Perceptual Analysis of Haptic Icons:
an Investigation into the Validity of Cluster Sorted MDS

Jerome Pasquero*
Center for Intelligent Machines
McGill University
Montreal, Qc, H3A 2A7, Canada

Joseph Luk, Shannon Little, Karon MacLean†
Department of Computer Science
University of British Columbia
Vancouver, B.C. V6T 1Z4, Canada

ABSTRACT

The design of usable haptic icons (brief informational signals delivered through the sense of touch) requires a tool for measuring perceptual distances between icons that will be used together as a set. Our experiences with one potentially powerful approach, Multidimensional Scaling (MDS) analysis of perceptual data acquired using an efficient cluster sorting technique, raised questions relating to the methodology for data collection. In this paper, we review key issues relating to perceptual data collection method, describe an example data set and present its initial MDS analysis, and then examine the impact of collection method on MDS outcome through a secondary analysis of the data and the inherent structure of the algorithm components. Our analysis suggests that an understanding of these issues is important for the method's effective use, but has not exposed any major flaws with the process.

CR Categories: H.1.2 [Information Systems]: User/Machines Systems—Human Information Processing; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Haptic I/O; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Prototyping

Keywords: tactile display, haptic icons, handheld device, multidimensional scaling, tactile perception, tactile feedback

1 INTRODUCTION

Using synthetic tactile feedback, we can communicate information about the state or content of a system, or the occurrence of an event, to a human user via cutaneous touch. Mobile devices which make use of tactile feedback (e.g. vibrating pagers and cell phones) have been available on the consumer market for several years. The tactile stimulation that they generate, however, is limited in its expressive capability; typically they make use of a single tactile element that vibrates at a unique frequency. More recently, mobile device manufacturers have been integrating tactile actuators with increased degrees of freedom into their products. This is possible, for instance, by using actuators for which the vibrating frequency can be controlled. The goal, inspired by the common use of audio icons in desktop interfaces and mobile telephony [1], is to allow for the design and construction of specific and short abstract tactile messages that can easily be interpreted by users with minimal cognitive effort [7].

At the center of our research lies the goal of understanding how to maximize the level of useful information content that tactile/haptic feedback devices can convey. To that end, we aim to develop a systematic analytic method to characterize the stimulus

---

*e-mail: jay@cim.mcgill.ca
†e-mail:{luk, shlittle, maclean}@cs.ubc.ca

space of tactile feedback devices in order to inform the design of tactile icons.

In this paper, we present our progress to date in clarifying and validating the use of multidimensional scaling (MDS) for this purpose. While it seems clear from past efforts that MDS analysis can provide relevant and (in some sense) accurate information about the perceived relationships among a set of test stimuli and thus guide stimulus set design, we were not satisfied with our ability to interpret and quantify these relationships, nor with our understanding of the impact of perceptual data collection method on the MDS analysis outcome.

Here, we begin by reviewing key issues relating to data collection method, then proceed to examine the impact of collection method on the MDS result via a detailed analysis of an example data set. While an exhaustive critique will be a substantial project, the insights presented here should aid in the wise use and accurate interpretation of results provided by a potentially powerful tool. Our analysis does not expose any major flaws with the process, and we thus conclude that it can be a valuable tool to evaluate the expressive capability of haptic devices.

2 BACKGROUND

2.1 Multidimensional Scaling

MDS comprises a set of methods that display the structure of a dissimilarity matrix as a geometrical representation in a given number of dimensions [9].

A dissimilarity matrix contains $N * (N-1)/2$ individual perceived dissimilarity scores between N items of a set. MDS optimization algorithms try to map the perceived dissimilarity scores δ_{ij} between items i and j from a dissimilarity matrix onto an M-dimension geometrical space (usually Euclidean) within which distances d_{ij} between items represent relative levels of dissimilarity:

$$d_{ij} \approx f(\delta_{ij}) \ \ i,j \in [1,2\cdots,N] \tag{1}$$

where $f()$ is a monotonic function. On the Euclidian space, distances are calculated with

$$d_{ij} = \sqrt{\sum_{a=1}^{M}(x_{ia}-x_{ja})^2} \tag{2}$$

where x_{ia} and x_{ja} are the item coordinates that the MDS algorithm is solving for.

Dissimilarity data are always mapped with a variable degree of error to the geometrical space, and several measures are used to evaluate the goodness of fit. For instance, one popular method introduced by Kruskal [5] and used to report goodness-of-fit factors in this paper, consists in minimizing the *stress* function:

$$S = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N}(f(\delta_{ij})-d_{ij})^2}{\sum_{i=1}^{N}\sum_{j=i+1}^{N}d_{ij}^2}} \tag{3}$$

Interpreting the resulting MDS plot and extracting the underlying perceptual axes governing the signal space can be a frustrating task. The dimensionality of the space under investigation and the exact meaning of its prominent axes cannot be inferred easily from any goodness-of-fit factor or from simple observation of the MDS plot. Allowing the MDS algorithm to use a large number of dimensions will result in a geometrical representation that is matched closely to the data, but which can be very difficult to visualize and interpret. On the other hand, limiting the MDS algorithm to a small number of dimensions might not lead to the most exact representation of the data, but it could offer a more intuitive view of its primary underlying structure. This is of particular importance when dealing with the analysis of perceptual data, which is known to be both very noisy and the product of an intricate system. The mechanisms that govern human sensorial perception are complex and it is unlikely that any MDS analysis will reveal how they work. However, our premise, confirmed qualitatively by past work, is that MDS analysis can still expose the salient perceptual dimensions and provide guidelines to the design of new meaningful and easily distinguishable artificial stimuli.

2.2 Obtaining Data for The Dissimilarity Matrix

The conventional means of acquiring dissimilarity values is to directly compare all possible pairs of stimuli in the set of interest. Subjects are presented with each pair, then asked to rate their degree of similarity. Similarity scores are then converted to dissimilarity scores through a simple transformation. Unfortunately, this paired comparison method suffers from a lack of consistency as the number of pair-wise comparisons increases (e.g. a stimulus set of 25 requires 300 comparisons and a stimulus set of 35 requires 595 comparisons): subjects forget their grading scale, and experiment time lengthens dramatically.

Cluster-Sorting Method: To avoid the problems mentioned above, Ward developed a cluster-sorting technique for an experiment designed to study the salient properties of the physical environment as perceived by humans [8]. He presented a set 20 photographs representing real physical environments to a group of subjects, and asked them to categorize the photographs into different clusters according to a set of rules. The sorting task was repeated a number of times with a different number of clusters for each trial. Dissimilarity matrices were constructed based on the number of times a particular pair of stimuli was sorted into the same cluster. Stimulus pairs that were often grouped together received a low dissimilarity score, whereas pairs that were never grouped together were assigned a maximum dissimilarity score.

While Ward's method was originally developed for visual stimuli, MacLean, Enriquez, Chan and others applied a similar technique to sets of haptic stimuli. Brief computer-generated haptic signals, or haptic icons, were constructed by varying parameters such as frequency, magnitude and shape and were presented to subjects through a force feedback knob [7] or more recently, a vibrotactile mouse [2]. Users were asked to classify the haptic icons into different clusters; the sorting task was repeated five times varying the number of clusters. Results from an MDS analysis of this type of data has demonstrated a separation of the icons that follows intuition while lending extra structural detail. In the knob-generated data set, for example, while frequency seemed to be the salient dimension overall, other perceptual dimensions such as shape also emerged from the MDS plots when only signals pertaining to a certain range of frequencies were considered.

Hollins et. al. have also made use of a group-sorting strategy to obtain dissimilarity data for the study of the underlying perceptual dimensions of tactile sensations [4]. However, in this case, the data for the MDS analysis was the result of a single sorting task during which subjects were asked to classify 17 real-life surface textures in a minimum of 3 but maximum of 7 groups. In a second phase of the experiment, subjects were asked to rate the textures according to different assumed properties (e.g roughness, temperature, flatness) and the resulting scales were mapped to the 3D MDS space obtained in the first phase. Hollins et al. found that two dimensions of the MDS space corresponded closely to a roughness-smoothness and hardness-softness scales respectively. The third dimension could not be related to any of the remaining scales. While valuable for understanding tactile perception, their perceptual data collection strategy differs from ours in its implementation. Moreover, we are more concerned with the understanding of the perception of synthetic computer-generated tactile icons, rather than the existing tactile properties of real objects.

2.3 Objectives and Organization

The strength of the cluster-sorting method lies both in its greater speed of execution - reducing the period over which subjects must retain a calibration - and, we hypothesize, in the more global comparative nature of the sorting task as compared to pairwise comparisons. However, we have recently observed that the method's global nature may also be a weakness, because it creates a complex pattern of correlations among the elements of the dissimilarity matrix. It is the need to understand the impact of this restriction on independence on MDS outcome (which nevertheless seems to show plausible patterns) that motivated the investigation we report here.

Our analysis is illustrated using actual data obtained from a cluster-sorting task using a new tactile device that we have built. Because the device differs from other displays in the way it displays tactile stimuli to the thumb tip, the resultant sensations were novel and subjects should have no pre-existing categorizations.

In Section 3, we present our example data set, including an overview of tactile device and experimental method. Section 4 introduces our key critical questions, and in section 5 we explore each of the question in turn with secondary analyses and discussion. Finally, we close with conclusions and recommendations.

3 DATA: HARDWARE, STIMULI AND PROCEDURE

3.1 Hardware

3.1.1 Tactile Stimulation by Lateral Skin Stretch

Applying a traveling wave of local tangential deformations against the fingertip skin can induce the sensation of a small-scale shape sliding on the finger [3]. During such interaction, no indentation normal to the skin surface occurs. The moving pattern of lateral skin stretch is perceived as a traveling feature. This illusion was previously exploited to display truncated Braille characters to the visually impaired with a prototype device called the Virtual Braille Display (VBD) [6].

3.1.2 Tactile Handheld Miniature Bimodal Device

Following encouraging results obtained with the VBD, we built a miniature version of the tactile display (TD) and integrated it into a prototype for a mobile handheld device. The result, referred to as the Tactile Handheld Miniature Bimodal (THMB) device (Figure 1), comprises a miniature tactile display for the thumb tip and an LCD screen, both assembled inside a small plastic case. Because our focus here is on the data analysis, detailed implementation information is reserved for later publication.

Typically, the THMB device's casing is held in the left palm and is secured in place between the thumb and the four remaining fingers, similarly to how a PDA or cell phone is held (Figure 2). The tactile display consists of a stack of eight piezoelectric benders intercalated between brass rods, which protrudes slightly through a narrow slit on the left side face of the device's case. The user's

Figure 1: THMB device.

Figure 3: Five waveforms used for the experiment.

(a)　　　　　　(b)

Figure 2: Interaction with the device: (a) overview of the THMB interface and (b) close-up on the thumb against the tactile display.

Table 1: Overall animation durations calculated from waveform and speed parameters.

waveform	wavelength	duration (ms) at speed		
(Fig. 3)	(in samples)	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{20}$
TRI	143	480	720	960
ROLL	62	221	331	442
SAW	20	86	130	173
BUMP	16	74	110	147
EDGE	16	74	110	147

thumb tip rests against the summit of the the TD. When activated, the TD's piezoelectric actuators induce lateral skin deformation to the thumb tip by bending.

The amount of bending of a particular piezoelectric actuator is controlled with a voltage applied across its electrodes. A PC host running Linux generates the 8 control signals, one for each piezo-actuator and sends them to interfacing electronics where they get filtered and amplified before being applied across the piezoelectric benders. The resulting control voltages range from $\pm 50V$ and are updated at 3125 samples/sec. They are encoded with a single byte and therefore can only take 256 different values.

3.2 Design of the Haptic Icon Stimuli

Because each of the 8 piezo actuators can be controlled independently, the device may be programmed to display 8-pixel animations, each pixel capable of 256 activation levels. Sets of 8 piezo activation states can be thought of as a frame; a time series of such frames define a given tactile stimulus, which thus varies across both time and space. With the THMB, we wanted to explore alternative shape configurations as well as ways to animate the stimuli over time.

Selection of Stimuli: Using a custom-built prototyping application to explore the stimulus space, we generated a wide variety of waveforms, defined as the displacement of one piezo element as a function of time. We selected five waveforms for further investigation based on the criteria that they produced qualitatively different tactile sensations while remaining as simple as possible. The intent was to begin with a candidate set of canonical basic waveforms referred to in this paper as TRI, ROLL, SAW, BUMP and EDGE (Figure 3). A waveform was animated across the display using a tweening method that consisted of playing the waveform on successive piezo elements with a phase delay of 1 sample. The resulting animation could then be slowed down by reducing the frame rate; or it could

be played backwards, resulting in the sensation of the traveling tactile pattern moving in the opposite direction.

Stimuli: The study used a 5x3x2 combination of stimulus factors, resulting in 30 total tactile animations used as stimuli. The five waveforms mentioned above, three animation speeds, and two amplitudes were varied to produce the patterns. Animation speeds consisted in $\frac{1}{10}^{th}$, $\frac{1}{15}^{th}$ and $\frac{1}{20}^{th}$ of the maximum 3125 frames/sec rate imposed by the hardware. Amplitude could either be full scale (F) or half scale (H). Because some waveforms were longer than others, the total duration of the animation (i.e., the amount of time a stimulus was present under the user's finger) was calculated from the waveform and animation speed. We kept track of this duration meta-parameter throughout the data analysis because it was a readily observable characteristic of the stimuli (Table 1). Stimuli varied in duration between 74ms and 960ms. For the purpose of the data analysis, the stimuli were labeled according to the following indexing: $< \mathtt{waveform} >< amplitude >< duration >$.

3.3 Experiment

3.3.1 Subjects

Ten participants (7 male) aged 19 to 31 were recruited for the experiment; none had any prior experience with the THMB device or similar displays. The participants were told to hold the device in their left hand with their left thumb resting lightly on the tactile display, and were allowed to take breaks as necessary.

3.3.2 Method

The participants used the method described in [7] to conduct the cluster-sorting task, wherein a GUI displayed graphical tiles for each stimulus which, when clicked, played the respective stimulus, and which could be moved about and sorted into the cluster boxes. The software was modified to present the 30 tactile stimuli used in this experiment, and the onscreen tiles representing the stimuli were not marked.

Each participant performed five similar cluster-sorting trials. In the first trial, the stimuli were sorted into a user-selectable number

(from 2 to 15) of clusters. In the subsequent four trials, subjects were required to sort the stimuli into 3, 6, 9, 12, or 15 clusters presented in random order, with the trial containing the number of clusters closest to the user-selected first trial clusters being eliminated. For example, if the user selected 8 groups for the first trial, their subsequent trials would consist of {3, 6, 12, 15} groups in random presentation order.

3.3.3 Data

A similarity matrix was calculated and converted to a dissimilarity matrix in the same way described in [7]. The cluster-sorting method assigns similarity points to a pair of stimuli each time they are put together in the same cluster. The number of similarity points given to a pair of stimuli placed together in a given sorting trial is equal to the number of clusters in that trial. For instance, if two stimuli were present in the same cluster during the trials involving 3 and 9 clusters, the pair is assigned 12 points $(3 + 9)$. For the first trial for which subjects get to choose the number of clusters, the amount of points alloted is adjusted to fit the closest integer of the set {3, 6, 9, 12, 15}. The similarity points are summed over a subject's trials and a simple inverse linear transform is applied to obtain a matrix with dissimilarity scores that range from 0 to 1000. A score of 0 indicates that the pair of stimuli always appeared in the same cluster across all trials, whereas a score of 1000 indicates that they were never paired.

3.3.4 Results

Dissimilarity matrices from all subjects were combined to create an average dissimilarity matrix (shown in Table 2) and the data was submitted to a 2D MDS analysis using SPSS^{TM} 13.0 (Euclidean distance algorithm with ordinal/untied data [1]). Figure 4 depicts the stimuli's resulting spatial arrangement, upon which three groupings have been manually overlaid by the authors. Orientation of the graph and its axes returned by SPSS^{TM} are irrelevant since MDS analysis is only concerned with the relative positions of the stimuli in space, rather than their absolute positions. Therefore, the plot has been rotated for clarity. Five lab colleagues uninvolved in the project were also provided with the plot without stimulus labels and asked to group the stimuli into as many natural clusters as desired based on spatial layout; similar configurations resulted in all cases. MDS plots in 3D were also generated but no extra structural information could be extrapolated.

To consider sensitivity of results to MDS algorithm, the individual subject matrices were also run through a replicated MDS (RMDS) algorithm. The resulting plots and goodness-of-fit stress factors for the the non-metric RMDS analysis were similar (but not identical) to the ones reported in this paper.

Subjects assigned labels to the different sorting boxes based on what we have reduced to four essential categories: duration (e.g. quick, very long, slow, short, fast), magnitude (e.g. moderate, weak, medium, strong), multiplicity (e.g. single, double, multiple, one, two) and description (e.g. rhythmic, beat(s), thud, heart beat, broken, vibes, sensation). While not immediately analyzable, these categories may lend insight into the perceptual dimensions subjects used in their classifications.

3.3.5 Preliminary Discussion

A few features are evident through inspection of Figure 4. The stimuli seem clustered naturally in three major groups ranged along the x-axis, with the shorter BUMP, SAW and EDGE signals to the left, ROLL

Table 2: Average dissimilarity matrix obtained from all 10 subjects' dissimilarity matrices.

Figure 4: Ordinal untied MDS plot from the average dissimilarity matrix obtained with the cluster-sorting experiment ($stress = 0.083$). Signal durations (from Table 1) have been rounded to the nearest multiple of 10ms for clarity.

in the middle and TRI to the right. There thus seems to be some influence of either pattern and/or overall duration on the major grouping; however, a duration argument may be at odds with the longer groups' internal structure. For the TRI and ROLL groups, duration (221-960ms) increases on the y-axis, while for the BUMP/SAW/EDGE cluster, overall duration seems roughly aligned horizontally. One of the longest ROLL stimuli is associated with the TRI cluster.

Neither amplitude nor duration visibly affect the TRI stimuli; whereas both the ROLL and BUMP/SAW/EDGE clusters seem strongly differentiated by amplitude (on the y-axis) and slightly by duration (on the x-axis).

The confounding of the relatively brief SAW, EDGE and BUMP stimuli is at least partially due to electromechanical filtering of the hardware; wave shape detail at these durations was later confirmed to be beyond the device's temporal display resolution.

While interesting and clearly meaningful at some level, the complex structure of these results immediately raised questions. For example: what is the actual perceptual *amplitude* of the differences found between these signals – can it be quantified? The y-axis seems to be interpreted as amplitude for the shorter signals, and as duration for another cluster which happens to contain the longest signals. Is this dimension perceptually contiguous, or is it used independently by each cluster? These uncertainties lead to closer examination of the data collection technique, and the more detailed analysis that follows.

4   GUIDING QUESTIONS

The cluster-sorting method for collecting perceived dissimilarity data is a framework that drastically decreases the time required to obtain a dissimilarity matrix from a large set of stimuli. Rather than requiring subjects to gradually construct a personal dissimilarity scale through successive direct comparisons between all possible pairs of stimuli – to give an order of magnitude, the experiment reported in section 3.3 would have required 435 direct comparisons – it provides them with an implicit reference scale that facilitates consistency in their answers. On the other hand, the resulting dissimilarity matrix is the outcome of a process with strong structural constraints where, at the same time, there may exist multiple grouping strategies which would express the same degree of dissimilar-

ity. Our curiosity about how the resulting interdependence in the dissimilarity matrix might impact the MDS outcome included these questions:

Q1 When applying the cluster-sorting technique, what assumptions are made concerning the structure of the collected data?

Q2 How does a dissimilarity matrix obtained from the cluster-sorting technique differ from a dissimilarity matrix constructed from a set of classic pairwise comparisons?

Q3 Does data obtained via the cluster-sorting technique preserve the structural information about the stimulus space at every level (for each sub-space)? Similarly, if one perceptual dimension is far more prominent than the others, will the other minor dimensions still be represented accurately in the dissimilarity matrix?

Q4 How is noise in the data-gathering process reflected in the dissimilarity matrix obtained with the cluster-sorting method?

In the following section, we examine each of these questions with further discussion and where possible, analysis of the cluster-sorting technique itself, data collected with it, or both.

5   ANALYSIS OF CLUSTER-SORTING METHOD

5.1   Assumptions on the Nature of the Dissimilarity Data Collected

MDS algorithms differ by assumptions made regarding input data distributions. Therefore, it is important to clarify relevant attributes of the data obtained with cluster-sorting technique. MDS algorithms are either metric or non-metric. Metric algorithms try to map the stimuli in a Euclidean space by keeping the distances between the stimuli as much like their corresponding dissimilarity scores as possible. Non-metric algorithms focus more on preserving the rank of the dissimilarity scores.

Therefore, by applying a metric algorithm to the data obtained from the cluster-sorting method, we formulate the assumption that the connection strength between two stimuli grouped together is linearly dependent on the number of clusters present. For instance, an association between two stimuli grouped in a sorting trial with 15 clusters is considered to be 5 times more powerful than a association between any other two stimuli grouped in a sorting trial with 3 clusters. While both Ward and MacLean obtained positive results when they applied a metric MDS analysis to their data, it is hard to demonstrate that a metric assumption is justified at all times.

On the other hand, the data in the dissimilarity matrix can also be considered to be non-metric. In this case, it is is assumed that summing the similarity points that accrue over successive sorting trials does not destroy the monotonicity of the data. This weaker assumption relaxes the constraints on the MDS analysis. However, an MDS analysis now results in a graphical representation that can only give an idea of the extent to which stimuli differ, without providing as much quantitative meaning to the graphical inter-stimuli distances.

Results reported in 3.3.4 are produced with a nonmetric analysis on the assumption that the data is ordinal. We also considered results generated with different metric MDS algorithms (including the INDSCAL algorithm). The MDS metric plots shared some similar features with their nonmetric counterparts, such as proximity in space of all the TRI stimuli. However, the plots did not exhibit any clear visual clustering of the stimuli into the 3 major groups described in the discussion section 3.3.5 and the stimuli appeared to be more continuously spread out across the perceptual space, making it harder to interpret the results. Without invalidating any
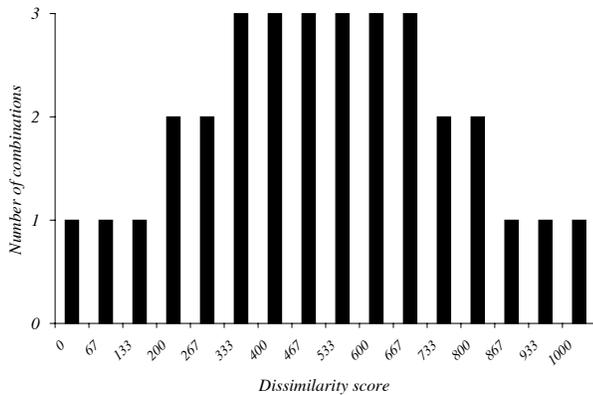
Figure 5: Number of possible combinations to obtain a single dissimilarity score with a cluster-set of $\{3, 6, 9, 12, 15\}$.

Subject index	Nb. init. clusters	Nb. of 1000's	Nb. of 0's
1	6	210 (48%)	0 (0%)
2	5	241 (55%)	2 (0%)
3	8	245 (56%)	2 (0%)
4	5	238 (55%)	9 (2%)
5	12	253 (58%)	3 (1%)
6	8	233 (54%)	4 (1%)
7	2	203 (47%)	1 (0%)
8	7	269 (62%)	1 (0%)
9	6	258 (59%)	8 (2%)
10	6	207 (48%)	5 (1%)

Table 4: Key properties of the average dissimilarity matrix. The average dissimilarity matrix, shown in Table 2, is formed by averaging each score element of the 10 subjects' matrices. Percentages in parentheses indicate a proportion of the average dissimilarity matrix's 435 elements.

	Nb. of 1000's	Nb. of 0's
Avg. Matrix	107 (25%)	0 (0%)

metric analysis of the data, our results indicate that perceptual data obtained using the cluster-sorting technique seems best assumed to be ordinal.

5.2 Dissimilarity Matrix Properties

A matrix obtained from a set of classic pair-wise comparisons contains score-elements that are independent of each other. Each dissimilarity score reflects a dissimilarity intensity that is in relation to the others, but changing one score element won't affect the other elements of the matrix. On the other hand, a dissimilarity matrix obtained with the cluster-sorting method is structurally constricted. One cannot re-adjust a single score for a particular pair comparison, without having to readjust other elements of the matrix. This is a direct consequence of the requirement for a minimum of one-stimulus-per-cluster imposed by the experimental procedure and the fact that assigning a stimulus to a cluster also means that it is not put into any of the other clusters of the same trial.

Dissimilarity scores obtained with the cluster-sorting method have one of a set of discrete values. For the cluster sets used here, these vary from 0 to 1000 with a resolution of 66.6 (Figure 5). For 10 of the 16 possible cases, there exists more than one combinatorial way to obtain the same score. Moreover, the random distribution of the set of score values across the entire dissimilarity matrix is not a normal distribution since the matrix elements (i.e., all the different pairs) are not independent.

This was observed in post-analysis and shown in Table 3, which displays the number of clusters actually chosen by our subjects on the first sorting trial, as well as the number of maximum dissimilarity scores of 1000 and of 0. The latter provide some indication of the dissimilarity matrix structure: a score of 1000 indicates two stimuli that were never grouped together across the five trials whereas a score of 0 indicates that they were always paired. Across all 10 dissimilarity matrices, 20 associations (ranging from 2 to 4 stimuli in size) with scores of 0 were made and all but one involved SAW, EDGE and BUMP stimuli exclusively - for instance, subject #4 systematically put the stimuli BUMPF74, BUMPF110 and EDGEH147 in the same clusters across all trials. Moreover, only 2 out of these 20 0-dissimilarity associations included stimuli of different amplitudes (one of the exceptions is the example given above for subject #4). Conversely, the average dissimilarity matrix indicates that 107 of the 435 possible pairs that could have been formed across the five trials were never made (Table 4).

The high occurrence of dissimilarity scores of 1000 is easily explained. Over all five sorting trials, numerous pairs of stimuli are never grouped together and therefore are assigned a maximum dis-

similarity score. Caution must be taken when interpreting a 1000 score. It cannot represent infinite dissimilarity because there is no absolute zero of perceptual similarity. Humans are very good at finding associations between items, irrelevantly of the amplitude of the contrast between the items presented. However, it can either stand for the maximum non-infinite difference between two stimuli among the set of stimuli displayed, or simply represent a measure of uncertainty of the data. The latter occurs, for instance, if not enough cluster-sorting trials have been run. This dual nature of 1000 scores strengthens the argument that an MDS analysis should assume cluster-sorted data are ordinal/untied.

The actual distribution of scores obtained through a simulation of completely random sorting for this method is shown in Figure 6[2]. Also shown is the score distribution for a typical subject who chose 6 clusters for the initial trial, and whose individual MDS result exhibited a clear and plausible clustering.

In both cases, high dissimilarity scores are much more present than low dissimilarity scores, with a peak occurrence of the maximum dissimilarity score of 1000 (accounting for more than 40% of all the scores in the matrix). Interestingly, the random distribution is almost monotonic, except for a score of 867 that occurs less often than a score of 800, explained partially by the fact that there exists more combinatorial ways for a stimulus pair to obtain a score of 800 than a score of 867.

The first trial allowed subjects to choose a preferred number of sorting clusters. This extra degree of freedom in the experimental procedure is intended to expose what subjects feel is the natural number of clusters for the set of stimuli presented (and which shows substantial and interesting variety, as documented by Table 3). However, it adds yet another degree of complexity to the analysis of the dissimilarity data. Figure 7 illustrates how the choice of the initial number of clusters affects the random distribution of the 1000-scores across the matrix. We chose to show the effect on the 1000-score frequency since it is the highest, but it is easy to understand that the entire random score distribution is depen-

---

[2]The random distribution was generated by averaging the results of 10000 iterations of the replication of a random classification of the stimuli following the sorting rules described in 3.3.3.
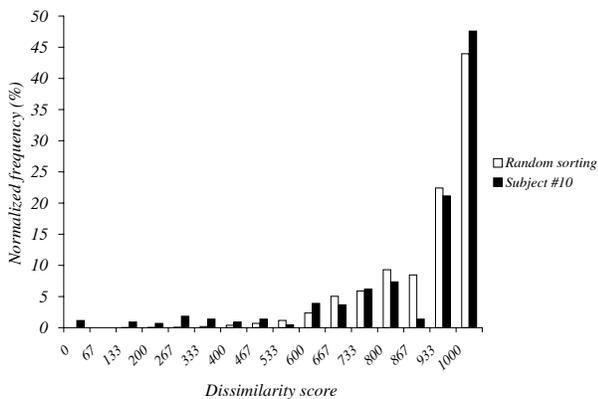
Figure 6: Score distributions for a typical subject and random sorting. Frequency indicates a proportion of the 435 score-elements present in a 30-stimulus dissimilarity matrix.
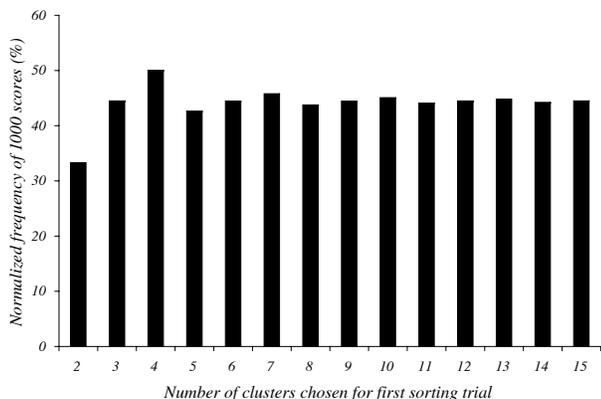
Figure 7: Average normalized frequency of 1000-scores obtained in a randomized cluster sorting execution, as a function of the number of clusters chosen for the first trial. Frequency indicates a proportion of the 435 score-elements present in a 30-stimulus dissimilarity matrix.

dent on the initial number of clusters. Of all the sorting trials, the one with the least number of clusters affects the distribution of the 1000-scores across the matrix the most because it forces a higher average number of stimuli per cluster. Therefore, the dissimilarity matrix obtained from a subject who selected two clusters for the initial sort will contain a wider distribution of scores than that of a subject who selected four clusters.

These results indicate that a validation of the subjects' data – for instance obtaining p-values that denote with high probability that the subjects's clustering were not random – is not a trivial task because the random distribution is ill-defined. This also suggests that comparing the scores of the individual dissimilarity matrices through standard techniques such as standard deviation might not be relevant, and should probably not be used to discard subjects.

For the analysis reported here, we validated the individual dissimilarity matrices obtained by trying to find any flagrant inconsistencies in subjects' answers. For instance, we examined the individual MDS plots of each subject. We also compared the number of 1000's between the subjects individual matrices and the random distribution. These inspections did not reveal any inconsistency and therefore we made use all of our subjects' data for our analyses.

Further work is required to assess the significance of the dependencies among the matrix cell values, the discretization of the scores and the unusual shape of the score distribution because they

are likely to have implications on the results and validation of any analysis (MDS or other).

5.3 Capture of MDS Group Internal Structure

For our example results, the MDS plot's y-axis seems to absorb different stimulus design parameters for different large groupings (Section 3.3.5). This made us wonder about the extent to which the cluster-sorting method can capture and reveal sub-dominant stimulus distinctions. Clearly, a 2D MDS analysis cannot provide enough detail to resolve group internal structure (if it varies from group to group) because of the dimensionality restriction it imposes on the algorithm; but does a dissimilarity matrix acquired with the cluster-sorting method nevertheless contain the more subtle distinctions specific to a given major group found by a low-dimensional analysis? If present, can it be recovered without resort to a less-intuitive, higher-dimension MDS analysis? To address these questions, we ran a second experiment.

Method: The experiment protocol was similar to the one described in 3.3.2, with the following differences and key parameters.

We tested five subjects, none of whom had taken part in the first experiment. For the first part of this experiment, subjects were again asked to classify the entire set of 30 stimuli, but this time with a cluster-set of $\{5, 10, 15\}$ rather than the set used for 3.3.2. In the second part, subjects proceeded to execute the same classification task three more times, on the 6 stimuli of three different particular waveform types (TRI, ROLL and EDGE, respectively) and for a cluster-set of $\{2, 3, 4\}$.

Data: From the first part of the experiment, we generated a control dissimilarity matrix (sized $30x30$) from which we isolated the three $6x6$ submatrices belonging to the subsets being examined for internal structure. From the second part, we produced three full $6x6$ dissimilarity matrices, one for each waveform type. We then ran a 2D MDS analysis for the $30x30$ control matrix and each of these six $6x6$ matrices, and compared the resulting plots.

Results and Discussion: The resulting 2D MDS plot for the control dissimilarity matrix (Figure 8(a)) is very similar to our MDS plot obtained with the original experiment (Figure 4). Since these were derived from cluster sorts using different cluster-sets, this suggests that cluster-set does not strongly impact the outcome of the MDS analysis.

Moreover, MDS plots (not shown) generated from submatrices of the $30x30$ control matrix strongly resemble their intra-waveform matrix counterparts ((b), (c) and (d) of Figure 8). Specifically, all graphs display a clear spatial demarcation between stimuli of half amplitude and stimuli of full amplitude. Furthermore, most of the gradation in duration as well as its orientation present on the intra-waveform plots is also visible on the submatrix plots. Thus, our data seems to offer evidence that the cluster-sorting method does capture the underlying structure of non-dominant distinctions.

A key observation is that these subtleties are not apparent from inspection of the global MDS plot based on the control $30x30$ dissimilarity matrix. This suggests that re-running the MDS algorithm on sub-sets of the entire stimulus space (thus relaxing the *stress* on the algorithm when it is required to find consistent dimensions for an entire diverse set of stimuli) can offer valuable information on lower structural levels of the entire stimulus set that are not visible from the global MDS plot.

5.4 Effects of Noise

Perceptual experiments always suffer from noisy data due to subject fatigue and calibration drift. The effects of perceptual noise on data obtained from a cluster-sorting experiment are hard to quantify, even if assumed to be random, because a fixed level of randomness will weigh differently on the overall noise from one trial

△ *TRI*	✛ *ROLL*	⊟ *EDGE*	◇ *SAW*	○ *BUMP*

□ *half amplitude*   ■ *full amplitude*   *number: duration (x10ms)*
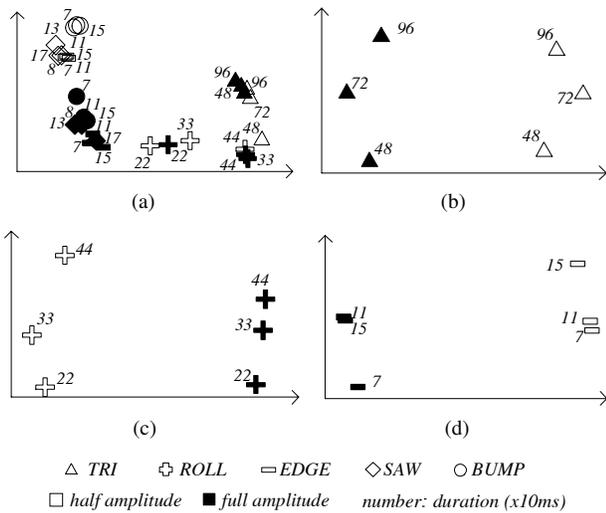
Figure 8: Ordinal untied MDS plots obtained from grouping the different waveforms into cluster-sets of {2,3,4}: (a) control with all 30 stimuli and with isolated (b) TRI, (c) ROLL and (d) EDGE stimuli.

to another. The similarity point assignment process tends to amplify noise for trials with a large number of clusters. For instance, the noise-induced erroneous sorting of a stimulus when 15 clusters are present is more significant than with only 3 clusters because an improperly matched pair is assigned 15 similarity points in the former case and only 3 in the latter. To complicate things, it can be argued that a 15-cluster discrimination task will generate more uncertainty than a 3-cluster task because it imposes a heavier cognitive load on the subject. If this is true, it would mean that the most noisy part of the data is also precisely the one that accounts for the largest similarity scores (or lowest dissimilarity scores).

Despite the fact that we are currently unable to account for the level of noise in the data, we are not too alarmed by this potential problem since, as reported in 5.2, our results were reproducible. This suggests that, even though noise might get amplified for the trials with a large number of clusters, the initial amount of noise generated by applying the cluster-sorting technique is not important enough to significantly distort the dissimilarity data. However, if this was to become a problem in future experiments using the same technique, one possible solution would consist in applying a corrective factor that decreases the weight of trials containing a large number of clusters.

6 CONCLUSIONS AND FUTURE WORK

The advantages of the cluster sorting algorithm for generating perceptual MDS input data are significant: in particular, it is superbly efficient in both time and consistency of subject response. The concerns we have identified are at minimum important to be aware of. Based on the reported investigation, advantages seem to outweigh the concerns, and it appears justifiable to continue using this technique with well-informed caution. With respect to each of our Section 4 questions, we specifically conclude:

Q1 Perceptual data obtained using this cluster-sorting technique is best assumed to be ordinal, and analyzed with a non-metric MDS algorithm.

Q2 Key characteristics of a cluster-sorted dissimilarity matrix, in particular a restriction on element independence, may impact MDS output. Further investigation is required to quantify both

the magnitude of this restriction, and the robustness of MDS algorithms to such inter-matrix correlations.

Q3 Hidden patterns in the resulting plot of a MDS analysis carried out on the entire stimuli set can become apparent when selected submatrices of the full dissimilarity matrix are submitted to the same MDS algorithm. This tends to indicate that the cluster-sorting technique also captures detailed information about sub-level of stimuli distinction that are not visible by sole inspection of the global MDS plot.

Q4 It is difficult to make sense of the distribution of noise across the dissimilarity matrix obtained from the cluster-sorting method (mainly due to noise amplification for trials with a large number of clusters). However, the fact that we were able to replicate our results over two independent experiments seems to indicate that the data-gathering technique is not generating enough noise to significantly distort the data.

This work represents the first step in an ongoing effort to develop a suitable mechanism for analyzing the perceived differences and usability of haptic icons. Our results for questions 2 and 4 in particular need further development. Finally, we are also considering alternatives to MDS for perceptual analysis of haptic icons.

REFERENCES

[1] S. A. Brewster and L.M. Brown. Tactons: Structured tactile messages for non-visual information display. In *5th Australasian User Interface Conference (AUIC2004), Dunedin.*, pages 15–23, Dunedin, Australia, 2004. Read: 2005-06-01.

[2] A. Chan, K. E. MacLean, and J. McGrenere. Learning and identifying haptic icons under workload. In *Proc. of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. WHC 2005, Pisa, Italy, 2005.

[3] V. Hayward and M. Cruz-Hernandez. Tactile display device using distributed lateral skin stretch. In *Proc. of the Haptic Interfaces for Virtual Environment and Teleoperator Systems Symposium*, volume DSC-69-2, pages 1309–1314. ASME IMECE2000, Orlando, Florida, 2000.

[4] M. Hollins, R. Faldowski, S. Rao, and F. Young. Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis. *Perception & Psychophysics*, 54:697–705, 1993.

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

[6] V. Levesque, J. Pasquero, Hayward V., and M. Legault. Display of virtual braille dots by lateral skin deformation: Feasibility study. *ACM Transactions on Applied Perception*, Vol. 2, No. 2:pp. 132–149, 2005. Read:.

[7] K. MacLean and M. Enriquez. Perceptual design of haptic icons. In *Eurohaptics*, Dublin, Ireland, 2003.

[8] L. Ward. Multidimensional scaling of the molar physical environment. *Multivariate Behavioral Research*, 12:23–42, 1977.

[9] F.W. Young. *Encyclopedia of Statistical Sciences*, volume 5, chapter Multidimensional Scaling. John Wiley & Sons Inc., 1985.