Sound meets Image: Freedom of Expression in Texture Description

Reinier J. Jansen*, René van Egmond, Huib de Ridder
Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands

ABSTRACT

The use of sound was explored as means for expressing perceptual attributes of visual textures. Two sets of 17 visual textures were prepared: one set taken from the CUReT database[1], and one set synthesized to replicate the former set. Participants were instructed to match a sound texture with a visual texture displayed onscreen. A modified version of a Product Sound Sketching Tool[2] was provided, in which an interactive physical interface was coupled to a frequency modulation synthesizer. Rather than selecting from a pre-defined set of sound samples, continuous exploration of the auditory space allowed for an increased freedom of expression. While doing so, participants were asked to describe what auditory and visual qualities they were paying attention to. It was found that participants were able to create sounds that matched visual textures. Based on differences in diversity of descriptions, synthetic textures were found to have less salient perceptual attributes than their original counterparts. Finally, three interesting sound synthesis clusters were found, corresponding with mutually exclusive description vocabularies.

Keywords: textures, expressivity, free labeling, CUReT database, frequency-modulated sounds

1. INTRODUCTION

Visual, auditory, and tactile textures provide information about events and objects in the environment. For example, when walking from one room to another, we may recognize a transition from a stone floor to a carpet by seeing a shiny versus a mat texture, hearing reverberant versus dampened sounds, and feeling differences in softness. Designers are considering more and more non-visual product experiences. Therefore, it is interesting to comprehend how sensory modalities can influence each other. It has been found that modified auditory feedback can influence the perceived roughness[3]. As people rubbed their fingers on sandpaper, the resulting sound was played back over headphones. Boosting high frequency content caused an increase in perceived roughness. Indeed, roughness is an interesting factor to study in multi-modal research, as it is present in different perceptual modalities.

In Van Egmond, *et al.*[4] the perceived roughness of visual and of auditory materials was investigated. Forty-nine images of real world surfaces of the CUReT database[1], as well as a selection of frequency-modulated tones, were judged on roughness using a paired-comparison paradigm. A follow-up study[5] was run to find an objective predictor for perceived roughness of visual textures. Seventeen CUReT images of the previous study were selected, and synthetic textures with uniformly distributed white noise of the same variance were generated. A perceptual experiment employing a rank-ordering paradigm established that the distribution of energy in subbands is a strong and simple predictor for subjective roughness measures on uniformly distributed textures. However, a systematic deviation in perceived roughness was found between synthetic and original textures. It was suggested that texture synthesis may change attributes of textures (e.g., glossiness), which may have an effect on perceived roughness. To investigate potential relations between visual roughness and other perceptual texture attributes, it is first necessary to know which attributes are perceived at all.

Given our multimodal interaction with most everyday objects, it would be interesting to examine how sound can be used as means for expressing these attributes. This exploratory study is a first step in providing participants with sound as additional dimension of expression for visual texture description. To assess whether this approach is viable, a test will be conducted to investigate whether people can use a sound synthesizer to generate auditory textures that match visual textures, and formulate words that describe this match. The resulting set of descriptions will be used in our in understanding of differences between original CUReT textures and corresponding synthetic textures.

*r.j.jansen@tudelft.nl; phone +31 (0)15 27 84956

Human Vision and Electronic Imaging XVII, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8291, 829119 • © 2012 SPIE-IS&T
CCC code: 0277-786X/12/$18 • doi: 10.1117/12.914698

2. INTERACTIVE PHYSICAL INTERFACE

Two topics need to be addressed before setting up a test to explore how sound can be used as means of expression: the sounds themselves, and an interactive physical interface. Ideally a sound is chosen with a known effect on an auditory perceptual dimension, and this dimension should be present in the visual system as well. Roughness is known to be such a dimension, yet little is known about other audiovisual perceptual dimensions[6]. Frequency-Modulated (FM) sounds have been used with success in subjective roughness evaluation studies[4,7]. Therefore, FM sounds will be used as starting point. To allow a large degree in freedom of expression, participants will be asked to explore an auditory space by trial and error within a limited amount of time. This in order to elicit associations they consider suitable for a given visual texture. Buxton[8] views 'explorative', 'quick', and 'suggestive' as typical attributes of a sketching process. Therefore, this act of creating sounds to visual textures can be viewed as an act of sketching. In a previous study by Jansen, *et al.*[2] an interactive physical Product Sound Sketching Tool was developed. Findings suggest that people are able to use this tool to explore and generate auditory concepts for a predefined product character (e.g., 'energetic'). Therefore, an interactive physical interface may be an appropriate method for sound exploration evoked by the texture of images. Additionally, the authors suspect a playful interface is more fun to work with for a longer period of time.

On the software side, the 'Operator' FM synthesizer of Ableton Live 8.2.2. was used to generate auditory textures. The synthesizer was set up with one carrier waveform (Oscillator A), two frequency-modulating waveforms (Oscillators B & C) and a band pass filter. Users could modify the carrier frequency, modulation frequencies, modulation depths, and the filter center frequency (see Table 1 for parameter ranges. On the hardware side, a table with a camera and a glass plate was constructed, on which four glasses and a loudspeaker-like object could be placed (see Fig. 1). A dedicated program with the visual programming language Max was designed. This program would relate the distance between a glass and the loudspeaker object to the oscillator or filter frequency of the corresponding Operator section. Increasing this distance by sliding a glass would result in a lower frequency, and vice versa. Additionally, the modulation depth of oscillators B & C could be controlled by rotating their glasses. Users were free in choosing whether to apply modulation and/or filtering by placing or removing corresponding glasses. The system allowed for sliding and rotating multiple glasses simultaneously. A demonstration with example sounds of the system can be found at Jansen[9].

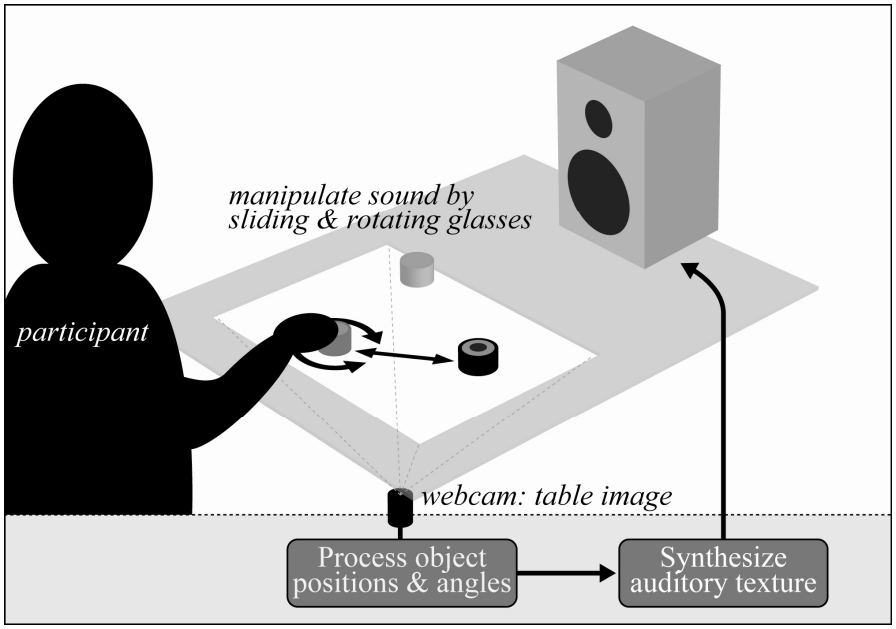

Figure 1: Components of and interactions with the physical interface.

Table 1: Parameter ranges of the Operator FM synthesizer in Ableton Live 8.2.2. Frequency parameters (*) and modulation depth parameters (**) are adjustable by sliding and rotating the associated glass of the physical interface, respectively.

Live Operator section	Oscillator A	Oscillator B	Oscillator C	Filter
Function	Carrier wave	Modulator of Osc A	Modulator of Osc A	Bandfilter 'SVF'
Frequency range (Hz)	Variable*: 10 to 240	Variable*: 1 to 69.3	Variable*: 1 to 69.3	Variable*: 30 to 18.5k
Level (dB)	Fixed: 0	Variable**: -4 to -24	Variable**: -4 to -24	Slope: 12 dB/oct, Q: 5
Associated glass color	Yellow	Green	Blue	Brown

3. EXPERIMENT

The goal of this experiment is to investigate: a) whether people are able to create sounds that match visual textures, b) whether people are able to describe textures based on similarity and dissimilarity between image and created sound, and c) if the chosen interactive approach facilitated the freedom of expression.

3.1 Method

A combination of interactive sound synthesis and a free-choice labeling paradigm was employed to create sounds and descriptions for seventeen original and synthesized visual textures stemming from the CUReT database.

3.1.1 Participants

Eight students and employees of the Faculty of Industrial Design Engineering volunteered (4 males, 4 females, 24 to 54 years old, average 30 years). All participants were native Dutch speakers, reported normal hearing, and normal or corrected-to-normal vision. None reported experience in describing auditory textures. Two participants said to have selected visual textures for product development during their education.

3.1.2 Stimuli

The seventeen original and seventeen synthesized visual textures from Van Egmond et al.[5] were used. One set was taken from the CUReT database, and one set consisted of images synthesized to replicate the former set. In Fig. 2 the sixteen original textures are shown in the top of each panel, and the sixteen corresponding synthetic textures are displayed in the bottom of each panel.

3.1.3 Apparatus

The experiment took place in a quiet room. Participants had to create sounds matching the visual textures by operating the interactive physical interface described in Section 2. One Behringer Truth B2301 active monitor was connected to a MacBook Pro 15" for monophonic amplification. The monitor was placed at a distance of approximately 1m in front of the participant, and set at a height of 1.40m to accommodate listening in both seated and standing position. The screen (resolution: 1440x900px) was placed at a distance of approximately .70m from the participant.

A dedicated Max patch was used to present visual textures centered on the screen against a white background (texture size: 240x240px). On the top side of the screen one would find the task description: 'Make a sound having the same texture as the image' (Dutch: 'Maak een geluid met eenzelfde textuur als de afbeelding'). Two text fields were used to collect terms describing sound and image. In the left field a participant could indicate how similar sound and image were, and in the right field how dissimilar. It was not obligatory to fill in text fields in order to proceed to a next trial. The Max patch recorded the synthesizer parameter settings with a sample rate of 100ms. In addition, the final sound at the end of each trial was saved in .wav format. Oscillator A was a prerequisite to hear a sound, but participants were free to choose whether or not to apply modulation and/or filtering. Consequentially, eight logical combinations of synthesizer sections were possible: 1{only Oscillator A}, 2{Oscillator A + Filter}, 3{Oscillators A + C}, 4{Oscillators A + C + Filter}, 5{Oscillators A + B}, 6{Oscillators A + B + Filter}, 7{Oscillators A + B + C}, and 8{Oscillators A + B + C + Filter}.

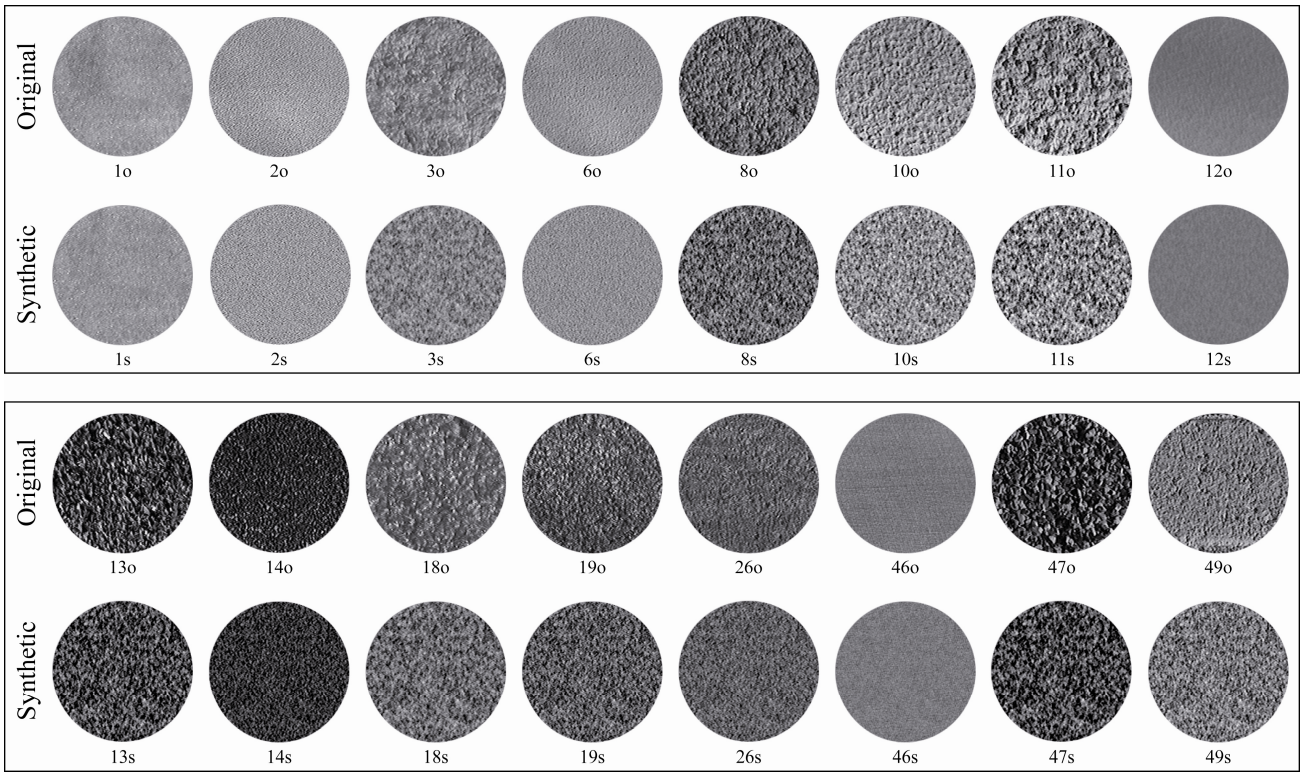

Figure 2: 16 original textures of the CUReT database and 16 synthetic textures created using the energy distributions in the subbands and uniformly distributed white noise. Numbers refer to the CUReT database, whereas 'o' and 's' refer to original and synthetic, respectively. Note that the 17th texture (CUReT #54) is left out only because of graphical display issues. Image adapted from Van Egmond et al.[5].

3.1.4 Procedure

Participants were informed that they had to create sounds that matched the visual textures, but were left unaware of the nature of these textures. Printed instructions on possible interactions with the interface were provided (e.g., on dragging and rotating glasses). Participants would start with five minutes of free manipulation to become acquainted with the interface and its sound space. The monitor volume was adjusted to a comfortable listening level.

A participant received either the CUReT texture set, or the resynthesized CUReT texture set. Prior to the start of the actual experiment, all textures of a set were presented sequentially (750ms each) to become familiar with the perceptual visual domain. The actual experiment consisted of twenty trials, during which participants had to create a sound matching the visual texture and fill in the text fields with a keyboard. Participants were instructed to use their own vocabulary, and were told that certain descriptions could be used for several images. Text fields could be left blank in case no fitting description could be made up.

Before the actual experiment started, three textures were presented that were found most rough (texture #47), least rough (#12), and medium rough (#3) in Van Egmond, *et al.*[5]. In the twenty experimental trials all seventeen textures, either original or synthesized, were presented in random order. Three textures out of the set of seventeen were randomly chosen to be repeated. Participants would never see the same texture in adjacent trials. Upon pressing the 'Next' button, the final sound of that trial was recorded for a duration of three seconds. During this time a blank screen was shown, and synthesizer parameters could not be changed until the start of the next trial.

The participants were asked to describe the textures and the way they would sonify the texture aloud. The experimenter wrote down these utterances and would ask for an explanation when applicable. During debriefing participants were asked about how they felt it went, about their approach, and further questions regarding observations made by the experimenter during the session. There were no time restrictions.

3.2 Results

This section will examine freedom of expression in the chosen method of texture description, and will conclude with an application of this method to understand roughness judgments in a previous study.

3.2.1 Experience of Freedom in Expression

All participants succeeded in generating sounds that they found to match the visual stimuli. Each session lasted approximately forty-five minutes. Various strategies were identified, both between and within participants. For example, some participants (pp 2,3,5,6,8) commented that they often first explored the sound space according to intuition, before making up a description. Other times, participants (pp 1,2,4,6,7) remarked that they first experienced an association (e.g., images, words), after which they started to create a sound, and finally made up a description. Two participants (pp 3,8) sometimes started by entering a description, and then continued by creating a corresponding sound. Finally, all participants would occasionally iterate between refining the sound and refining the description. Two participants (pp 2,8) commented positively on the versatility of the sound space.

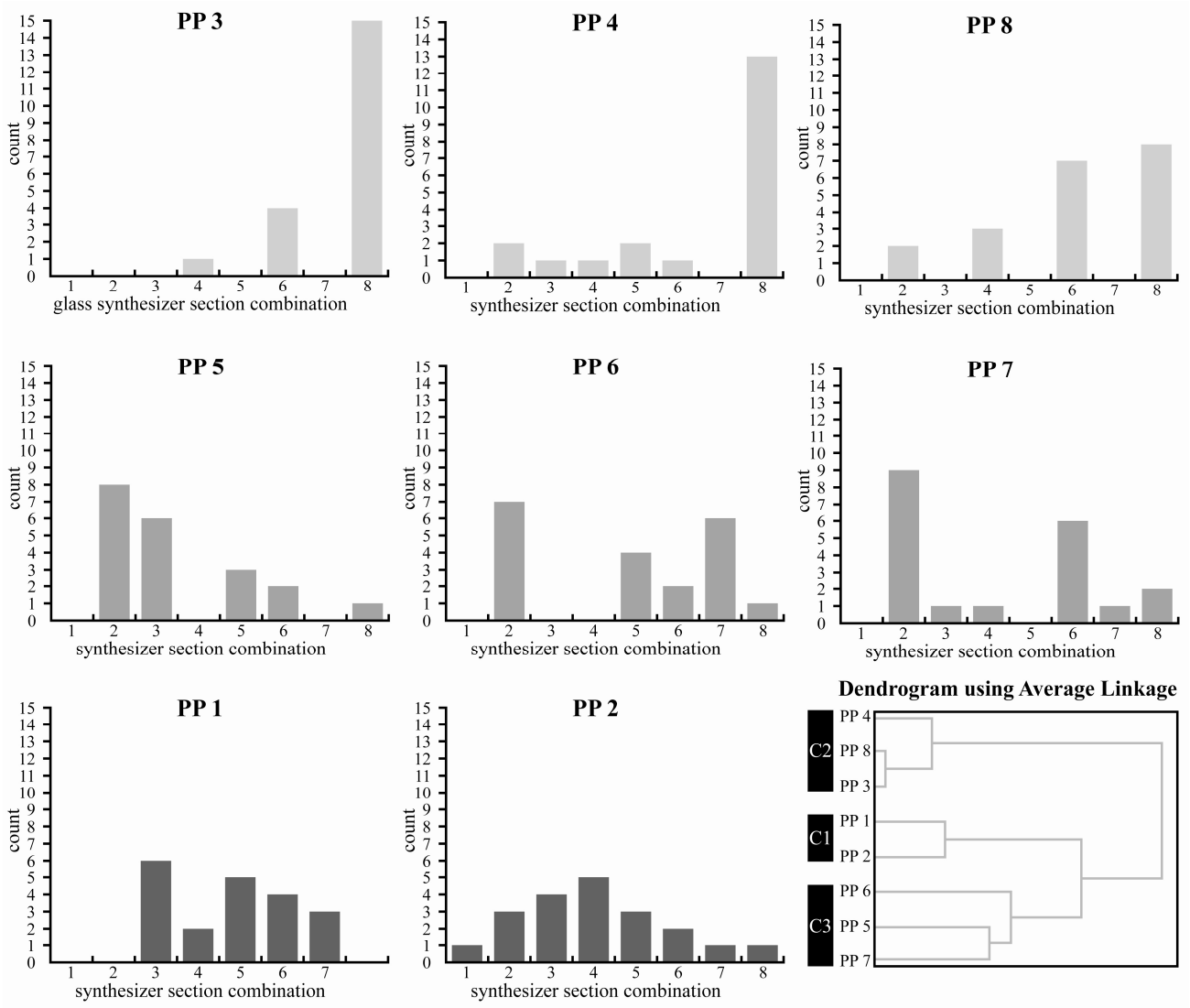

Figure 3: Frequencies of logical combinations of synthesizer sections per participant (see §3.1.3 for the contents of each combination). Bottom right: dendrogram of the hierarchical cluster analysis of the synthesizer section combinations, leading to clusters C1 (pp 1+2), C2 (pp3+4+8), and C3 (pp 5+6+7).

A first analysis on how participants explored the sound space is shown in Fig. 3. Each bar represents the frequency of a combination of synthesizer sections (see §3.1.3). The shapes of the bar charts suggest that participants may have had different preferences regarding these combinations. A hierarchical cluster analysis (SPSS 19, method: average linkage, measure: Chi-square) was performed on the frequencies of each combination. The dendrogram in Fig. 3 suggests a solution of three clusters. Cluster C1 (participants 1 & 2) has a bell-like shape, which corresponds with always using Oscillator A and one or two synthesizer. In cluster C2 (participants 3, 4, 8) participants mostly use the filter section, with a strong preference for using all sections. The distribution in cluster C3 (participants 5, 6, 7) contains a peak with a strong preference for using only the filter section, and a peak with a preference for using at least Oscillator B.

Participants expressed various effects of glasses on the sound. For example, pp1 related the green glass (i.e., Oscillator B) with 'bumpy' (Dutch: 'bobbelig'), blue (Oscillator C) with 'sharp' ('scherp'), and brown (filter) with 'dark and light' ('donker en licht'). Additionally, relations between image and sound attributes were uttered (e.g., dark images & low frequencies vs. light images & high frequencies). Most participants were enthusiastic about using the interface, even though they expressed difficulties in making up descriptions. Nevertheless, all participants were able to complete the similarity text fields, using either separate words or full sentences. Across participants, opposing text fields at four different images were left blank.

All descriptions were stripped of articles, pronouns, verbs, prepositions, and conjunctions. The words 'image', 'sound', and synonyms thereof were removed, as they do not refer to a subjective characteristic of the multimodal combination. Additionally, superlative and comparative adjectives were turned into normal adjectives (e.g., 'darker' became 'dark'). In case of an unequal amount of descriptions in both text fields the authors chose the most appropriate pairs. In many cases, one description could have multiple opposing descriptions (i.e., 'smooth' was opposed by 'bumpy', 'coarse', 'fluffy', 'grainy', 'grating', 'rigid', 'rough', 'sharp', and 'speckled').

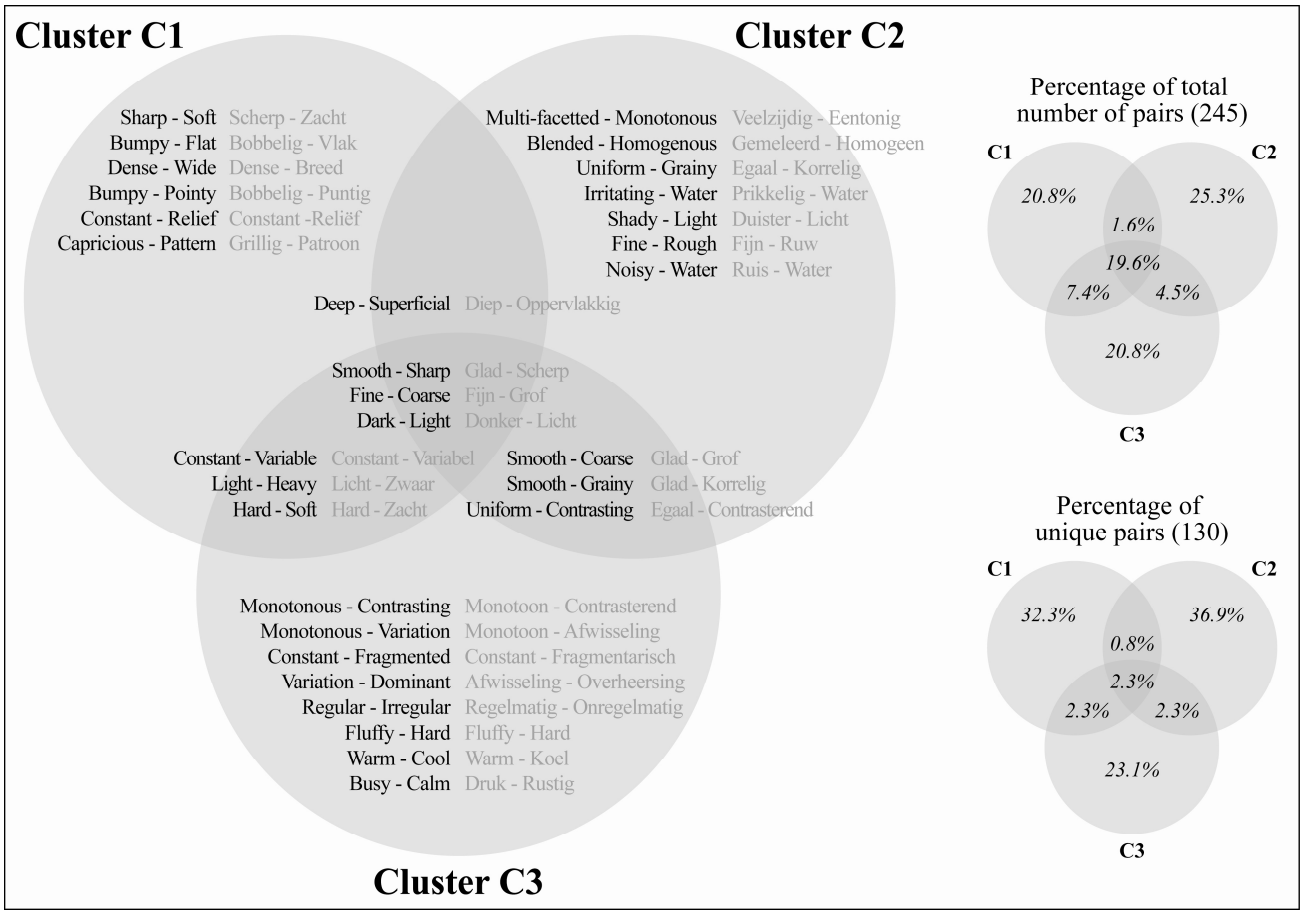

Figure 4: Venn diagrams of overlapping clusters with non-directional description pairs. For readability only terms with frequency >1 are displayed.

Including repetitions, this resulted in a total of 245 pairs to describe similarities and dissimilarities between images and sounds. Many pairs of descriptions were found in two directions (e.g., 'coarse-fine' and 'fine-coarse'). Leaving out this directionality (i.e., regarding 'coarse-fine' and 'fine-coarse' as identical), and filtering out repetitions, participants used 130 unique pairs. The richness of a vocabulary can be judged through quantity and diversity. The highest amount of pairs was found in cluster C2 (100), whereas C1 (71) and C3 (74) consisted of considerably less pairs. On the other hand, a similar amount of unique pairs was found in clusters C1 (50), C2 (54), and C3 (50). Compared to C1 and C3, this means that in C2 more pairs were used at higher frequencies. A Venn diagram was constructed to examine overlap of clusters in non-directional description pairs, see Fig. 4. In total 66.9% of the 245 pairs were mutually exclusive between the clusters. On the other hand, 33.1% of the pairs occurred in multiple clusters, of which three pairs (19.6%) were featured in all clusters: 'smooth-sharp', 'fine-coarse', and 'dark-light'. However, these pairs represent only 2.3% of all unique pairs. In total 92.3% of the 130 unique pairs were mutually exclusive between the clusters. These results suggest different vocabularies were used within each cluster.

3.2.2 Texture descriptions of original versus synthetic images

In a previous study by Van Egmond, *et al.*[5] the same set of stimuli was rank-ordered by participants on perceived roughness. A 1D solution was found, on which three main categories could be visually distinguished: Smooth, Rough, and Medium Rough. Additionally, a systematic deviation between the judgments on the original and synthetic textures was found: synthetic textures were judged to be less rough than the original ones and this difference increased with the roughness of the original texture. The descriptions of the current experiment were matched to these roughness categories. In addition, the differences in descriptions between original and synthetic textures were determined.

Table 2: Frequencies of audiovisual texture descriptions, distributed according to the roughness clusters described in Van Egmond, *et al.*[5]. Descriptions were translated from Dutch for readability.

Roughness cluster Smooth				Roughness cluster Medium rough				Roughness cluster Rough			
Original	n	Synthetic	n	Original	n	Synthetic	n	Original	n	Synthetic	n
fine	7	uniform	4	dark	7	dark	4	coarse	12	coarse	3
light	6	grainy	3	coarse	6	shady	4	sharp	6	grainy	3
soft	6	gray	2	dense	3	irritating	3	irregular	5	dark	2
regular	4	monotonous	2	irregular	3	variable	3	bumpy	4	distorted	2
constant	2	noisy	2	blended	2	contrasting	2	dark	3	gray	2
homogeneous	2	boring	1	calm	2	light	2	fluffy	3	irritating	2
warm	2	constant	1	grainy	2	sophisticated	2	blended	2	noisy	2
blended	1	distorted	1	hard	2	deep	1	capricious	2	shady	2
bumpy	1	dull	1	messy	2	demure	1	deep	2	chaotic	1
calm	1	elegant	1	soft	2	dented	1	heavy	2	cheerful	1
coarse	1	fine	1	alien	1	fine	1	rough	2	contrasting	1
grainy	1	fragmented	1	black	1	fragmented	1	round	2	deep	1
heavy	1	light	1	bumpy	1	grating	1	soft	2	fierce	1
relief	1	multi-facetted	1	busy	1	irregular	1	busy	1	fragmented	1
ribbed	1	soft	1	capricious	1	multi-facetted	1	dense	1	grains of sand	1
rotating	1	speckled	1	changeable	1	repeating	1	earthly	1	grating	1
short-haired	1	stained	1	chunky	1	sand	1	fine	1	light	1
sludgy	1	structureless	1	concentrated	1	scattered	1	flowery	1	mat	1
smooth	1	wave	1	deep	1	sharp	1	hard	1	multi-facetted	1
speckled	1			differing	1	smooth	1	high	1	sharp	1
tight	1			fine	1	snow	1	light	1		
				heavy	1	worked up	1	mossy	1		
				light	1			mountain tops	1		
				movement	1			smooth	1		
				organic	1			warm	1		
				regular	1			watery	1		
				rough	1						
				strange	1						
				uneven	1						
Total	43	Total	27	Total	50	Total	35	Total	60	Total	30

Textures were allocated to roughness categories based on the original versions, to enable comparison with their synthetic counterpart. As a consequence, synthetic textures 8s, 13s, and 18s were allocated to the Rough category, instead of Medium Rough. Textures are now distributed as follows: Smooth {1,2,6,12,46}, Medium Rough {3,14,19,26,49}, and Rough {8,10,11,13,18,47,54} (see Fig. 2 for visual reference). In the previous section non-directional pairs were examined to identify the amount of unique pairs constituting a vocabulary. For this application only directional pairs are of interest (i.e., the order in which similarity and opposing text fields were entered). Table 2 displays the frequencies of descriptions for each roughness category, each of which is further split into original versus synthetic descriptions. For readability, all descriptions of similarity text fields were translated from Dutch by the authors, and opposing descriptions have been omitted (note: some similarity descriptions may have multiple opposing ones).

When comparing original versus synthetic textures, Table 2 shows that within each roughness category the original textures feature a higher description count, as well as a higher number of unique terms. This suggests that the current texture synthesis algorithm may cause textures to be perceived with less salient perceptual features than their original counterparts. This is in line with Van Egmond, *et al.*[5], who mention that "texture synthesis may change the glossiness or other attributes of textures, which may have an effect on perceived roughness." Zooming in on the Rough category, Table 2 shows that the words 'coarse' (15), 'sharp' (7), dark (5), and 'irregular' (5) were most used. Surprisingly, the word 'rough' itself was only used twice within this category, and three times overall. On the other hand, within the Smooth category the descriptions 'fine' (8), 'light' (7), 'soft' (7), 'grainy' (4), 'regular' (4), and 'uniform' (4) were most used, whereas 'smooth' was only mentioned once. This suggests roughness, interpreted as 'rough' versus 'smooth', may not have been the most salient feature for these combinations of visual and auditory textures. Finally, within the Rough and Medium Rough categories descriptions semantically related to 'rough' (i.e., 'coarse', 'sharp', 'grating') appear more often for original textures. This confirms the systematic deviation found in Van Egmond, *et al.*[5].

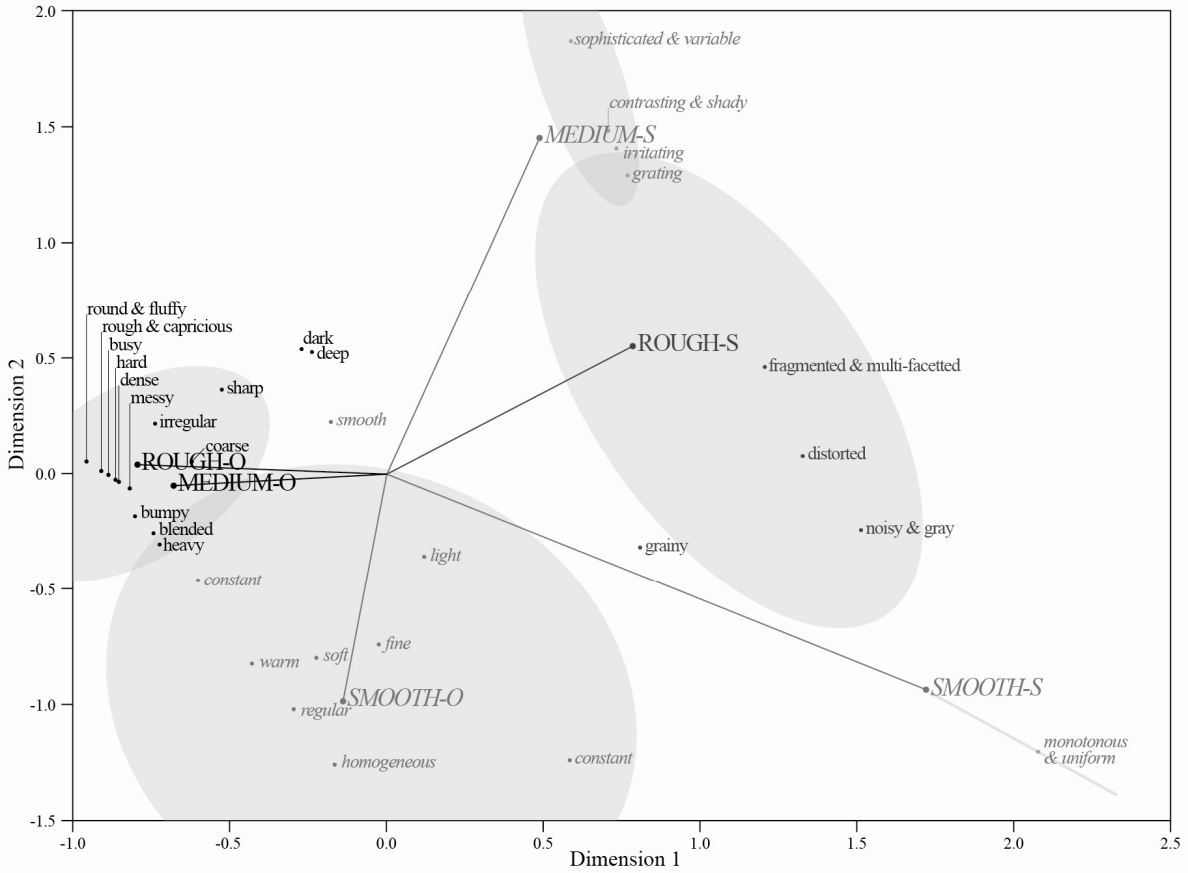

Figure 5: A five-cluster solution based on a correspondence analysis of descriptions and three roughness categories. Roughness categories are appended with 'O' and 'S' to differentiate between original and synthetic textures. Only descriptions that occurred more than twice in the total data set are displayed.

A correspondence analysis (JMP v9.0.2) was conducted to verify if the descriptions themselves converge to the same roughness categories, and original versus synthetic sub-categories. The coordinates of the 3D solution were used in an hierarchical cluster analysis using Ward's method. The cluster analysis yielded 5 clusters. In Fig. 5 the roughness descriptions and the attributes are presented in a space spanned up by the first two dimensions of the cluster analysis. The clusters are indicated by bivariate normal ellipses (p=.95). The three roughness categories of Van Egmond, *et al.*[5] are appended with 'O' and 'S' for original and synthetic textures, respectively. Based on the description count, the ROUGH-O and MEDIUM-O categories are joined in one cluster, whereas SMOOTH-O, SMOOTH-S, MEDIUM-S, and ROUGH-S are found to be dissimilar. The descriptions within each cluster appear to correspond with the distribution of descriptions in Table 2. For example, in Table 2 the column of original smooth textures mostly contains the descriptions 'fine', 'light', 'soft', and 'regular'. This is reflected by the proximity of these descriptions to the label SMOOTH-O in Fig. 5.

Closer inspection of Fig. 5 shows that all original clusters are found on the left side of Dimension 1, whereas synthetic clusters are located on the right side. This suggests that synthetic textures elicit different descriptions than original textures. Another explanation to explain this difference may be that participants were either in the original, or in the synthetic stimulus condition. Logically, this should be reflected by the use of certain synthesizer sections, as descriptions were given to the combination of sound and image. However, clusters based on synthesizer sections in §3.2.1 include participants of both conditions. Therefore, additional research is required in which participants describe both types of textures.

3.3 Discussion

There are four main findings in this study. First, the proposed interactive physical interface and accompanying FM-synthesizer made it possible to create sounds that participants found to match visual textures. Furthermore, participants were able to formulate descriptions on the similarity (e.g., 'light') and dissimilarity (e.g., 'dark', 'heavy') between sound and image, of which the latter description helped in interpreting the former. However, some combinations are still ambiguously defined (i.e., 'light-heavy' may refer to physical weight, emotional load, or rigidness of a mechanism). The activity was generally experienced as difficult, but exciting.

Second, three different synthesis strategies were identified. The descriptions of the vocabularies for each cluster shared a similar richness in diversity, and were largely mutually exclusive. The descriptions of the vocabularies that overlapped appear to be universally applicable, as opposed to specific terminology. This suggests that each FM-synthesizer section helps in describing different perceptual dimensions of visual textures. Additionally, the relation between sound and image can be explored in more detail by comparing synthesizer parameter values (as opposed to binary measurement of synthesizer sections), descriptions, and (psycho-)physical measures.

Third, findings suggest that the current texture synthesis algorithm may cause textures to be perceived with less salient features than their original counterparts. This was found by comparing the richness in diversity of descriptions allocated to three roughness categories described in Van Egmond, *et al.*[5], subdivided in original versus synthetic textures. This distribution of descriptions was confirmed by means of a correspondence analysis. So far, no synonymous descriptions were taken into account. However, some descriptions may be regarded as similar (e.g., 'distorted' / 'noisy', or 'dark' / 'shady').

Fourth, the Dutch word 'ruw' (English: 'rough') was used only three times throughout the experiment. This indicates a limitation of the previous study, in which participants were forced to think in terms of roughness, whereas they appear to differentiate mostly on other perceptual attributes.

Furthermore, future explorations could investigate if a similar vocabulary is used to describe visual textures when no auditory textures have to be created. Also, the description pairs can be used to construct a set of semantic differential scales. A rating experiment with three stimulus conditions (i.e., sound, image, sound & image) can be useful to examine the influence of each modality on texture descriptions.

4. CONCLUSION

This exploratory study was a first step in using sound as facilitator for unveiling perceptual dimensions of visual textures. From the results it is clear that this approach is worth further investigation. People enjoyed using the interactive physical interface to control a frequency modulation synthesizer, and were successful at creating auditory textures and formulating descriptions that match visual textures. Based on a difference in diversity of vocabularies, it was suggested that the current texture algorithm causes synthetic textures to be perceived with less salient perceptual features than their original counterparts. Finally, three interesting clusters of sound synthesizer use were found, corresponding with mutually exclusive description vocabularies. This cluster analysis warrants further research with more participants, and under controlled experimental conditions.

ACKNOWLEDGMENTS

We gratefully acknowledge Elif Özcan and Ans van Doorn for their valuable comments, and all participants volunteering in the experiment.

REFERENCES

[1] Dana, K. J., van Gineken, B., Nayar, S. K., Koenderink, J. J., "CUReT database," ACM Transactions on Graphics, 18, 1-34 (1999).
[2] Jansen, R. J., E. Özcan, and Egmond, R. van, "PSST! Product Sound Sketching Tool," Journal of the Audio Engineering Society 59(6): 396-403 (2011).
[3] Spence, C. and M. Zampini, "Auditory contributions to multisensory product perception," Acta Acustica United With Acustica 92(6): 1-17 (2006).
[4] Egmond, R. van, Lemmens, P., Ridder, H. de, "Roughness in sound and vision," Proceedings of Human Vision and Electronic Imaging (2009).
[5] Egmond, R. van., Pappas, T. N., Ridder, H. de, "Subband analysis and synthesis of real-world textures for objective and subjective determination of roughness," Proceedings of Human Vision and Electronic Imaging (2010).
[6] Klatzky, R. L. and Lederman, S. J., "Multisensory Texture Perception," In: Multisensory Object Perception in the Primate Brain. J. Kaiser and M. J. Naumer, Springer New York: 211-230 (2010)
[7] Fastl, H. and Zwicker, E. [Psychoacoustics: Facts & Models], Springer, Berlin (2007)
[8] Buxton, W. [Sketching User Experiences: Getting the Design Right and the Right Design], Morgan Kaufmann, San Francisco, CA (2007).
[9] Jansen, R. J., "SPIE demonstration video," http://studiolab.ide.tudelft.nl/studiolab/jansen/research/sound-meets-image/, last visited: January 2012.