.

Figure 6.3 displays the complete subset of 30 textures selected from the ALOT database as obtained after the processing steps (normalization, cropping, re-synthesis and color-removal). Using this image collection $\mathcal{I}$, we employed again the IM algorithm to compute all possible interpolations $T_{ij} \in \mathcal{T}$ between image pairs (I_i, I_j) with $I_i, I_j \in \mathcal{I}$ and $i \neq j$. The method is capable of searching for similar patches in both input images to produce a sequence that renders a strip presenting a smooth transition

Parameter	Re-synthesis	Interpolation
Patch size (in pixels)	10	10
Gradient weight	0.1	2
Random search	**off**	**off**
Uniform scale (range)	$[0.9, 1.2]$	$[0.9, 1.3]$
Relative scale (range)	$[0.9, 1.1]$	$[0.9, 1.1]$
Rotation (range)	$[-\pi/4, \pi/4]$	$[-\pi/4, \pi/4]$
Normalized weight	**off**	**off**
Interpolate gradients	–	**on**

Table 6.2: *The main parameter values from the Image Melding algorithm for the re-synthesis and the interpolation steps, respectively.*

between the original textures. However, this transition is computed based on the linear interpolation of computational features which does not correspond to a perceptually linear interpolation required for the estimation of a perceptual distance metric. This means that the interpolations do not represent a smooth, pleasant, evenly distributed transition between two textures to human observers (see Figure 6.7). The latter can particularly be observed for exemplars with features of different sizes or large variations in the magnitudes of the vertical and horizontal image gradients, in which IM relies in order to ensure the sharpness of the final sequences. Consequently, the interpolations do not represent a quantitative measurement scale and cannot be directly used as stimuli for the derivation of a perceptual similarity measure. Instead, an additional perceptual linearization of these interpolations is required.

6.4 Measuring perceptual similarity

In this section, we provide the details regarding the derivation of a novel perceptual similarity measure based on the concept of just-noticeable differences (JND). As illustrated in Figure 6.2, the interpolation stimuli generated according to the descriptions in the previous section have to be perceptually linearized. Then, the analysis of the user variance in the localization of patches in-between the interpolation sequences allows the derivation of a matrix $D_{\mathcal{I}}$ that contains the perceptual distances between pairs of textures (I_i, I_j) in the dataset. Finally, we also examine human ratings regarding concrete texture features to later evaluate the correspondence between such ratings and the dimensions of the derived PTS. In

the following, we describe these steps in more detail.

6.4.1 Perceptual linearization of texture interpolation

Originally, IM requires the patches in the interpolated image T_{ij} to be similar to both source images I_i and I_j, where the relative contribution of each source is modulated by a linear numerical interpolation. In order to achieve a more evenly distributed transition between the sources in accordance with human perception, we propose to guide the synthesis process with the following bijective power function instead:

$$g(x) = \begin{cases} x^{(\alpha+1)} & \text{if } \alpha \geq 0 \\ 1-(1-x)^{(1-\alpha)} & \text{else} \end{cases}, \tag{6.1}$$

In order to determine the values of α that produce perceptually linear transitions, we designed a preliminary psychophysical experiment to compute this mapping for each texture combination. To that end, we generated a set of interpolated stimuli by fixing the values of α to $\alpha \in \{-10, -5, -2.5, -1, -0.5, -0.25, -0.1, 0, 0.1, 0.25, 0.5, 1, 2.5, 5, 10\}$, resulting in a total of $435 \times 15 = 6{,}525$ image strips $T_{ij}(\alpha) \in \mathcal{T}_{ext}$, each having a resolution of $1{,}152 \times 256$ pixels. The values of the synthesis parameters were kept during all calculations, yet they differ slightly from those in the re-synthesis step (see Table 6.2).

To obtain the user judgments, we developed a custom website with an interface in which all the 15 interpolations for a concrete T_{ij} were presented to the users (one-at-a-time), together with a slider widget and text with detailed instructions (see Figure 6.4). By moving the slider, the user was able to select the value of α for a concrete texture pair, and ultimately control which particular stimulus $T_{ij}(\alpha)$ was displayed on the screen at a certain moment. The instructions encouraged the users to interact with this widget in order to *"find the best slider position where the transition becomes as smooth (pleasant, even, linear) as possible"*. Each of these elections were considered to be a psychophysical assignment or task.

Then, we arranged a psychophysical pipeline in which the study was divided into several micro-batches consisting of 20 assignments each, where half of them consisted of the same stimuli with the interpolated strip reversed horizontally. Besides, 5 additional assignments were included for training purposes and to ensure the validity of the participants' answers (see paragraph on data verification). As crowdsourcing using Amazon Mechanical Turk (AMT) has been demonstrated to be a convenient practice to collect perceptual data, we integrated our pipeline

into AMT, and every individual micro-batch was presented to groups of between 12 and 15 Amazon workers. In total, 643 workers participated in this initial experiment. In the following, the pre-processing and verification of the answers given by the participants will be explained.

Data verification. During early trials with AMT, we observed that an important amount of workers did not try to perform the task correctly or did not understand the instructions adequately. As reported by earlier investigations in the vision community, these sort of problems are not unfamiliar to crowdsourcing platforms due to the to the limited amount of control over the potential subjects. For instance, Bell et al. [BBS14] devoted quite some effort to automatically identify and discard 31% of the workers on a brightness localization task. In a similar manner, we incorporated several tools into our pipeline to facilitate the realization of the assignments, automatically detect ineligible workers and, in short, ensure the quality of the resulting data. These tools, are listed next:

1. A training step prior to the actual experimental procedure. In this step, the same interface with more detailed instructions and convenient feedback indicating the correctness of the answer was presented to help the realization of the **two** initial assignments. Although the precise α-value for this stimuli was unknown, a loose confidence interval around a plausible answer was established. These training assignments were kept constant for all the batches.
2. A testing step for the **three** assignments subsequent to the training. These assignments were also fixed for all the batches. Again, a loose confidence interval around a reasonable answer was set. In this case, however, we rejected the workers whose selection of α lied outside that interval. The discarded workers were free to try the task again from the beginning.
3. Examination of the worker's Root Mean Square (RMS) error between the ratings for the original and horizontally-reversed images. If the RMS exhibited is larger than a generous threshold, the worker's results were excluded from further analysis.
4. Inspection of the mean elapsed time per worker and assignment. The answers from workers who present significantly lower values that the overall mean time in the same batch were equally removed from the posterior analysis.
5. Analysis of the inter-participant correlation coefficients. Workers whose coefficient is significantly lower than the overall mean in the same batch were considered to be outliers, and their results were

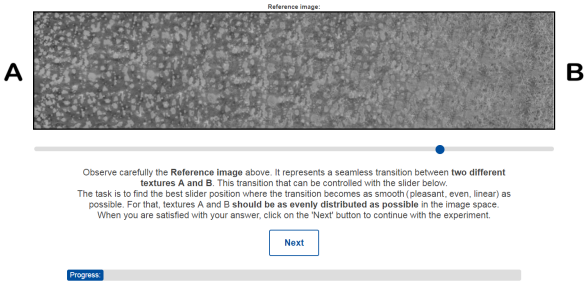


Figure 6.4: *Experimental interface for the perceptual linearization of texture interpolations. The users were asked to find the best slider position where the transition becomes as smooth, linear and evenly distributed as possible.*

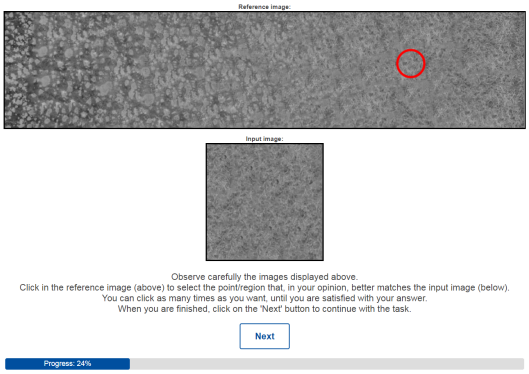


Figure 6.5: *Experimental interface for the formulation of perceptual similarity based on JND. The users had to localize the region in the interpolation strip shown above that best corresponds to the image below.*

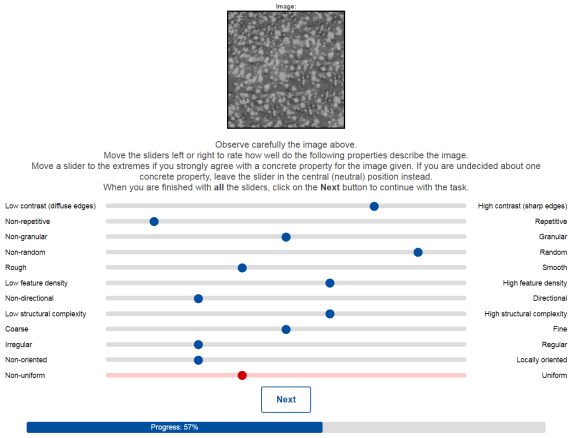


Figure 6.6: *Experimental interface for the perception of texture features. The users rated a set of texture qualities on a bipolar scale.*

excluded from further analysis.

Indeed, we observed a high degree of correlation between the values for the RMS between similar stimuli, mean elapsed time and inter-participant correlation. In other words, the subjects that did not perform the experiment adequately, tended to exhibit low average elapsed time, high RMS error and low inter-participant correlation values. This fact greatly facilitated the automatic detection and removal of ratings coming from ineligible workers. In general, we preferred discarding doubtful candidates and collecting additional ratings and we set the thresholds accordingly, in favor of the efficiency and reliability of the final data. Altogether, the ratings provided by 88 workers were not considered for further analysis, implying that we discarded 13.68% of the collected data, of which roughly a third were related to workers that passed the preliminary tests and then submit default answers (i.e. went through the stimuli leaving the slider in the default position). This user practice was also reported by Bell et al. [BBS14].

From the remaining data, we obtained a total of $555 \times 20 = 11{,}100$ answers, 25.51 on average per texture combination T_{ij}. The final texture synthesis will then be guided by the numerical interpolation resulting from $g(x)$ where the value of α is the average of the users' judgments, for each $T_{ij} \in \mathcal{T}_{ext}$. A comparison between the newly generated, perceptually linear interpolations and the original ones is illustrated in Figure 6.7 for two example combinations. As it can be observed, the new strips present a more evenly-distributed transition between the two material sources. In the scope of our evaluation, we demonstrate that these transitions do not only appear visually linear, but also represent an almost accurate interval scale which is well-suited for perceptual measurements. Therefore, we can use the generated strips as input stimuli for our main experimental setup as described in the following section.

6.4.2 Formulation of perceptual similarity based on just-noticeable differences

The main purpose of this investigation is the computation of perceptual similarities between texture pairs while mitigating some of the shortcomings from analogous methods. These points in question are given by the complexity, parallelizability and time duration of the perceptual assignment, the sparsity of the PSM and its extension to unseen samples. With this in mind, we lean on the concept of JND, which has been extensively used in the domain of experimental psychology. The fundamental princi-

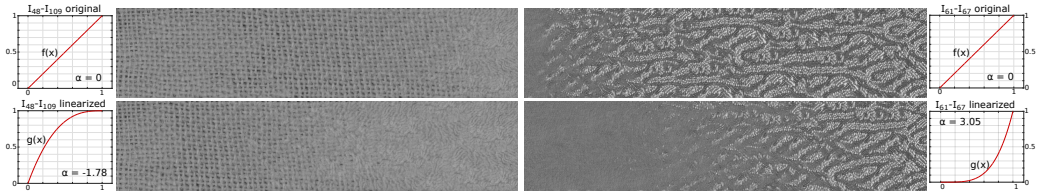


Figure 6.7: *Two example interpolated strips (I_{109}, I_{48}) on the left and (I_{205}, I_{67}) on the right, displaying the original transitions (top) and the linearized transitions (bottom). The adjacent plots depict the numerical interpolation guiding the synthesis, $f(x)$ (original) and $g(x)$ (linearized). Note how the in the images above the texture sources do not contribute equally to the final image, while the images below present a smooth and evenly distributed contribution to the final interpolation.*

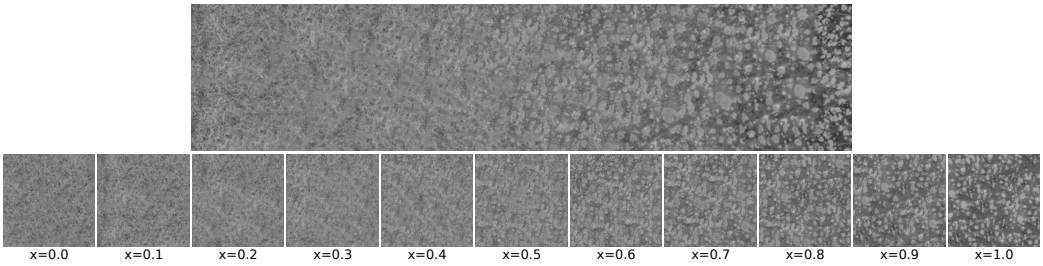


Figure 6.8: *An example reference transition between two textures (I_{79}, I_{182}) utilized in the main experimental setup (top row) as well as the intermediate interpolating patches to be located by the user in the reference image (bottom row).*

ple is the measurement of how much we need to change a stimulus along a given dimension before people notice the change [CW11]. To exploit this notion with regard to our investigation, we analyze the accuracy with which users are able to locate intermediate patches in a linearly-changing continuum, obtained from the preceding perceptual linearization of interpolated texture sequences. The underlying intuition is that users' precision when performing this task correlates with the pairwise texture similarity. More concretely, if the user responses for a particular pair show a low accuracy, the textures at the end of the gradients are more likely to be perceptually similar, as the interpolants cannot be distinguished. Contrarily, a high accuracy would indicate that the textures are perceived more dissimilar, as the intermediate patches can be located precisely. This way, we avoid direct ratings of the similarity or dissimilarity between textures, which is an ill-defined task that depends on subjective user characteristics.

While the preceding perceptual linearization allows us to produce visually smooth transitions between pairs of material textures, we additionally generated intermediate interpolating patches (resolution of 256×256 pixels) by synthesizing fixed intervals of the interpolation $g(x)$ where $x \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. An exemplar of these patches $T_{ij}(x)$ for a concrete texture pair (I_i, I_j) together with the corresponding linearized strip $T_{ij} \in \mathcal{T}_{linear}$ is depicted in Figure 6.8.

With this experimental setup, we are able to parallelize the similarity judgment task into a collection of micro-batches to be integrated in human intelligence platforms (AMT in our case), mitigating possible traces of tiredness while performing the exercise. Each of the batches consisted of 40 assignments (including 6 repetitions to test user consistency) in which a randomized **reference** image (the interpolation sequence) was displayed, together with one of the corresponding **input** textures (an intermediate patch), whose value for x was also randomized. Furthermore, we allowed the reference images to be shown horizontally reversed during the assignments. Through additional explanatory instructions, the workers were asked to *"locate the region within the interpolation sequence shown above that best corresponds to the image shown below"*. Figure 6.5 shows the experimental interface for a concrete stimuli assortment.

A total of 1,183 workers participated in this experiment. Similarly to the previous study, we incorporated several mechanisms in our pipeline to facilitate the completion of the exercise and, ultimately, to automatically discard ineligible workers. In this occasion however, we were not able to use the inter-participant correlation, as the stimuli presented to the workers were fully randomized. With this process, we detected 58 unsuitable workers whose results were not considered in the posterior analysis, which corresponds to discarding only 4.9% of the participants. From the rest of the workers, we gathered $1{,}183 \times 40 = 47{,}320$ user clicks, 108.78 on average per texture combination. For each T_{ij}, the uncertainty of the participants was measured by computing the Root Mean Square *(RMS)* of the differences between the user clicks and the centroids calculated for each set of clicks $\mathcal{C}_{ij}(x)$, where x is fixed to the values given earlier. Figure 6.9 exemplifies this calculation for two texture pairs with large and small perceived similarities respectively. The normalized RMS values shaped our perceptual similarity matrix $S_{\mathcal{I}}$, where s_{ij} describes the perceived similarity between textures I_i and I_j.

Moreover, we introduced a user satisfaction survey at the end of the study that requested participants to rate, on a Likert scale from 1 to 7, whether the task was understandable, whether it was easy to accomplish, if they enjoyed the realization of the experiment and whether they con-

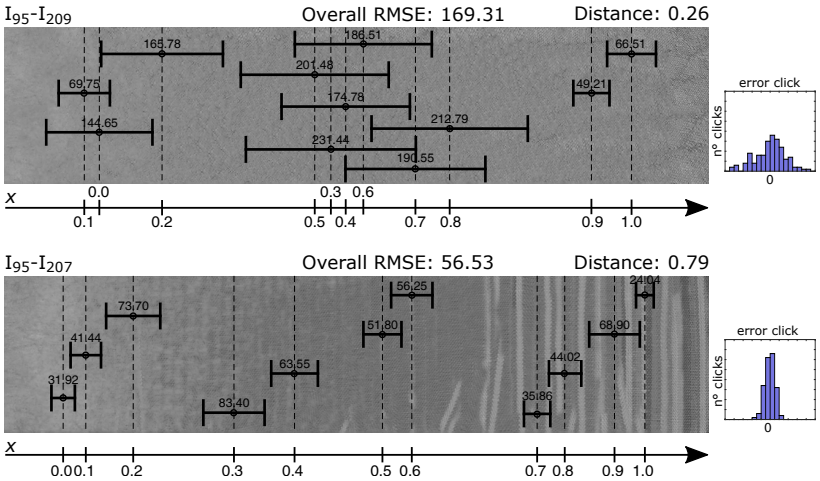


Figure 6.9: *Uncertainty of the users represented by the RMSE across the measurement scale for two example pairs which exhibit low* (I_{95}, I_{209}) *and large* (I_{95}, I_{207}) *perceived distance. The error bars correspond to the RMSE for a particular x value, while the vertical lines coincide with the centroids. The respective histograms of the clicks are shown to the right.*

sider the payment to be fair. The average answers to this questionnaire showed that the exercise was well understood (6.27), relatively easy to accomplish (5.54), enjoyable (5.66) and also adequately paid (4.48). The vast majority of positive comments left by the workers confirmed the good reception of our experimental setup and did not support possible evidences of disinterest, fatigue or similar issues.

From the derived similarity matrix and its corresponding distance matrix $D_{\mathcal{I}} = 1/S_{\mathcal{I}}$, we can compute a low-dimensional texture space by applying dimensionality reduction techniques such as multidimensional scaling (MDS). To also evaluate the degree of correlation between the main dimensions of this texture space with certain textural features, we perform a respective analysis following the descriptions in the next section.

6.4.3 Perception of texture features

Whereas the thorough analysis of perceptual texture features is a challenging problem itself that is beyond of the scope of our study, we analyze the plausibility of our approach by comparing whether the dimensions of the texture space derived from our similarity matrix correlate to well-known textural features used in previous investigations. For this

purpose, we collected user ratings regarding a group of perceptual feature scales for each of the texture samples in $\mathcal{I}$. We employed the set of 12 perceptual qualities defined by Rao and Lohse [RL96] (contrast, repetition, granularity, randomness, roughness, feature density, directionality, structural complexity, coarseness, regularity, local orientation, and uniformity), which were rated using a 9-point Likert scale. This bipolar method measures the positive or negative response to a particular quality by labeling the two ends of the scale (e.g. low contrast (1) and high contrast (9)). Thus, 24 opposite feature adjectives were given to characterize our database of textures.

Again, we leaned on AMT to gather 93 workers in order to perform this psychophysical experiment whose corresponding web interface is depicted in Figure 6.6. At the same time, this study was divided into three separated sub-batches, each of them having a group of 31 workers rating this set of features for 10 particular textures. From the user responses, a feature matrix $F_{\mathcal{I}}$ was constructed by averaging these ratings, where f_{ij} represents the j-th perceptual feature for the texture i.

6.5 Results

In this section, we firstly evaluate our perceptual similarity metric based on the distribution of the user clicks obtained for the task of locating texture patches along the corresponding texture sequences, paying particular attention to the *level of measurement* or *scale of measure* [Ste46]. This is followed by the analysis of the perceptual texture space derived from the subjective data.

6.5.1 Evaluation of perceptual similarity metric

To validate our metric, we analyze the distribution of the ratings obtained from the main experimental study. This includes the validation of the hypothesis that the derived perceptual similarity correlates with larger variances in the user ratings (as described before), the analysis of whether the texture similarity measure shows approximated interval scale behavior as well as the evaluation of the distance metric itself.

In order to verify the presumed correlation of perceptual similarity with the user variances, we inspected the distributions of the positions in the horizontal axis selected by the participants when locating intermediate texture patches, exemplified in Figure 6.9 for two sample interpolations. Indeed, the variances in the user judgments were lower for visu-

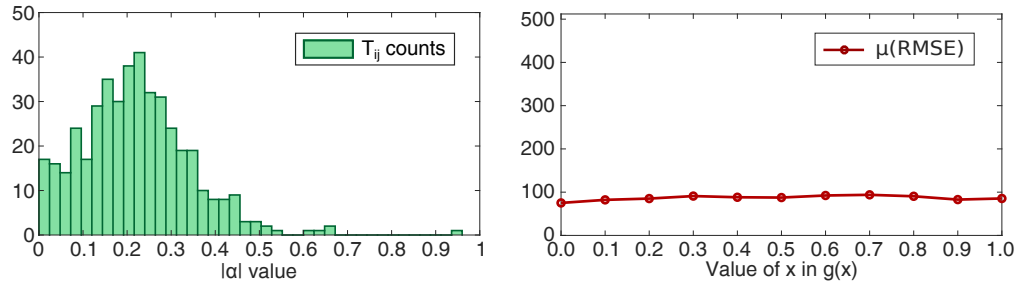


Figure 6.10: *The left figure shows an histogram with the α-values from fitting $g(x)$ to the user clicks. Note how the majority of them have rather low values (below 0.5). The average uncertainty across all workers and $\mathcal{T}_{ij}$ separated by the value of x is displayed on the right. This is measured by the RMSE between the user clicks and the centroids.*

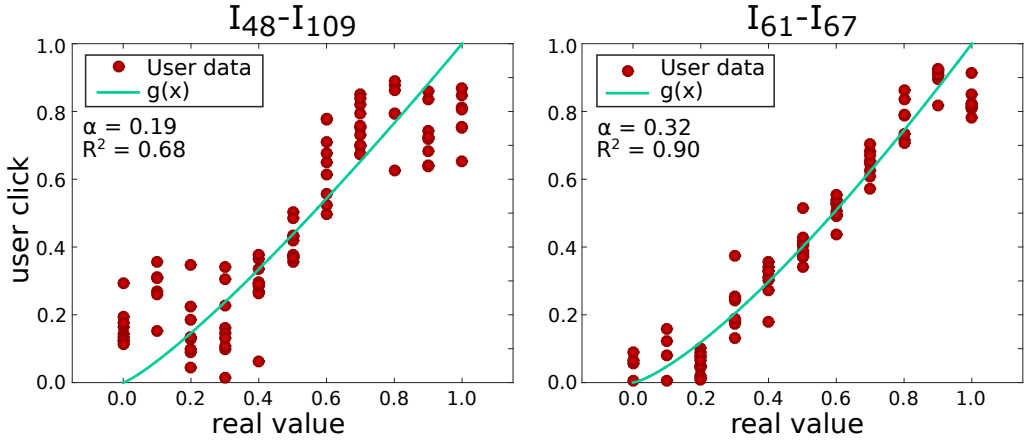


Figure 6.11: *User clicks from the main experimental study in the horizontal axis, for two example image pairs. The plots show how fitting $g(x)$ to the points produces almost-linear functions ($\alpha \approx 0$). In addition, we obtain adequate R^2 values when fitting a linear function.*

ally different textures e.g. (I_{95}, I_{207}), which indicates that the interpolants could be located more accurately. In contrast, the interpolants were less reliably located for similar materials pairs e.g. (I_{95}, I_{209}), as the visual differences at neighboring locations along the sequence were hardly noticeable.

By relying on perceptually linear interpolation stimuli, the user performance in locating certain patches along these interpolation strips is expected to follow measurement scale characteristics. Indeed, the *level of measurement* [Ste46] for a concrete pair is intimately related to the un-

certainty with which users perform the task, as it can be observed when inspecting the positions of the centroids in Figure 6.9. For small overall RMSE values, the texture sequence defines an almost accurate interval-like scale, whereas for large RMSE values this scale becomes less precise.

Assuming that the precision of the linearized images in representing an interval scale is associated with the accuracy of approximating the user ratings with a linear function, we fitted $g(x)$ to the data points collected for each $T_{ij} \in \mathcal{T}_{linear}$. Here, a value of $\alpha \approx 0$ would result in an almost-linear function $g(x) \approx x$. In general, the obtained results indicate that the user responses are well approximated by functions close to linear, since nearly all $g(x)$ present α-values that approach zero (see Figure 6.10), with an average of 0.21 across $\mathcal{T}_{linear}$, hence confirming our expectations. Figure 6.11 shows the user clicks and fitted functions for the two texture sequences shown in Figure 6.7. As illustrated by these plots, fitting $g(x)$ to the data points produces almost-linear functions. In addition, a high coefficient of determination (R^2) when fitting $f(x) = x$ to the user ratings would indicate that the variability in the data is appropriately explained by a linear model. In fact, (R^2) shows acceptable values, i.e. an average of $R^2 = 0.724$ is obtained across all T_{ij}. Both indicators support the approximation of an interval scale behavior for most of the texture combinations.

Furthermore, we examined the average RMSE across all pairs and users, separated by the x values of the intermediate patches (see Figure 6.10). According to the figure, the errors are stable across the values of x in the transition, confirming our belief that the subjects' uncertainty is not affected along the scale. The robustness of the measuring scale was further evaluated by comparing the matrices obtained when taking into account either the user data coming from original stimuli or the data from horizontally-reversed stimuli only. Calculating the RMSE between these two matrices produces a relatively low error value (0.138), implying that the experimental results are not significantly affected by reversing the measurement scale.

Finally, we also analyzed whether the perceptual metric derived from our studies satisfies the requirements of a metric. In fact, the obtained distance matrix $D_{\mathcal{I}}$ as shown in Figure 6.12, follows the conditions of non-negativity ($d_{i,j} > 0 \ \forall \ i \neq j$) and symmetry ($d_{i,j} = d_{j,i}$). Besides, the triangle inequality $d_{ij} + d_{jk} \geq d_{ik}$ is satisfied for 99.68% of the triplets i, j, k. However, it does not fulfill the condition of identity of indiscernibles ($d_{i,j} = 0 \ \forall \ i = j$), since our method would yield distances $d_{i,j} = \epsilon$ where $\epsilon > 0$ for identical textures. To solve this limitation, we simply set the elements in the diagonal to zero.

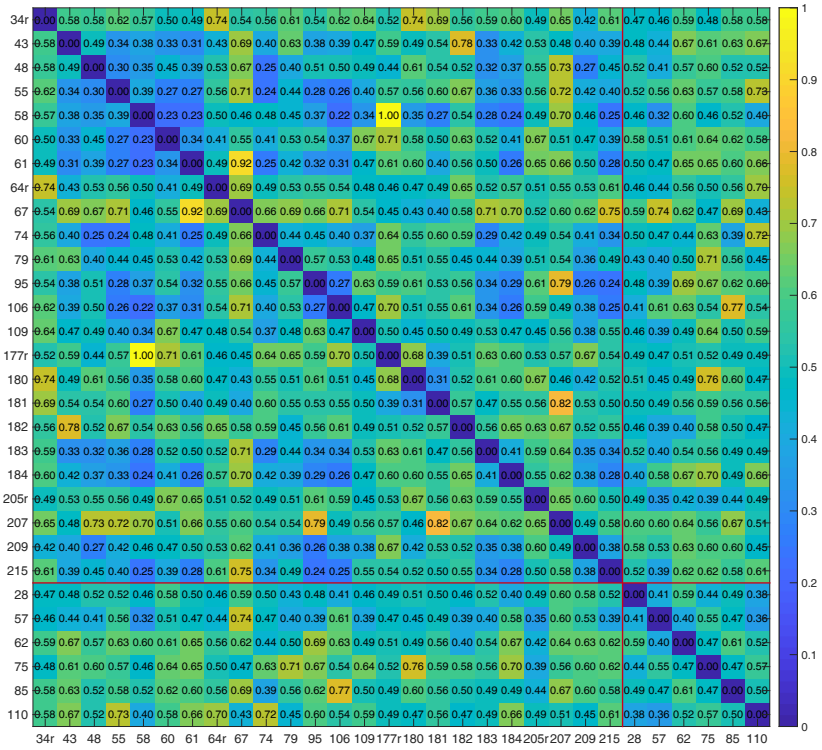


Figure 6.12: *Normalized dissimilarity matrix $D_{\mathcal{I}}$ resulting from the user judgments in the perceptual similarity experiment. The red lines separate the samples obtained from non-fabric samples.*

6.5.2 Analysis of the perceptual texture space

Subspace transformation techniques have been often employed to reveal the underlying dimensions of datasets [GF01, LDC+15, LQD+16, RL96, RFS+98, WAKB09]. Although linear approaches like principal component analysis (PCA) and classical multidimensional scaling (MDS) have been widely used in early texture perception studies, they lack the capability to unveil the non-linear structures that are very likely to be hidden in the data coming from user experiments. In contrast, isometric feature mapping (Isomap) [TdSL00] is a non-linear method which computes the distance between points based on a weighed graph, which makes it a powerful technique when the PSM is sparse. Since we chose to collect user ratings for all possible pairs of textures in $\mathcal{I}$, we obtain a dense distance matrix $D_{\mathcal{I}}$ by design. Hence, we opted for applying non-classical non-metric MDS to $D_{\mathcal{I}}$, which attempts to estimate a subdimensional embedding for the data while maintaining the ranks of the dissimilarities.

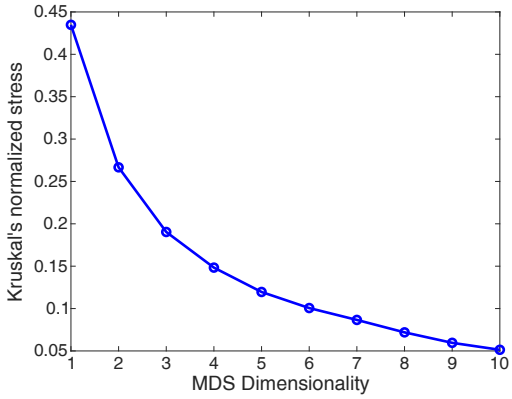


Figure 6.13: *Kruskal's normalized stress for dimensions one to ten. Although there is no obvious choice for the intrinsic dimensionality, three dimensions seem to be a reasonable choice.*

Firstly, in order to analyze the intrinsic dimensionality of the underlying PTS, we computed Kruskal's normalized stress measure [Kru64] for a number of dimensions up to ten (see Figure 6.13). The number of dimensions can be then estimated by looking at the *elbow* where the graph's strong slope ceases and the remaining stress values even out. In our case, there is no obvious decision as this elbow is not evident in the plot, although three dimensions seemed a reasonable and easy-to-visualize choice to represent the dimensionality of the data, with a fair Kruskal's stress value (0.19).

Figure 6.14 presents the output of applying non-classical non-metric MDS to our data in the XY, XZ and YZ plane, where X, Y and Z represent the three main dimensions of the perceptual space respectively. The distribution of the projected samples along the axes in the PTS indicates the similarity or dissimilarity of the textures according to our metric respectively. As it can be observed, the general distinctness between samples is well-preserved by our model, as textures that have comparable appearances under visual inspection lie nearby in the computed perceptual space, e.g. (I_{180}, I_{181}), (I_{95}, I_{183}), (I_{61}, I_{106}) or (I_{209}, I_{215}). On the other hand, pairs with large visual differences such as (I_{61}, I_{67}), (I_{106}, I_{177}) or (I_{85}, I_{207}) are situated in opposite sides of the axes that conform the main dimensions in the PTS. This observation is in accordance with the initial hypothesis regarding the close relationship between user variance and perceived similarity between textures.

When considering the non-fabric texture samples, it becomes apparent that most of them have relatively large perceptual distance values with

respect to the majority of the remaining fabric textures. As a consequence, those whose structure is largely dissimilar to the rest of the samples (i.e. I_{62}, I_{75}, I_{85}, I_{110}) are located in outer regions of the sub-dimensional space, while the samples I_{28} and I_{57} are situated closer to the bulk of the sample points. Even though the distinctness between the fabric and non-fabric materials is preserved in the PTS to a great extent, our metric did not fully capture some obvious texture differences of certain pairs such as (I_{75}, I_{177}), (I_{28}, I_{57}) or (I_{110}, I_{67}). A possible reason for this behavior may be the fact that, despite their visual differences, these pairs of textures share some characteristics (e.g. scale of distinctive features or distribution of the gradients) that grant them a comparable synthesizeability in the IM algorithm.

Moreover, we investigated a possible interpretation for the perceptual dimensions based on the attributes in $F_{\mathcal{I}}$. For this purpose, we followed previous investigations and calculated the correlation coefficient between the average feature ratings for each texture sample (the columns in $F_{\mathcal{I}}$) and their respective coordinates in the three-dimensional space. These coefficients are incorporated in Figure 6.14, and are depicted with greater detail in Figure 6.15. As it can be observed, each dimension seems to correspond to a combination of features, where *Dimension #1* shows a strong correlation with the structural complexity, feature density, contrast and granularity. *Dimension #2* corresponds to the qualities describing randomness, roughness and coarseness. Lastly, *Dimension #3* is dominated by the randomness, uniformity, repetitiveness, directionality, local orientation and regularity. Interestingly, despite the fact that the stimuli tested in our experiments differ, the arrangement of salient features in each dimension resemble those obtained in similar earlier investigations [LDC+15, RL96], with variations in their weight and location in the PTS.

6.6 Discussion

In this paper, a novel formulation for measuring the mutual perceptual similarity or distance between textures is presented, which builds upon the concept of just-noticeable differences and patch-based texture synthesis techniques. Thereby, we exploit the hypothesis that, given a synthesized interpolated sequence between two images, the confidence with which subjects are able to locate intermediate specimens in the continuum will correlate with the perceived similarity of the corresponding textures. To successfully apply this insight, we proposed a perceptual linearization of the texture interpolation results that otherwise only represent a linear

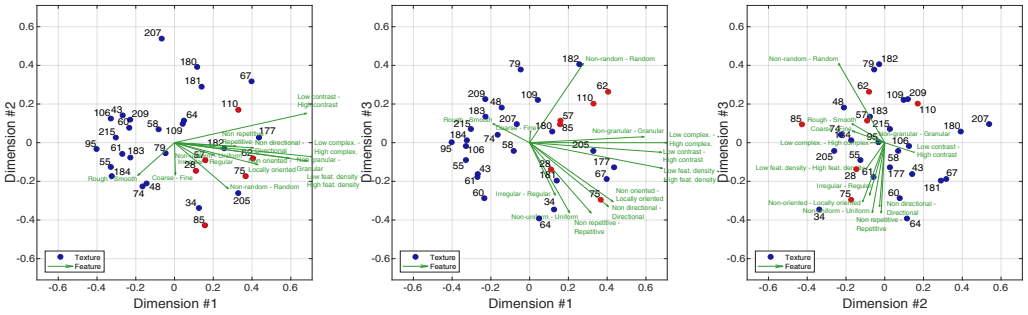


Figure 6.14: *The three first dimensions resulting from applying non-classical, non-metric MDS to the dissimilarity matrix $D_{\mathcal{I}}$. The blue points represent the projected positions of the fabric material samples, while the red points illustrate the non-fabric subset. The correlation coefficients between ($F_{\mathcal{I}}$) and the main perceptual dimensions are represented with green arrows.*

	Low contrast - High contrast	Non-repetitive - Repetitive	Non-granular - Granular	Non-random - Random	Rough - Smooth	Low feat. density - High feat. density	Non-directional - directional	Low complex. - High complex	Coarse - Fine	Irregular - Regular	Non oriented - locally oriented	Non-uniform - Uniform
X	**0.68**	0.32	**0.59**	0.28	-0.19	**0.69**	0.41	**0.72**	-0.00	0.13	0.43	0.21
Y	0.15	-0.02	-0.08	**-0.24**	**-0.17**	-0.09	-0.03	-0.07	**-0.17**	-0.08	-0.11	-0.06
Z	0.05	**-0.36**	0.03	**0.41**	0.10	-0.13	**-0.33**	0.00	0.06	**-0.27**	**-0.31**	**-0.36**

Figure 6.15: *Correlation coefficients between the feature ratings ($F_{\mathcal{I}}$) and the three main dimensions of the PTS. The larger values for each dimension are highlighted in red.*

transition in the space of the used computational features. This perceptual linearization produces a consistent quantitative scale. To validate our approach, we have provided evidence that our methodology yields a meaningful, low-dimensional and perceptually uniform texture space, whose underlying interpretation of the main dimensions with respect to different textural features is in line with earlier research.

In order to generate the interpolated image strips, we made use of a fully automatic, state-of-the-art texture synthesis algorithm (i.e. Image Melding [DSB+12]) which represented a great compromise between time performance and visual quality. However, this algorithm is less reliable when the structural content of the textures is quite diverse, and consequently, the outcome cannot be used as a scale to measure texture

similarity. The perceptual linearization step resolves the problem to a large extent. The soundness of the scale is further confirmed by fitting an almost-linear function to the user ratings and additionally by observing the average RMSE for all pairwise combinations. However, even after the linearization process, some texture pairs still present relatively abrupt transitions, suggesting that the interpolation between rather dissimilar samples may not always be meaningful. On the other hand, we also observed a reduced number combinations between visually different textures that however shared comparable feature size and produced smooth transitions. This resulted in mistakenly predicting small perceived distances for otherwise rather dissimilar textures. More recent developments in texture synthesis have employed neural networks trained for object recognition in order to transfer styles between natural or artistic images [GEB16], as well as to synthesize interpolated images by means of a simple linear interpolation of deep convolutional features [UGP+17]. Still, these methods are not optimized for preserving the details in the appearance of material textures and the derived images are still noisy, containing artifacts produced by the neural networks.

By and large, our methodology intends to alleviate some of the obstacles posed by earlier approaches when gathering image/texture similarity data, namely the parallelizability and duration of the perceptual task itself, the extensibility to new sample points and the sparsity of the resulting matrix. In order to additionally account for the quality of the derived perceptual distances we exploited the noise inherent to such measurements into our benefit, instead of directly asking subjects for similarity measurements. As one of our goals is to study fine-grained similarities between fabric textures, we opted for collecting all the possible pairwise similarities in our database. With this design, the complexity of our solution quickly increases with the size of the database ($O(n^2)$), as adding a new material k involves conducting an initial linearization plus additional experiments for all possible combinations between k and the existing data set. Still, our proposed metric can be equally utilized to collect similarities for an incomplete set of pairs, hence resulting in a sparse distance matrix. In this case, alternative dimensionality reduction techniques which can deal with sparsity, like Isomap [TdSL00] or the MDS methodology proposed by Wills et al. [WAKB09] would have to be employed.

Admittedly, our experimental pipeline could largely benefit from the utilization of regression techniques combined with deep image features from neural networks, in order to automatically compute both the linearization parameter (α) and the perceptual distances ($d_{ij} \in D_{\mathcal{I}}$) between pairs of textures. Nonetheless, for this automated method to be tested

with a minimum reliability, a far larger database of material samples would need to be collected, for these methods require large amounts of training data to make accurate predictions. Thus, a future avenue of research may include the acquisition of a larger, meaningful database of material images from retailing and fashion stores, captured under controlled lighting and viewing conditions. Perceptual distances using our methodology would be then computed and used as the ground-truth for the training of a predictive model based on deep image features in order to learn the perceived texture similarity, in accordance with the approach by Lou et al. [LQD+16]. The examination of the potential results would indicate to what extent fine-grained perceptual distances can be predicted by dense computational features. Finally, the arrangement of the data points in Figure 6.11 (left) has a noticeable sigmoidal (s-shaped) residual. Higher-order linearization could be incorporated in following iterations of our methodology. This would probably produce more uniform transitions between the original exemplars and hence more reliable measurements.

6.7 Acknowledgments

This work was partially funded by the X-Rite Graduate School on Digital Material Appearance at the University of Bonn. We further thank Julian Iseringhausen and Douglas Cunningham for the valuable discussion.

Part IV

Closure

CHAPTER 7

Summary, Discussion and Future Work

Digital materials are present in a wide range of every-day situations. However, despite the extraordinary material reproduction quality achieved by the current technology, this is still distant from accurate photo-realism. By investigating how humans perceive material appearance in digital environments we can narrow the perceptual gap between real materials and their digital counterparts.

This final chapter begins presenting an outlook of the contributions provided in the scope of this thesis (Section 7.1). This is followed by a thorough discussion of the results introduced in the previous chapters (Section 7.2), including the possible limitations of the proposed techniques and a landscape of potential future research concerning the given topics. To conclude, several final remarks are given (Section 7.3).

7.1 Summary and contributions

This manuscript presents several investigations relevant to the field of digital material appearance and its perception. Such topics together with a summary of the respective contributions are listed in the following lines:

1. We explored the interplay between different sensing modalities (vision and hearing) in the perception of material qualities (Chapter 3). As the main finding, we discovered that the tapping and rubbing sounds employed are able to facilitate the estimation of tactile material qualities and that such estimation process could be distorted

by the modification of the audio cues.

2. We then studied the effects of a material sonification system for tactile devices in the assessment of physical and emotional material qualities (Chapter 4). For this investigation we utilized audio feedback produced only when rubbing the material sample with the fingertip. Despite not affecting the overall perception of the material, we determine that the sound cues employed and consequently the proposed sonification system are not able to complement the visual information. According to this insights, relevant research directions in interactive material sonification are given.

3. Later, we examined the ability of state-of-the-art digital material models in transmitting material appearance in comparison to simple photographs and real-world materials (Chapter 5). Our major finding is the verification of an existing gap between digitized materials and their physical counterparts. Furthermore, we hypothesized and confirmed that the studied appearance model suffers from a significant loss of information at grazing angles, where geometric cues are not adequately represented.

4. Lastly, one of the shortcomings of psychophysical studies in material perception is the lack of suitable metrics to quantify certain aspects of material appearance. For this reason, we considered the investigation of perceptually plausible metrics for materials. To simplify the large dimensionality inherent to such problem, we focused our investigation on textures obtained from real material samples and developed a methodology to compute the perceived pairwise similarity between textures (Chapter 6). Our technique is able to mitigate some of the obstacles from previous data collection methods and produces a meaningful, low-dimensional and perceptually uniform space of material textures.

In the next section, the set of contributions stated above will be discussed in greater detail.

7.2 Discussion and future work

This section provides extended discussion to the topics that have been covered throughout the previous chapters of this dissertation.

7.2.1 Multimodal perception of materials

The findings in the scope of multimodal material perception confirm touch-related material sound as a relevant actor when perceiving material attributes, specially those characteristics associated to the haptic experience. However, the similar effect was not detected in a follow-up investigation, in which we considered interactive sound generated by a sonification system. We contemplate several reasons for this unanticipated outcome, which mainly concern the stimuli utilized in each experiment and the specific interactions employed to generate touch-related material sound.

For the studies proposed in Chapter 3, we recorded sound from a sequence of material interactions, which included rubbing and tapping gestures. Likewise, samples from three diverse material classes were employed in our studies: leathers, papers and fabrics. However, during the following investigation described in Chapter 4, tapping sounds and paper stimuli were disregarded. Such decision was made in order to reduce the range of synthesized sounds that the sonification algorithm has to deliver. Furthermore, the visual stimuli (photographs) in both studies also differ, as in the follow-up experiment the borders from the material samples, which provide additional volumetric information, had been digitally removed. As a result of these decisions, the ability to communicate material appearance from the visual stimuli presented in Chapter 4 was reduced significantly. The most notorious case is the "hard–soft" dimension, where the inter-participant correlation values drop down by almost 0.3 between the audiovisual conditions of the two mentioned investigations (see Figures 3.5 and 4.4). Presumably, the information concerning the hardness of the material was largely contained in the impact sounds from tapping interactions, and not so much in the audio cues resulting from scrapping/rubbing interactions. Likewise, the full-modal condition presents a similar drop on the correlation coefficient for this concrete dimension ("hard–soft") between the two experiments, despite the fact that the users had available the actual material samples to interact. In this case, we hypothesize that the presence of paper samples, having rather different characteristics as leathers or fabrics, established an upper bound when assessing the hardness of the material. In the absence of such anchoring, participants' assessments for this concrete quality pair were less reliable. The former assumptions would also apply, to some extent, to the relatively low (or lower) agreement found for other investigated dimensions (i.e. "thick–thin", "stiff–flexible" or "natural–synthetic")in the second experiment, as it is implied by the strong relationship between

the ratings for the mentioned qualities (see Figure 4.6).

Following research regarding the audiovisual perception of materials should not disregard the value of impact/collision sounds when discriminating between materials categories or when evaluating their perceptual qualities, in accordance with previous investigations [AR01, CAKP98, GM06, KPK00, LH12, FGM+14, MIWH15], all of them including some sort of striking sounds in their experimental studies. Nonetheless, an investigation from Giordano and McAdams [GM06] questions the capacity of humans to discriminate between samples within the same class based on auditory cues. In light of these findings, a fruitful avenue of future research could evaluate the impact of employing high-resolution audio and video from more complex audiovisual interactions in the perception of material qualities and material categorization. In this regard, specific gestures like two-finger stroking or sample scrunching have been documented to be the most repeated when interacting with textile materials [AOP+13].

All in all, the conclusions arising from the analysis of the respective experiments would only apply to the collection of materials and the set of perceptual qualities considered in these investigations. The selection of pairs of material attributes was initially accomplished by choosing those most regularly employed in related literature (see Section 3.3.1). In accordance with this choice, only 4 principal components are enough to represent 95.1% and 94.8% of the variance from the full-modal conditions, respectively for the two mentioned investigations. Meanwhile, analogous studies with a similar or higher number of perceptual dimensions present a much more moderate eigenvalue decay. For example, in the 42-dimensional space from Fleming et al. [FWG13], seven PCs could only explain 50% of the variance, roughly the same amount as only the three main factors from by Fujisaki et al. [FTK15], despite the fact that the perceptual space utilized was much wider (69 dimensions). Analogously, Rao and Lohse [RL96] found no single PC that explained more than 10% of their 12 visual dimensions. The latter highlights the fact that the results from similar studies in perception of material qualities are highly dependent on the selected attribute set. For that reason, the development of a standard space of perceptual or computational features to describe material appearance remains essential for future investigations in material perception.

7.2.2 Digital transmission of subjective material appearance

This manuscript also examines the ability of an advanced digital material model (Spatially Varying BRDF) to transmit material information (represented by a rigorously selected set of material qualities) in comparison to simple photographs, rendered motion videos and real materials. The conducted analysis revealed the current perceptual gap between digital models and photos, which is mostly due to the geometrical abstractions imposed by the SVBRDF model.

In general, none of the evaluated stimuli (photos, static renderings and motion renderings) proved to be fully capable of transmitting advanced material properties, as significant differences were found between the user ratings for the full-modal condition (physical materials) and the rest of the models, as observed in Figure 5.6. These effects are especially noticeable in the perception of those qualities categorized as tactile, being the "thick-thin" and "stiff-flexible" the dimensions which are more poorly represented by digital models. Likewise, an important number of dissimilarities emerge for the perceived transparency, which is to be expected, since the SVBRDF measurements are not able to represent this particular characteristic. When considering the existing meaningful effects between digitized materials and photographs, the largest misperceptions arise again for tactile qualities together with the shininess and brightness dimensions. Undoubtedly, the evaluated digitized materials are not able to depict some of the volumetric and reflectance information revealed by the photographs' distinctive border regions. This insight is confirmed by the fact that digitally removing such regions from the images greatly reduces their the ability to transmit the set of qualities. Finally, we did not find a significant impact of motion in the perception of material qualities. Still, further experiments would need to be conducted on this matter, where the concrete attributes to be studied are properly isolated as Wend et al. [WFEM10] did for motion gloss.

One of the points in question regarding the analysis of the experimental data concerns to which statistical tests should be employed to analyze the effects between the experimental conditions. In general, parametric tests are utilized when the studied population fulfills several assumptions and when the variable of interest is computed on an interval scale. Yet, there is some controversy about whether the variables measured by Likert scales (as those employed in our studies) are interval or ordinal [CW11]. To avoid any risk, we opted for employing non-parametric tests in our experimental analysis to uncover significant effects between the

tested conditions.

Once again, the evidences concerning the transmission of material appearance found in this investigation apply only to the considered stimuli. In order to extrapolate these findings to additional material classes (wood, stone, plastic), supplementary experiments would be need to be conducted. Besides, some aspects concerning the digitized materials (i.e. small-scale differences in illumination, geometry and scale) could have limited their expressiveness. Additionally, one would need to take into account the suitability of SVBRDF representations to depict each individual material category. In this regard, the two classes examined in the scope of our research had unequal performance. While leathers were mostly well represented by this model, some digital fabrics, specially those from samples exhibiting protruding fibers and fluffy appearance, were considered somewhat unrealistic (see Figures 5.11 and 5.12).

By and large, this investigation supports the insight that a great deal of the appearance of materials is condensed in very specific regions, concretely those areas visible when observing the sample under grazing angles. This conclusion could be further explored through a experimental setup in which eye gaze tracking is employed to study the regions of visual attention when observing digital materials, which remains a promising direction for future work. Moreover, due to the relatively low angular resolution provided by the employed acquisition device and the limitations inherent to the model, such fine effects of appearance could not be fully captured. Future research strands could contemplate higher order appearance models. Concretely BTF and BSSRDF representations are able to depict additional volumetric and reflectance material characteristics (self-occlusion, intereflections or subsurface scattering) with finer detail. In turn, their acquisition process becomes more complex, significantly lengthier and presents additional storage issues that would need to be addressed.

7.2.3 Perceptual similarity metrics

Lastly, we developed a perceptually-based metric for similarity or dissimilarity between materials, where these materials are represented by textures from real samples acquired under controlled viewing and illumination conditions. In our experiments, we exploited the public ALOT database [UA09] to collect a set of 30 textures from fabrics and other alternative material classes. The election of material textures over surface reflectance models arises from the idea of estimating pairwise distances based on observable perceptual features, and not reparameterizing com-

putational features in terms of a given model (e.g. Ward BRDF). Besides, the latter direction has been already covered in greater or lesser extent by previous research efforts to compute a subdimensional space for gloss i.e. [PFG00, WAKB09] or transparency i.e. [CWFS07], in order to find a suitable perceptual space for BRDF material editing i.e. [SGM+16] and to investigate the perceptual effects of BTF filtering i.e. [JWD+14]. Furthermore, textures from real samples is the predominant way of displaying materials in online retailing stores (see Figure 1.1). Hence, the potential findings would have a more direct application in the domain of e-commerce and could serve as the basis for more advanced metrics.

The novelty of this investigation lies in the methodology employed to compute such perceptual similarities. Previous methods either directly queried the similarity between textures to the participants or relied on time-consuming, not parallelizable experimental processes which did not allow the later incorporation of unseen points. Our technique avoids these problems with an indirect approach in which similarities are considered as proportional to the inherent noise resulting from a localization task. Besides, the proposed approach is parallelizable, easy to scale and avoids additional human issues (e.g. fatigue or boredom). The fact that different methodologies, like the one proposed in this thesis and those from previous studies, lead to perceptual texture spaces (PTSs) whose main dimensions largely correlate with each other supports the validity of our method. In fact, the X dimension in our experiments seems analogous to the Z dimension in those from Rao and Lohse [RL96] and Lou et al. [LQD+16], while our third dimension (Z) is largely comparable to dimensions X and Y in the former studies, respectively.

One of the drawbacks of the suggested computation for the perceptual similarity between textures is the preliminary linearization step, necessary to carry out the experiment on a consistent interval scale. For this reason, this procedure is composed by two individual steps, thus, complicating the inclusion of unseen points in the PTS. Even after this procedure, few interpolated strips still exhibited non linear characteristics, translated into transitions between textures which still exhibit slightly larger α-values when fitting $g(x)$ to the user clicks (see Figure 6.10 left). In order to address this issue, one feasible solution would be the utilization of higher-order linearization functions instead of the employed one (see Equation 6.1). Another aspect that limits the practical applicability of the proposed method to large databases is its quadratic complexity ($O(n^2)$). However, our approach can be equally utilized to gather similarities for an incomplete set of pairs, resulting therefore in a sparse matrix. In the context of this investigation, we opted for the computation of a

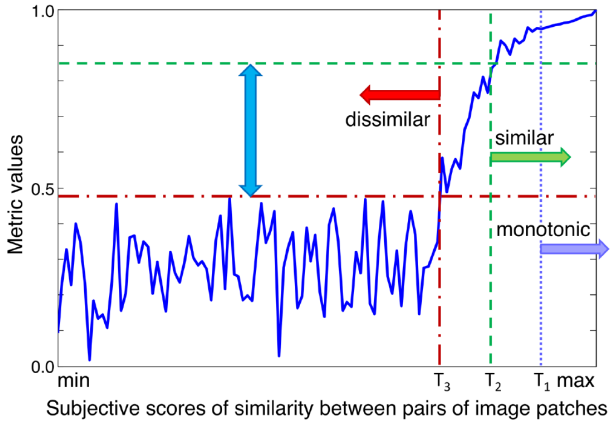


Figure 7.1: *Conceptual plot of desired metric performance: metric values vs. subjective similarity scores. From Zujovic et al. [ZPN+15].*

fully-connected matrix in order to sample densely the fine-grain relationships between textures from the same category (fabric).

Precisely, previous research has determined that humans are consistent while judging similarities between almost identical textures (Figure 7.1, right of T_1) or while discriminating similar against dissimilar textures (left of T_3) [ZPN+15]. Our research attempts to provide a metric for those remaining textures in the illustrated similarity scale (between T_1 and T_3) which are only comparable in some respect (e.g. directionality, scale) but different in another (e.g. regularity, contrast) and whose subject-to-subject similarity agreement tends to be poor. Still, the present methodology is constrained to those intermediate-to-fine texture distances and cannot provide a monotonic relationship between measured and perceived similarity for every possible pair of textures in the given scale.

In practice, we would like to have at our disposal a fully computational way to calculate perceptual similarities between materials that does not require additional psychophyisical studies for pairs of points out-of-sample. To that end, there is a boundless number of regression algorithms which, in combination with deep image features, could be used to predict image similarity/dissimilarity values. Nonetheless, the reliability of such algorithms depends on the training process in a considerably large database of material textures annotated with ground-truth pairwise similarities.

7.3 Final remarks

In summary, with the compilation of research introduced in this manuscript, we have provided novel insights in how humans perceive real and digital materials and how the latter are communicated through digital devices. We believe that the given findings have direct applications in the scope of online commerce, concretely in the development of more effective user interfaces which are consistent with human perception and take into account the multimodal nature of the interactions with real material samples. Additional areas of research that could benefit from the results presented in this thesis regarding the transmission of digital material information and their multisensory perception are product design, virtual reality, medical imaging, physical rehabilitation or sensory substitution for the impaired.

To conclude, the increasing range of interactions with digital commodities and the rise of devices and applications linked to Internet of Things (IoT) demands innovative and more efficient ways to depict digital products. Indeed, the perception materials from which such products/objects are made have a decisive role in these increasingly complex process. By making steps towards unveiling how humans perceive real and virtual materials, we would greatly contribute to the massive task of developing more engaging digital environments and user-oriented interfaces to be used in a large spectrum of applications.

Bibliography

[AB91] E. Adelson and J. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991.

[Ade01] E. Adelson. On seeing stuff: the perception of materials by humans and machines. *Proc. SPIE*, 4299:1–12, 2001.

[AG15] J. Ackermann and M. Goesele. *A Survey of Photometric Stereo Techniques*. Now Foundations and Trends, 2015.

[AJM12] S. An, D. James, and S. Marschner. Motion-driven concatenative synthesis of cloth sounds. *ACM Trans. on Graphics*, 31(4):102:1–102:10, July 2012.

[And11] B. Anderson. Visual perception of materials and surfaces. *Current Biology*, 21(24):R978 – R983, 2011.

[AOP+13] D. Atkinson, P. Orzechowski, B. Petreca, N. Bianchi-Berthouze, P. Watkins, S. Baurley, S. Padilla, and M. Chantler. Tactile perceptions of digital textiles: A design research approach. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '13, pages 1669–1678, New York, NY, USA, 2013. ACM.

[AP96] E. Adelson and A. Pentland. The perception of shading and reflectance. In *Perception As Bayesian Inference*, pages 409–423. Cambridge University Press, New York, NY, USA, 1996.

[AR01] F. Avanzini and D. Rocchesso. Controlling material properties in physical models of sounding objects. In *Proc. of the International Computer Music Conf., La Habana*, pages 91–94, 2001.

[BA02] S. Barrass and M. Adcock. Interactive granular synthesis of haptic contact sounds. In *22nd Audio Engineering Society Conf.: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, Jun 2002.

[Bal08] B. Balas. Attentive texture similarity as a categorization task: Comparing texture synthesis models. *Pattern Recognition*, 41(3):972–982, Mar. 2008.

[BBS14] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Trans. on Graphics (SIGGRAPH)*, 33(4), 2014.

[BCFW08] D. Bartz, D. Cunningham, J. Fischer, and C. Wallraven. The Role of Perception for Computer Graphics. In *Eurographics 2008 - State of the Art Reports*. The Eurographics Association, 2008.

[Bli77] J. Blinn. Models of light reflection for computer synthesized pictures. *SIGGRAPH Comput. Graph.*, 11(2):192–198, July 1977.

[Bro66] P. Brodatz. *Textures: a photographic album for artists and designers*. Dover Pubns, 1966.

[BS+10] N. Bonneel, C. Suied, I. Viaud-Delmon, and G. Drettakis. Bimodal perception of audio-visual material properties for virtual environments. *ACM Trans. on Applied Perception*, 7(1):1:1–1:16, Jan. 2010.

[BSFG09] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), Aug. 2009.

[BSGA10] C. Barnes, E. Shechtman, D. Goldman, and A.Finkelstein. The generalized PatchMatch correspondence algorithm. In *European Conf. on Computer Vision*, Sept. 2010.

[BW95] T. Bijmolt and M. Wedel. The effects of alternative methods of collecting similarity data for multidimensional scaling. *Int. Journal of Research in Marketing*, 12(4):363–371, 1995.

[BWG13] E. Baumgartner, C. Wiebel, and K. Gegenfurtner. Visual and haptic representations of material properties. *Multisensory Research*, 26(5):429–455, 2013.

[BXBF13] K. Bouman, B. Xiao, P. Battaglia, and W. Freeman. Estimating the material properties of fabric from video. In *Computer Vision (ICCV), 2013 IEEE International Conf. on*, pages 1984–1991. IEEE, 2013.

[CAKP98] C. Carello, K. Anderson, and A. Kunkler-Peck. Perception of object length by sound. *Psychological Science*, 9(3):211–214, 1998.

[Cha94] M. Chantler. *The effect of variation in illuminant direction on texture classification*. PhD thesis, Heriot-Watt University, 1994.

[CHN+11] A. Clarke, F. Halley, A. Newell, L. Griffin, and M. Chantler. Perceptual similarity: A texture challenge. In *Proc. of the British Machine Vision Conf.*, pages 120.1–120.0. BMVA Press, 2011.

[CJSC03] A. Citrin, D. S. Jr., E. Spangenberg, and M. Clark. Consumer need for tactile input: An internet retailing challenge. *Journal of Business Research*, 56(11):915 – 922, 2003.

[CLG+16] G. Cirio, D. Li, E. Grinspun, M. Otaduy, and C. Zheng. Crumpling sound synthesis. *ACM Trans. on Graphics*, 35(6):181:1–181:11, Nov. 2016.

[CRST+15] S. Creem-Regehr, J. Stefanucci, W. Thompson, N. Nash, and M. McCardell. Egocentric distance perception in the oculus rift (dk2). In *Proc. of the ACM SIGGRAPH Symposium on Applied Perception*, SAP '15, pages 47–50, New York, NY, USA, 2015. ACM.

[CW11] D. Cunningham and C. Wallraven. *Experimental Design: From User Studies to Psychophysics*. A. K. Peters, Ltd., Natick, MA, USA, 1st edition, 2011.

[CWFS07] D. Cunningham, C. Wallraven, R. Fleming, and W. Strasser. Perceptual reparameterization of material properties. In *Computational Aesthetics '07*, pages 89–96, 2007.

[Dal93] S. Daly. The visible differences predictor: An algorithm for the assessment of image fidelity. In *Digital Images and Human Vision*, pages 179–206. MIT Press, Cambridge, MA, USA, 1993.

[DFY+11] K. Doerschner, R. Fleming, O. Yilmaz, P. Schrater, B. Hartung, and D. Kersten. Visual motion and the perception of surface material. *Current Biology*, 21(23):2010 – 2016, 2011.

[DGFH16] J. DeGol, M. Golparvar-Fard, and D. Hoiem. Geometry-informed material recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 1554–1562. IEEE, 2016.

[DMC14] X. Dong, T. Methven, and M. Chantler. How well do computational features perceptually rank textures? a comparative evaluation. In *Proc. of Int. Conf. on Multimedia Retrieval*, ICMR '14, pages 281–288, New York, NY, USA, 2014. ACM.

[DSB+12] S. Darabi, E. Shechtman, C. Barnes, D. Goldman, and P. Sen. Image Melding: Combining inconsistent images using patch-based synthesis. *ACM Trans. on Graphics*, 31(4):82:1–82:10, 2012.

[DvG+99] K. Dana, B. van Ginneken, S. Nayar, and J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans. on Graphics*, 18(1):1–34, Jan. 1999.

[Efr82] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*, volume 38. SIAM, 1982.

[EL99] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the Seventh IEEE International Conf. on Computer Vision*, volume 2, pages 1033–1038 vol.2, 1999.

[EL12] M. Ellens and F. Lamy. From color to appearance in the real world. In *Proc. SPIE*, volume 8291, pages 82910B–82910B–8, 2012.

[ESG14] R. Etzi, C. Spence, and A. Gallace. Textures that we like to touch: An experimental study of aesthetic preferences for tactile stimuli. *Consciousness and Cognition*, 29:178–188, 2014.

[ET86] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, pages 54–75, 1986.

[Fec60] G. Fechner. *Elemente der Psychophysik*. Number v. 1 in Elemente der Psychophysik. Breitkopf und Härtel, 1860.

[FGM+14] W. Fujisaki, N. Goda, I. Motoyoshi, H. Komatsu, and S. Nishida. Audiovisual integration in the human perception of materials. *Journal of Vision*, 14(4), 2014.

[FKN80] H. Fuchs, Z. Kedem, and B. Naylor. On visible surface generation by a priori tree structures. *SIGGRAPH Comput. Graph.*, 14(3):124–133, July 1980.

[Fle14] R. Fleming. Visual perception of materials and their properties. *Vision Research*, 94:62 – 75, 2014.

[FTK15] W. Fujisaki, M. Tokita, and K. Kariya. Perception of the material properties of wood based on vision, audition, and touch. *Vision Research*, 109:185 – 200, 2015.

[FVHK16] J. Filip, R. Vávra, M. Havlíček, and M. Krupička. Predicting visual perception of material structure in virtual environments. *Computer Graphics Forum*, 2016.

[FWG13] R. Fleming, C. Wiebel, and K. Gegenfurtner. Perceptual qualities and material classes. *Journal of Vision*, 13(8):9, 2013.

[GCLS02] S. Guest, C. Catmur, D. Lloyd, and C. Spence. Audiotactile interactions in roughness perception. *Experimental Brain Research*, 146(2):161–171, 2002.

[GEB15] L. Gatys, A. Ecker, and M. Bethge. Texture synthesis using convolutional neural networks. In *Proc. of the Int. Conf. on Neural Information Processing Systems*, volume 1 of *NIPS'15*, pages 262–270, Cambridge, MA, USA, 2015. MIT Press.

[GEB16] L. Gatys, A. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *IEEE Conf. on Computer Vision and Pattern Recognition*, CVPR '16, pages 2414–2423, 2016.

[Ges97] G. Gescheider. *Psychophysics: The fundamentals*. Lawrence Erlbaum Associates Publishers, 1997.

[GF01] R. Gurnsey and D. Fleet. Texture space. *Vision Research*, 41(6):745 – 757, 2001.

[Gib79] J. Gibson. *The ecological approach to visual perception*. Houghton, Mifflin and Company, 1979.

[GJMS+14] S. Garrido-Jurado, R. Munoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280 – 2292, 2014.

[GM06] B. Giordano and S. McAdams. 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, 2006.

[GMD10] D. Geisler-Moroder and A. Dür. Bounding the albedo of the ward reflectance model. *High Performance Graphics*, 2010.

[Gow75] J. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, Mar 1975.

[GP79] R. Gregory and H. P. Border locking and the café wall illusion. *Perception*, 8(4):365–380, 1979. PMID: 503767.

[GS90] P. Green and V. Srinivasan. Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54(4):3–19, 1990.

[GTR+06] J. Gu, C. Tu, R. Ramamoorthi, P. Belhumeur, W. Matusik, and S. Nayar. Time-varying surface appearance: Acquisition, modeling and rendering. *ACM Trans. Graph.*, 25(3):762–771, July 2006.

[Hal12] F. Halley. *Perceptually Relevant Browsing Environments for Large Texture Databases*. PhD thesis, Heriot-Watt University, Apr. 2012.

[HF13] M. Haindl and J. Filip. *Visual texture: Accurate material appearance measurement, representation and modeling*. Springer Science & Business Media, 2013.

[HIL+05] D. House, V. Interrante, D. Laidlaw, R. Taylor, and C. Ware. Design and evaluation in visualization research. In *VIS 05. IEEE Visualization.*, pages 705–708, Oct 2005.

[HJ10] C. Hallett and A. Johnston. *Fabric for Fashion: The Swatch Book*. Laurence King Publishing, 2010.

[HJK+13] C. Ho, R. Jones, S. King, L. Murray, and C. Spence. Multisensory augmented reality in the context of a retail clothing application. In *Audio Branding Academy yearbook 2012/2013*, pages 167–174, Germany, 2013. Nomos Publishers.

[HJZ13] A. Hope, M. Jones, and H. Zuo. Sensory perception in materials selection for industrial/product design. *The International Journal of Designed Objects*, 6(3):17–31, 2013.

[HLM06] Y. Ho, M. Landy, and L. Maloney. How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, 6(5):8, 2006.

[Hor89] B. Horn. Shape from shading. chapter Obtaining Shape from Shading Information, pages 123–171. MIT Press, Cambridge, MA, USA, 1989.

[HS13] S. Hossain and S. Serikawa. Texture databases - a comprehensive survey. *Pattern Recognition Letters*, 34(15):2007–2022, Nov. 2013.

[Joann18] Joann fabric and craft stores. http://www.joann.com/fabric/, 2018. Accessed on July 31, 2018. Images reproduced under the 'fair use' provision of the Copyright Law.

[JWD+14] A. Jarabo, H. Wu, J. Dorsey, H. Rushmeier, and D. Gutierrez. Effects of approximate filtering on the appearance of bidirectional texture functions. *IEEE Trans. on Visualization and Computer Graphics*, 20(6):880–892, June 2014.

[Kan79] G. Kanizsa. *Organization in vision: Essays on Gestalt perception*. Praeger Publishers, 1979.

[KEBK05] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra. Texture optimization for example-based synthesis. *ACM Trans. on Graphics*, 24(3):795–802, July 2005.

[KFB10] J. Křivánek, J. Ferwerda, and K. Bala. Effects of global illumination approximations on material appearance. *ACM Trans. Graph.*, 29(4):112:1–112:10, July 2010.

[KGS12] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[KL10] R. Klatzky and S. Lederman. Multisensory texture perception. In *Multisensory Object Perception in the Primate Brain*, pages 211–230. Springer New York, New York, NY, 2010.

[KPK00] R. Klatzky, D. Pai, and E. Krotkov. Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments*, 9(4):399–410, 2000.

[Kru64] J. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, Mar 1964.

[LDC+15] J. Liu, J. Dong, X. Cai, L. Qi, and M. Chantler. Visual perception of procedural textures: Identifying perceptual dimensions and predicting generation models. *PLOS ONE*, 10(6):1–22, 2015.

[LFKW06] H. Li, S. Foo, T. K.E., and S. Westin. Automated three-axis gonioreflectometer for computer graphics applications. *Optical Engineering*, 45:45 – 45 – 11, 2006.

[LG04] M. S. Landy and N. Graham. Visual perception of texture. *The Visual Neurosciences*, 1:1106, 2004.

[LH96] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. of the 23rd Annual Conf. on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 31–42, New York, NY, USA, 1996. ACM.

[LH12] G. Lemaitre and L. Heller. Auditory perception of material is fragile while action is strikingly robust. *Acoustical Society of America Journal*, 131:1337, 2012.

[LQD+16] J. Lou, L. Qi, J. Dong, H. Yu, and G. Zhong. Learning perceptual texture similarity and relative attributes from computational features. In *2016 Int. Joint Conf. on Neural Networks (IJCNN)*, pages 2540–2546, July 2016.

[LWM+16] C. Li, M. Wand, J. Matas, N. Sebe, and M. Welling. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *Computer Vision – ECCV 2016*, pages 702–716. Springer International Publishing, 2016.

[MB10] L. Maloney and D. Brainard. Color and material perception: Achievements and challenges. *Journal of Vision*, 10(9), 2010. doi: 10.1167/10.9.19.

[MC09] A. Misra and P. Cook. Toward synthesized environments: A survey of analisis and synthesis methods for sound designers and composers. In *Proc. of the International Computer Music Conf. (ICMC)*, pages 155–162, San Francisco, CA, USA, 2009. Int. Computer Music Association.

[Mea97] Meastex image texture database. http://www.texturesynthesis.com/meastex/meastex.html, 1997. Accessed on July 31, 2018.

[MIWH15] R. Martín, J. Iseringhausen, M. Weinmann, and M. Hullin. Multimodal perception of material properties. In *ACM SIGGRAPH Symposium on Applied Perception*, SAP '15, pages 33–40, New York, NY, USA, Sept. 2015. ACM.

[MM96] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837–842, Aug 1996.

[MMK+06] J. Meseth, G. Müller, R. Klein, F. Röder, and M. Arnold. Verification of rendering quality from measured btfs. In *Proc. of the 3rd Symposium on Applied Perception in Graphics and Visualization*, APGV '06, pages 127–134, New York, NY, USA, 2006. ACM.

[Mood18] Mood designer fabrics. online fabric store. https://www.moodfabrics.com/, 2018. Accessed on July 31, 2018. Images reproduced under the 'fair use' provision of the Copyright Law.

[MPBM03] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. Graph.*, 22(3):759–769, July 2003.

[MWH17] R. Martín, M. Weinmann, and M. Hullin. Digital transmission of subjective material appearance. *Journal of WSCG*, 25(2):57–66, June 2017.

[MWH18a] R. Martín, M. Weinmann, and M. B. Hullin. Evaluating the effects of material sonification in tactile devices. *arXiv preprint arXiv:1807.01106*, 2018.

[MWH18b] R. Martín, M. Weinmann, and M. B. Hullin. A study of material sonification in touchscreen devices. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, ISS '18, pages 305–310, New York, NY, USA, 2018. ACM.

[MXK+19] R. Martín, M. Xue, R. Klein, M. B. Hullin, and M. Weinmann. Using patch-based image synthesis to measure perceptual texture similarity. *Computers and Graphics*, 81:104 – 116, 2019.

[NRH+77] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometrical considerations and nomenclature for reflectance. National Bureau of Standards Monograph #160, U.S. Dept. of Commerce, 1977.

[OIM+16] A. Owens, P. Isola, J. McDermott, A. Torralba, E. H. Adelson, and W. T. Freeman. Visually indicated sounds. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2405–2413, June 2016.

[ONY13] S. Okamoto, H. Nagano, and Y. Yamada. Psychophysical dimensions of tactile perception of textures. *EEE Trans. on Haptics*, 6(1):81–93, Jan. 2013.

[OPM02] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, Jul 2002.

[Osg52] C. Osgood. The nature and measurement of meaning. *Psychological bulletin*, 49(3):197, 1952.

[PDVG03] D. Picard, C. Dacremont, D. Valentin, and A. Giboreau. Perceptual dimensions of tactile textures. *Acta Psychologica*, 114(2):165 – 184, 2003.

[PFG00] F. Pellacini, J. Ferwerda, and D. Greenberg. Toward a psychophysically-based light reflection model for image synthesis. In *Proc. of SIGGRAPH*, pages 55–64, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[Pho18] Photex database. http://www.macs.hw.ac.uk/texturelab/resources/databases/photex/. Accessed on July 31, 2018.

[Pho75] B. Phong. Illumination for computer generated pictures. *Commun. ACM*, 18(6):311–317, June 1975.

[PN+13] T. Pappas, D. Neuhoff, H. de Ridder, and J. Zujovic. Image analysis: Focus on texture similarity. *Proc. of the IEEE*, 101(9):2044–2057, Sept 2013.

[PS00] J. Portilla and E. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, Oct. 2000.

[RFS+98] B. Rogowitz, T. Frese, J. Smith, C. Bouman, and E. Kalin. Perceptual image similarity experiments. In *Human Vision and Electronic Imaging III*, volume 3299, pages 576–591. Int. Society for Optics and Photonics, 1998.

[RFWB07] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. Visual equivalence: Towards a new standard for image fidelity. *ACM Trans. Graphics*, 26(3), July 2007.

[RL96] A. Rao and G. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, 36(11):1649 – 1669, 1996.

[RSK10] R. Ruiters, R. Schnabel, and R. Klein. Patch-based texture interpolation. *Computer Graphics Forum*, 29(4):1421–1429, 2010.

[Sch94] C. Schlick. An inexpensive brdf model for physically-based rendering. *Computer Graphics Forum*, 13:233–246, 1994.

[SGM+16] A. Serrano, D. Gutierrez, K. M., H. Seidel, and B. Masia. An intuitive control space for material appearance. *ACM Trans. on Graphics*, 35(6):186:1–186:12, Nov. 2016.

[SI06] J. Swan II. Experimental design and analysis for human-subject visualization experiments. In *IEEE InfoVis*, 2006.

[SLRA13] L. Sharan, C. Liu, R. Rosenholtz, and E. Adelson. Recognizing materials using perceptually inspired features. *Int. Journal of Computer Vision*, 103(3):348–371, 2013.

[SMdB+15] H. Steinhausen, R. Martín, D. den Brok, M. Hullin, and R. Klein. Extrapolation of bidirectional texture functions using texture synthesis guided by photometric normals. In *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)*, volume 9398, Feb. 2015.

[SN15] G. Schwartz and K. Nishino. Automatically discovering local visual material attributes. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 3565–3573, June 2015.

[SSW+14] C. Schwartz, R. Sarlette, M. Weinmann, M. Rump, and R. Klein. Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the university of bonn. *Sensors*, 14(5):7753–7819, 2014.

[Ste46] S. Stevens. On the theory of scales of measurement. *Science*, 103(2684):677–680, 1946.

[Ste57] S. Stevens. On the psychophysical law. *Psychological Review*, 64(3):153, 1957.

[SWRC06] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *Computer Vision – ECCV 2006*, pages 1–15, Berlin, Heidelberg, 2006. Springer.

[SZ06] C. Spence and M. Zampini. Auditory contributions to multisensory product perception. *Acta Acustica united with Acustica*, 92(6):1009–1025, 2006.

[TdSL00] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Ann. NY Acad. Sci*, 911:418, 2000.

[TJL+14] A. Tajadura-Jiménez, B. Liu, N. Bianchi-Berthouze, and F. Bevilacqua. Using sound in multi-touch interfaces to change materiality and touch behavior. In *Proc. of the 8th Nordic Conf. on Human-Computer Interaction*, NordiCHI '14, pages 199–202, New York, NY, USA, 2014. ACM.

[UA09] U. of Amsterdam. Amsterdam library of textures (ALOT). http://aloi.science.uva.nl/public_alot/, 2009. Accessed on July 31, 2018.

[UGP+17] P. Upchurch, J. R. G., G. Pleiss, R. Pless, N. Snavely, K. Bala, and K. Weinberger. Deep feature interpolation for image content changes. In *Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 6090–6099, 2017.

[VLD07] P. Vangorp, J. L., and P. Dutré. The influence of shape on the perception of material reflectance. *ACM Trans. on Graphics*, 26(3):77:1–77:9, July 2007.

[VZ05] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1):61–81, Apr 2005.

[WAKB09] J. Wills, S. Agarwal, D. Kriegman, and S. Belongie. Toward a perceptual space for gloss. *ACM Trans. on Graphics*, 28(4):103:1–103:15, Sept. 2009.

[War92] G. Ward. Measuring and modeling anisotropic reflection. *SIGGRAPH Comput. Graph.*, 26(2):265–272, July 1992.

[WBCB08] C. Wallraven, M. Breidt, D. Cunningham, and H. Bülthoff. Evaluating the perceptual realism of animated facial expressions. *ACM Trans. on Applied Perception*, 4(4):4:1–4:20, Feb. 2008.

[WBN+11] B. Williams, S. Bailey, G. Narasimham, M. Li, and B. Bodenheimer. Evaluation of walking in place on a wii balance board to explore a virtual environment. *ACM Trans. Appl. Percept.*, 8(3):19:1–19:14, Aug. 2011.

[WBSS04] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing*, 13(4):600–612, April 2004.

[WFEM10] G. Wendt, F. Faul, V. Ekroll, and R. Mausfeld. Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision*, 10(9):7, 2010.

[WGK14] M. Weinmann, J. Gall, and R. Klein. Material classification based on training data synthesized using a btf database. In *European Conf. on Computer Vision*, pages 156–171. Springer, 2014.

[WL00] L. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proc. of the 27th Annual Conf. on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 479–488, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[WLGK16] M. Weinmann, F. Langguth, M. Goesele, and R. Klein. Advances in Geometry and Reflectance Acquisition. In *EG 2016 - Tutorials*. The Eurographics Association, 2016.

[WLKT09] L. Wei, S. Lefebvre, V. Kwatra, and G. Turk. State of the Art in Example-based Texture Synthesis. In *Eurographics 2009, State of the Art Report, EG-STAR*, pages 93–117, Munich, Germany, Mar. 2009. Eurographics Association.

[WLL+09] T. Weyrich, J. Lawrence, H. Lensch, S. Rusinkiewicz, and T. Zickler. Principles of appearance acquisition and representation. *Foundations and Trends in Computer Graphics and Vision*, 4(2):75–191, Oct. 2009.

[Woo79] R. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. *Proc. SPIE*, 0155:0155 – 0155 – 8, 1979.

[XBJ+16] B. Xiao, W. Bi, X. Jia, H. Wei, and E. Adelson. Can you see what you feel? color and folding properties affect visual–tactile material discrimination of fabrics. *Journal of Vision*, 16(3):34–34, 2016.

[XJF09] Y. Xu, H. Ji, and C. Fermüller. Viewpoint invariant texture description using fractal analysis. *Int. Journal on Computer Vision*, 83(1):85–100, June 2009.

[XR16] X-Rite. Tac7 physical material scanner. https://www.xrite.com/categories/appearance/total-appearance-capture-ecosystem/tac7, 2016. Accessed on July 31, 2018.

[ZCS03] L. Zhang, B. Curless, and S. M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *CVPR*, 2003.

[ZPN+15] J. Zujovic, T. Pappas, D. Neuhoff, R. van Egmond, and H. de Ridder. Effective and efficient subjective testing of texture similarity metrics. *Journal Optical Soc. of America*, 32(2):329–342, Feb 2015.