HELSINKI UNIVERSITY OF TECHNOLOGY

Department of Mechanical Engineering

David Nicholai Johnson

Sensory Testing and Mechanical Perceptions of Quality – A Novel Application of Quick Individual Vocabulary Profiling

Masters Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering.

Espoo, May 30th, 2006

Supervisor: Professor Kalevi Ekman
Instructor: M.Sc. Päivi Kuusio

HELSINKI UNIVERSITY OF TECHNOLOGY — ABSTRACT OF THE MASTER'S THESIS

Author:	David Nicholai Johnson
Title of the Thesis:	Sensory Testing and Mechanical Perceptions of Quality – A Novel Application of Quick Individual Vocabulary Profiling
Date:	29 May 2006 **Number of Pages:** 58
Department:	Department of Mechanical Engineering
Professorship:	Kon-41 Machine Design
Professor:	Professor Kalevi Ekman
Instructor:	Päivi Kuusio, M.Sc. (Tech.)

PURPOSE: To create a method for evaluating and understanding perceived mechanical quality in products. While the method can be applied to any mechanical quality purpose, this thesis addresses the example of bistable slide mechanisms in mobile phones.

MATERIALS AND METHODS: Proposed herein is a method for evaluating perceived mechanical quality using sensory panels and quick individual vocabulary profiling, a variant of sensory profiling. Included in the proposed method are sensory evaluation theory, testing procedures, and data analysis methods. The main result of this thesis is a conclusion of whether or not quick individual vocabulary profiling is an appropriate tool for evaluating perceived mechanical quality.

RESULTS: After completing a complete iteration of sensory profiling using a panel of 13 persons, the thesis has shown that quick individual vocabulary profiling is an appropriate and reasonable means of evaluating and understanding perceived mechanical quality. In addition, this implementation of quick individual vocabulary profiling gives relatively quick and concise results. Furthermore, the results show with a reasonable degree of certainty that a degree of consensus was achieved between different assessors and some clear trends have emerged.

CONCLUSION: Quick individual vocabulary profiling combined with various analysis techniques has been conclusively proven to be a possible solution for evaluation perceived mechanical quality of bistable slide mechanisms in mobile phones.

Keywords:	Sensory Profiling, Flash Profile, Perceived Mechanical Quality, Principal Component Analysis, Subjective Testing, Mobile Phones

1. INTRODUCTION 7

1.1 Meandering, the wrong turn, and the road less traveled. 7

1.2 Goals 9

1.3 The importance of perceived mechanical quality 9
1.3.1 Perceived mechanical quality and actual quality 10

2. BACKGROUND & THE STATE OF THE ART 11

2.1 Sensory Evaluation 11
2.1.1 Assessor Types 11
2.1.2 Sensory Evaluation Scenario 12

2.2 Sensory Profiling Methods 13
2.2.1 Consensus Vocabulary Methods 13
2.2.2 Free Choice Profiling 14
2.2.3 Repertory Grid Method 15
2.2.4 Flash profile method 16

2.3 Data Analysis & Interpretation 17
2.3.1 General Procrustes Analysis 17
2.3.2 Principal Component Analysis 19

2.4 Internal and External Preference Mapping 23

2.5 State of the Art 24
2.5.1 Research using Consensus Vocabulary 24
2.5.2 Research using Individual Vocabulary 25
2.5.3 Sensory Profiling Applied to Mechanics 26

3. EXPERIMENTAL PROCEDURE 28

3.1 Sensory Panel Composition and Selection 28

3.2 Experiment Environment 28

3.3 Phones in Study 29

3.4 Testing Process 30
3.4.1 Sessions One and Two 30
3.4.2 Sessions three and four 33
3.4.3 Overall Process Diagram 34

4. RESULTS 35

4.1 Agglomerative Hierarchical Clustering 35

4.2 Principal Component Analysis correlation Loadings 37

4.3 Internal Preference Mapping 41

4.4 External Preference Mapping **43**
4.4.1 Attribute Listings 46
4.4.2 Attribute Type Breakdown 47

5. DISCUSSION **49**

5.1 Lingual hiccups in the process **50**

5.2 Possible Improvements **50**

6. CONCLUSIONS **53**

7. REFERENCES **55**

8. APPENDIX **59**

8.1 Appedix A **59**

8.2 Appendix B **60**

PREFACE

This Master's Thesis was written for Nokia Research Center in Helsinki, Finland.

I would like to express my sincere gratitude to the examiners of this thesis, Professor Kalevi Ekman of Helsinki University of Technology, Pekka Pihlaja from the Nokia Research Center, and M.Sc. Päivi Kuusio from Nokia Research Center. Without their guidance and support, this thesis would never have been possible.

Additionally, I would like to thank M.Sc. Gaëtan Lorho for his influence on the direction and execution of the thesis. He provided guidance and direction during the early development of the topic and was instrumental in the adoption of sensory profiling as a vehicle for evaluating perceived mechanical quality. Furthermore, his knowledge and support with regard to analysis techniques was critically important in facilitating the completion of my research.

David Nicholai Johnson

May 30th, 2006

ABBREVIATIONS AND ACRONYMS:

AHC	Agglomerative Hierarchical Clustering
ANOVA	Analysis of Variance between groups
ENSIA	Ecole supérieure des industries agricoles et alimentaires
GPA	Generalized Procrustes Analysis
GP3	GuineaPig 3 Subjective Audio Testing Software
ID	Identification
KVL	Danish Royal Veterinary and Agriculture University
LCD	Liquid Crystal Display
nPLS	Multilinear Partial Least Squres regression
PCA	Principal Component Analysis
PC	Principal Component(s)
QDA	Quantitative Descriptive Analysis
SVD	Singular Value Decomposition
XLSTAT	Statistical Software for Microsoft Excel
UI	User Interface

1. INTRODUCTION

1.1 Meandering, the wrong turn, and the road less traveled.

Initially, the goal of this thesis was to broadly investigate mechanical perceived quality. The idea came from a discussion regarding the dip in popularity of Mercedes automobiles and their resulting financial results. A statement was made to the effect that, previously, one could identify a Mercedes simply based on the sound of the door closing, whereas recently the build quality had been lacking. While it sounded crazy, this triggered a train of thought… What is this concept of perceived mechanical quality and what does it mean for design? Does the sound of a door closing actually affect the utility that the user derives from the automobile? No, of course not - but it creates a perception of quality that affects the user's satisfaction and sense of contentment with the product. Initially, this thesis started with the idea that perceived quality might be driver for initial customer satisfaction, and something that should be actively sought after in the product creation process.

Perceived quality, as such, is literally the perception of quality. The question then remains, how can we investigate perceived quality? One simple and straightforward way that was pursued early on in the process was simple, straightforward hedonic testing. Hedonic testing is a term used in psychometric circles to refer to a preference test employing average individuals without any special training, where the following question is posed – "Given this set of products, how would you rank them according to preference?" Hedonic Testing is an excellent way to test people's perceptions and preferences within a limited set of stimuli. (OMahony, 1986) Obviously, hedonic testing only tells us about people's affinity, or lack thereof, for a given set of stimuli. As such, information gleaned from hedonic testing cannot be applied to stimuli outside of this limited set. Another problem with simple hedonic testing is that, while we know some of the characteristics of the stimuli set and we can relatively easily determine the group's response to them – we have no idea why people prefer one over the other. In other words, we know what people like but have no solid basis for saying why they prefer one stimulus over another.

Given the limitations of hedonic testing, how might we go about pursuing the research of perceived quality? Perhaps the best way to begin to explore perceived quality is to understand the transfer function between products and preferences. Such a transfer function is dependant on how preferences vary from individual to individual, and also how perceived product characteristics vary from product to product. This thesis will strive to investigate both sides of mechanical perceived quality; perception being the human side and characteristics forming the product side.

Obviously, the list of factors affecting perceived quality is nearly limitless. A short, terse list might include mechanical feedback in mechanisms, general appearance, weight, material wear characteristics, brand, previous experience, anecdotal evidence, surface finish, color, shape, ergonomics, etc… Because the topic of perceived quality is prohibitively large, this thesis will focus mainly on the characterization of mechanical feel in mobile phones with a bistable slide mechanism. This thesis will approach the problem using a sensory panel to objectively evaluate product characteristics. This subset of products was chosen because of its prevalence in the mobile phone market, availability, and the hope that there might be sufficiently large variation between products for descriptive vocabulary generation. After extensive literature review, sensory evaluation using quick individual vocabulary profiling was selected as the method of choice for the characterization of perceived mechanical quality.

1.2 Goals

The goal of this thesis was initially to study perceived mechanical quality. However, since very little similar research relating to perceived mechanical quality has been released – one of the goals of this thesis was to identify and test a candidate method for evaluating perceived quality. The method selected is a rapid, individual vocabulary type of sensory profiling known as quick individual vocabulary profiling. Strictly speaking, the primary deliverable for this thesis is not an overview of perceived mechanical quality relative to the eight selected mobile phones, but rather an evaluation of whether or not quick individual vocabulary profiling is appropriate for evaluating perceived mechanical quality. Measures of this appropriateness might include success or failure in generating attributes, numbers of relevant attributes and their descriptive ability, descriptiveness and convergence of correlation loading plots, and saliency of resulting preference maps.

This thesis will study perceived mechanical quality of eight bistable slide mobile phone models. Bistable slide phones are mobile phones which have two main parts which slide relative to each other. The bistable portion of the name comes from the fact that these sliding mechanisms have an integrated spring or springs to assist motion and to prevent unwanted opening. Bistable slide phones were selected because of their subltly complex tactile and haptic feedback, availabililty, and prevalence in the current marketplace.

1.3 The importance of perceived mechanical quality

As more and more players enter the product market, there is an ever increasing demand for the consumer’s attention. As such, the need to differentiate one’s products from the competition is more important than ever, and better styling is simply not enough to win consumers. Increasingly, as people become more demanding consumers, the need for products delivering a ‘complete experience’ is becoming critical to being successful. For example, it is simply not acceptable to provide a beautiful and functional product that feels cheap in the hand and creaks and squeaks when used. Perceived mechanical quality research is one way to better understand the underlying perceptions and preferences of the consumer.

1.3.1 Perceived mechanical quality and actual quality

Plastics no longer have the novel and space age associations they carried in the middle of the 20th century. Their widespread use in products and cheapness mean that consumers have become jaded about some of the characteristics of plastic. For instance, if a person describes a mobile phone as having a ‘plastic feel’ overall, the implied connotations are not unanimously positive. However, by adjusting the perceived mechanical quality of such products, we can use the same materials and production techniques and end up with a product that elicits a much better response from consumers. It should be noted, again, that perceived mechanical quality has very little to do with actual quality. One can imagine a product with excellent mechanical quality that conveys an air of cheapness, or conversely a poor quality product might pass itself off as having good perceived mechanical quality. As such, it is critical to understand what factors are behind good or bad perceived mechanical quality.

2. BACKGROUND & THE STATE OF THE ART

2.1 Sensory Evaluation

Sensory evaluation is a science wherein one or more individuals attempt to define and quantify the physical attributes of a product. For example, a company launching a new type of pudding might use a sensory evaluation panel to help develop and tune the different components in the taste and the consistency of the product. Generally speaking, the food science and flavor industries have been especially active in the development and deployment of sensory evaluation over the decades. In its infancy, sensory evaluation was used primarily within food sciences, but recently, it has spread to other fields of research.

Typically, sensory evaluation includes the use of a sensory panel or large pool or separate individuals who work towards the common task of gaining a deeper understanding of product characteristics. Obviously, the use of sensory panels comprised of more than one individual allows a broader and more accurate determination of the sensory characteristics of the product or stimuli at hand. Multiple assessors tend to make up for the blind spots present in other assessors, and can interact dynamically to evolve the collective knowledge. As such, multiple assessors working together as a panel has been the most common way of performing sensory evaluation.

2.1.1 Assessor Types

Within sensory evaluation panels, there are standardized definitions for three different types of assessors: naïve, selected, and expert (ISO 1993, 1994). Naïve assessors are those who never participated in any sensory evaluations. Selected assessors are those who have already participated in sensory evaluation and have undergone some training to help them to be more effective and consistent in their sensory evaluations. Finally, expert assessors are assessors who have shown a high aptitude for sensory evaluation, have well developed memory (to allow comparison from previous experiences with similar stimuli), and may have special background in fields related to the stimuli set.

2.1.2 Sensory Evaluation Scenario

In order to illustrate how and why sensory profiling is used, it might be helpful to create a hypothetical scenario. Let us suppose that there is a brewery wishing to expand its product line by creating new beers. This brewery could benefit from using sensory profiling for several reasons; first and foremost a sensory panel would provide descriptive terms for characterizing new beer flavors. These descriptive terms could then be used, for example, in advertising campaigns or on packaging to communicate these characteristics to the customers. The descriptive terms, or attributes as we will refer to them, might also be used by the brewers themselves to understand what components of taste exist and how the ingredients are affecting the overall taste.

Additionally, one problem that arises is variation in raw ingredient characteristics, and this is especially important in food related industries. A sensory panel could help to pinpoint and quantify these taste variations, thereby allowing corrective action to be taken. Employing a sensory panel allows a more consistent product over time. In effect, sensory panels allow us to take sensory snapshots of products and to reproduce them consistently.

Finally, by pairing a sensory panel with information about consumer’s preferences, we can better understand the market. This helps to identify latent needs existing in the market, and could help the brewery to create a beer targeting an unrealized market segment.

In order to accomplish all of these things, there are some prerequisites for sensory profiling. The panel should be composed of panelists who have the ability to distinguish fine details of taste, smell, and color and are able to readily describe these things. Additionally, the panelists need to feel sufficiently passionate about the product or products and be willing to do their part to characterize it. Typically expert panelists are very good at remembering sensory characteristics and can be very sensitive to small variations in those sensory characteristics (Zamora et al, 2004).

Using the most common method of sensory profiling, consensus methods, the process might proceed as follows:

- The sensory panelists would meet as a group and taste the beer.
- Together, the panel would create attributes to describe the appearance, taste, mouthfeel, and smell of the beer and discuss in detail exactly the meaning of these attributes.
- These discussions might occur with or without a moderator, and depending on panel size and structure, the panel might be broken into one or several groups. Regardless, the panel selects attributes which will be common to all assessors.
- Assessors would be trained and evaluated in the proper use of the selected attributes.
- Finally, each of the panelists would individually rate the beer according to the attributes created by the panel.

This type of sensory profiling is described in more detail in Section 2.2.1.

2.2 Sensory Profiling Methods

2.2.1 Consensus Vocabulary Methods

Conventional Profiling is a very common method that requires skilled, trained assessors and a heavy investment of both time and effort. Conventional profiling is a sensory evaluation method based on consensus vocabulary, meaning that all assessors use the same attributes to describe the stimuli. In order to facilitate this, the panel must first meet as a group to generate a list of attributes. Then, as individuals or in smaller group sessions, the assessors are trained in what exactly each attribute means and how to apply them. This is ideally done by citing examples for each attribute. A good example of how this can be accomplished is illustrated in a study involving the perception of luxury and lighter sounds by Lageat, Czellar and Laurent (Lageat et al, 2003). Because the study dealt with sound, lighter sounds were recorded and could be relatively easily modified to represent low and high anchors for attributes using sound editing software (Lageat et al, 2003). Following this attribute training phase, the assessors evaluate the actual set of stimuli over one or several sessions. Evaluating the products more than once is frequently carried out to allow checks for repeatability and consistency of attributes meanings. Quantitative Descriptive Analysis is one very common scheme for carrying out conventional sensory profiling and is available as a set of guidelines for sensory profile trials (Stone et al, 1974).

Conventional profiling using consensus vocabulary can yield very good results if done properly. However, the downside of such consensus vocabulary methods is the time needed to generate the attributes as a group and the training phase wherein assessors become familiar with attribute definitions. Depending on the nature of the product being evaluated, sensory evaluation is sometimes complicated by order effects and fatigue. This is often the case with products with strong tastes or numbing effects such as beer or wine, where it is simply impossible to evaluate many samples in the same session due to diminished taste sensitivity.

2.2.2 Free Choice Profiling

Free Choice Profiling is a sensory profiling method that steps neatly around the time consuming and difficult task of consensus vocabulary generation and allows naïve assessors to develop descriptors using their own terminology. (Narain et al, 2003) When using free choice profiling, instead of lengthy and complicated group training sessions to define attributes, assessors tasked with exploring the sensory space individually and developing their own definitions. This exploration and attribute definition is usually steered and moderated somewhat by a structured testing program or interface, or with the help of a test moderator. Further advantages of this method include the freedom for subjects to use whatever language they are most comfortable with, less demanding training, and the possibility of using naïve assessors rather than experts.

Since assessors are using terms they themselves have created, the attribute training phase is less critical because assessors intrinsically understand the meanings of attributes. Also from an organizational standpoint, this method is much simpler than consensus vocabulary methods since there is no need for the sensory panel to meet all at once.

2.2.3 Repertory Grid Method

The repertory grid method is a sensory profiling technique used for eliciting sensory characteristics from panelists. It was developed in the 1960's as a method to allow assessors to use their own vocabulary to describe perceptual aspects (Kelly, 1955). The basis for repertory grid elicitation is that assessors are presented with three stimuli simultaneously and asked to find a sensory characteristic to link two of these stimuli. After selecting two of the three stimuli which he or she perceives to be more similar, the subject is asked to come up with a word or phrase describing the similarity and a word or phrase to describe how the third stimuli in the triad is different from the pair (Berg et al, 2000). These similarity and difference pairs were originally referred to as constructs, but the term used in this study is descriptors. The descriptors can then be analyzed for correlations and interrelations using one of several available methods, e.g. cluster analysis. However, repertory grid method was used in this study as a supplement to a quick individual vocabulary profiling method to to help naïve assessors generate attributes. Repertory grid method was using in the first two rounds of this sensory evaluation.

2.2.4 Flash profile method

The use of flash profile builds on free choice profiling, and further condenses it to allow rapid, flexible, and accurate sensory profiling to be carried out. The primary difference between the flash profile method and free choice profiling is that the flash profile method allows assessors simultaneous access to all stimuli in the stimuli set. It works by combining familiarization with the stimuli, attribute generation, and attribute use into one single action. In other words, during their first exposures to the stimuli, the assessors develop terms to differentiate between the stimuli and rate the stimuli according to those terms. It is thought that by allowing assessors simultaneous access to all the stimuli in the set, they are forced to consider perceptual differences between stimuli to examine and differentiate between them. Also, because the flash profile method is a condensed version of free choice profiling, several iterations of the flash profile method are typically used. Again, this allows checks for repeatability and attribute consistency (Delarue et al, 2004). As the assessors proceed through the various trials of flash profile method, they improve and focus their attributes, dropping those attributes that do not fit and adding attributes to fill gaps in the sensory space. A variant of the flash profile method was used throughout the sensory evaluation portion of this thesis, and is referred to as quick individual vocabulary profiling.

2.3 Data Analysis & Interpretation

2.3.1 General Procrustes Analysis

One of the main advantages that consensus vocabulary methods have over free choice profiling and other sensory profiling methods where assessors are free to develop their own terminology to describe the stimuli is that all subjects use exactly the same terms. Obviously, from an analysis standpoint, if all assessors use exactly the same terminology and agree on all the definitions, it is quite easy to derive some kind of assessor 'group average opinion.' Simply averaging together all assessors' opinions into a group average configuration is a bit overly simplistic, but it does well to illustrate the point. Also, consensus methods shine because common concepts can be defined and examples can be given, allowing the exact definitions of attributes to become very precise and their use to be very consistent.

Analyzing results from free choice profiling is much more complicated than consensus vocabulary methods for a variety of reasons. First off, all assessors are free to generate their own individual sensory attributes. Also, in some cases, different assessors have even used different languages within the same study. Another problem is variation in assessor's definitions of attributes. While one assessor might use dryness in wine to mean one thing, another assessor might use exactly the same term to mean something slightly, but still appreciably different. Finally, one assessor might have three sensory attributes to describe a certain stimuli set whereas another assessor uses five. Given all of these complications, how are we ever to arrive at a method for analyzing the results in an objective and repeatable manner? One solution to this problem is principal component analysis, or PCA.

2.3.1.1 Generalized Procrustes Analysis Background

First it is necessary to deal with variations in the number of attributes used by different assessors and their variations of scale. The solution comes in the form of general procrustes analysis, or GPA as it will be referred to henceforth. Originally, procrustes analysis was intended as a method for matching two configurations, for example a matrix of *n* samples by *m* attributes for two different assessors. The origin of Procrustes is taken from Greek mythology, and refers to a bandit who forced his victims to lie on an iron bed. Those who were too tall were cut to length; those who were too short were stretched to fit (Wikipedia, 2006). Procrustes analysis, as such, was developed as a means for matching solutions of two Factor Analyses (Catell et al, 1962), but nowadays it is widely used in sensory profiling and food science.

2.3.1.2 GPA Process

GPA is usually performed on column centered data, which means that it has been normalized to account for variation in levels of scales. This is accomplished by subtracting the column average (attribute average) from each of the entries in that column to account for this variation in scale. The result is a column with an average of zero. (Kunert et al, 1999)

GPA can be divided into two principle stages:

- Determining isotropic scaling factors to allow for differences in range scores
- Finding optimal rotations in the n dimensional space to minimize procrustean distance

The result is a group average configuration which can be used in the ensuing PCA analyses. (Kunert et al, 1999)

GPA works by simplifying different assessor's impressions of products into an Attributes [A] by Products [P] matrix. As mentioned previously, since different assessors have different numbers of Attributes [A], [A] is taken as the largest number of attributes by any single assessor – in our case 12. For assessors with less than 12 [A] attributes, the empty slots in their 12 x 8 ([A]x[P]) matrices are filled in with zeros, which is referred to as padding. Then the GPA algorithm begins by moving the centroid of each assessor's [A] by [P] matrix to the origin of this eight dimensional space. Then GPA scales the assessors attributes ratings isotropicly to account for the fact that some assessors naturally tend to use more or less of the scale. This isotropic scaling corrects for that variation in scale usage. Finally, the [A] by [P] matrix is rotated and scaled in the eight dimensional space with the goal of minimizing the procrustes distance. The isotropic scaling and rotation/reflection are applied iteratively until the procrustean distance meets some preset convergence criteria. (Arnold et al, 1986)

2.3.2 Principal Component Analysis

While GPA deals with alignment of attributes, normalizing of attribute scales, and differences in numbers of attributes between assessors – we still need some method for analying this 12 dimensional data. The solution to this analysis comes in the form of Principal Component Analysis, or PCA is it will be referred to henceforth. PCA is a multivariate analysis technique that is intended to reduce the dimensionality of data and give a smaller set of uncorrelated variables. In other words, PCA is a graphical means of showing relationships in multidimensional data. PCA is very useful for analyzing data in an [M] observations by [N] variables layout. PCA derives its basic functionality by manipulating matrices to create a covariance matrix [S], and eigenvalues [L] using the following formula:

$$U'SU = L \qquad (1)$$

Where U is an orthonormal matrix, and a covariance matrix [S] where:

$$\begin{pmatrix} S_1^{\,2} & S_{12} & \dots & S_{1P} \\ S_{12} & S_{22} & \dots & S_{2P} \\ \dots & \dots & \dots & \dots \end{pmatrix} \qquad (2)$$

$$S_{1P} \; S_{2P} \; S_{3P} \; S_{4P}$$

The strength of a relationship can be given by the following formula:

$$r_{ij} = s_{ij}/(s_i\, s_j) \qquad (3)$$

Finally, the principal component transform transforms P correlated variables in P new uncorrelated variables. The relation is as follows:

$$z = U' \,[\, x - \overline{x} \,] \qquad (4)$$

We can then plot the eigenvectors of the dominant eigenvalues, which gives an excellent representation of the interrelation of the different factors (Jackson, 1991).

An example might help to illustrate more clearly. Let's consider the following data from one individual's sensory evaluation, shown in Table 1.

Phones	MechLooseInSlide	MechStrengthInSlide	EaseOfSlide	EquilAtBiStablePoint	SymmetryInSlideSlope	StrengthOfClicking	SymmetryOfClicking	SlideFeeling
A	3	2,7	-3,9	3,2	-1,8	-0,8	-3,4	0,6
B	3,7	1	1,3	-2,1	1	-1,8	3,8	2,8
C	0,4	0,1	-2,4	-0,6	-0,1	1,3	2,9	-3,4
D	4,3	4	-1,4	1,6	-1,6	2,6	-5	1
E	-2,6	-2,8	3,8	-5	0,5	-0,8	1,5	-0,2
F	5	-0,5	1	-3,6	5	4,4	4,6	4,3
G	-3,7	-1,3	2,6	-1	2,1	-0,2	1,6	-1
H	3,4	2,5	-0,2	-2	1	3,2	2	-1,8

Table 1. Sensory attribute scores for each of the 8 mobile phones, marked with letters A through H

The letters in the leftmost column represent different sensory stimuli – in this case mobile phones with bistables slides. The attributes in the topmost row are attributes created by this assessor in his own words. The assessor has rated each phone according to his specified attribute, with the scale from -5 to 5. Using XLSTAT (Addinsoft, 2006) to perform the PCA, we arrive at correlation matrix shown in table 2.

Variables	MechLooseIn Slide	Mech StrengthIn Slide	EaseOf Slide	EquilAtBi StablePoint	SymmetryIn Slide Slope	StrengthOfClicking	SymmetryOfClicking	SlideFeeling
MechLooseIn Slide	**1,0**	**0,7**	-0,5	0,3	0,0	0,5	-0,1	0,5
Mech StrengthIn Slide	**0,7**	**1,0**	**-0,7**	**0,8**	-0,5	0,3	-0,6	0,0
EaseOf Slide	-0,5	**-0,7**	**1,0**	**-0,8**	0,6	-0,2	0,5	0,2
EquilAtBi StablePoint	0,3	**0,8**	**-0,8**	**1,0**	-0,7	-0,1	**-0,8**	-0,2
SymmetryIn Slide Slope	0,0	-0,5	0,6	-0,7	**1,0**	0,4	**0,8**	0,4
StrengthOfClicking	0,5	0,3	-0,2	-0,1	0,4	**1,0**	0,1	0,1
SymmetryOfClicking	-0,1	-0,6	0,5	**-0,8**	**0,8**	0,1	**1,0**	0,1
SlideFeeling	0,5	0,0	0,2	-0,2	0,4	0,1	0,1	**1,0**

Values in bold are significantly different from 0 with a significance level alpha=0,05

Table 2. Correlation matrix of attributes. Scale from -1,0 to 1,0. Note that the matrix is symmetric.

Figure 1 shows the eigenvalues and their loadings.

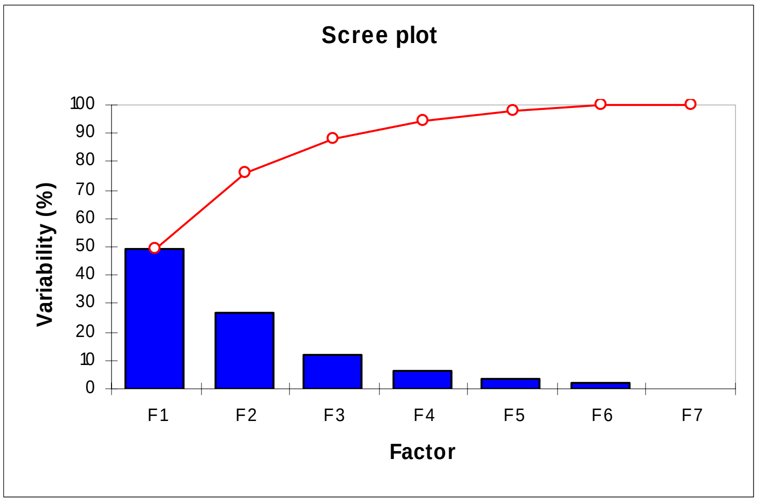


Figure 1. Scree Plot showing attribute loadings for the first seven principle components

Given this information, we see that by plotting the attributes and phones over the F1 and F2 eigenvalues, or the first two principal components, we would account for the most variance (49% and 27% respectively). The results shown in Figure 2.

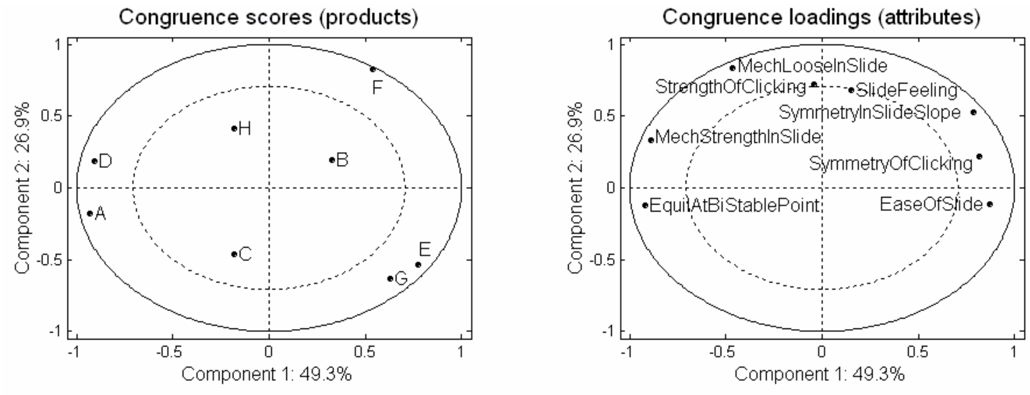


Figure 2. Congruence scores and loadings over the first two principal components for one assessor. Letters A through H represent products. The inner and outer circles indicate 50% and 100% explained variance. Note! These are only preliminary results.

The congruence scores of products and congruence loading of attributes are plotted over the same scale. The idea behind this plotting is that the two graphs: congruence scores (products) and congruence loadings (attributes) are presented over the same principal component space (Lorho, 2005). As such, we can make comparisons between products in the leftmost plot, and their loadings with respect to attributes. The two ovals represent 50% significance and 100% explained variance respectively. The further out radially attributes and products are, the stronger the correlation being represented is. In the component space given, attributes at opposite ends of the circle represent negatively correlated attributes, and those that are spacially closer are more highly correlated. For example, from the leftmost plot, we can say that there is an inverse correlation between the assessor's attributes 'SymmetryOfClicking' and 'EquilAtBiStablePoin't (Equilibrium at Bistable Point). Additionally, since the congruence scores and congruence loadings are plotted with the same principal components over the same space, we can make inferences between product locations and attribute locations. For example, sample D seems to correlate well with 'MechStrengthInSlide.' It should also be noted that the data presented here was taken from a session early in the training, and is not representative of the true final results of this study.

2.4 Internal and External Preference Mapping

As individuals, we all have unique preferences when it comes to products. As such, it would be a grave oversight to simply lump together a broad group of consumers and average their opinions. Given that fact, we need some means of classifying variations in preference from individual to individual and how this varies relative to a given set of products. Additionally, there is a need for some metric to track product improvements and their acceptance in different market segments. The answer to this problem is internal and external preference mapping. Preference analyses, simply put, are a means of visualing relationships between products, attributes, and preferences. Preference maps plot perceived product characteristics (attributes) and consumer preferences on the same perceptual space. The two basic types of preference mapping are internal and external preference mapping. External preference data is built on perceived characteristics (attributes) and preference data, while internal preference mapping is based solely on preference data. While both methods basically use PCA to carry out the underlying statistical work, internal preference mapping uses a variant called singular value decomposition, or SVD, and external preference mapping uses full PCA. (Kleef et al, 2005)

A simple, imaginary example of internal preference mapping is shown in Figure 3. The figure shows the first two principal components, spanning 49.3% and 26.9% of the variation respectively. In addition to that, the key to preference mapping is the vectors eminating from the origin of the plot. Each of these vectors represents one consumer’s preference, the longer vectors representing a higher correlation. For this example, we can see a small cluster clearly preferring products A and D, with some other smaller preferences heading in other directions. Excellent real world examples of internal and external preference mapping include a cigarette lighter sounds study (Lageat 2003) and several studies involving consumers beer preferences (Guinard et al, 2001), which are discussed in the section 2.5.1.

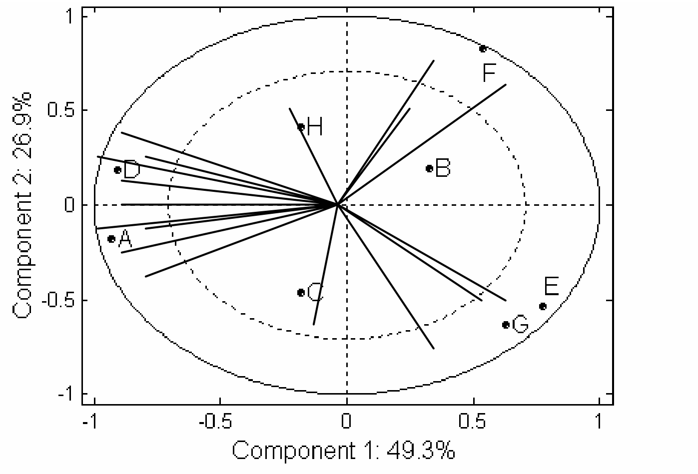


Figure 3. Example of an internal preference plot over the first two principal components. The inner and outer circles show 50% and 100% explained variance.

2.5 State of the Art

2.5.1 Research using Consensus Vocabulary

As mentioned previously, methods using consensus vocabulary have been the norm in food science and sensory profiling for decades. It is helpful to consider an example, and Guinard, Yip, Cubero, and Mazzuchelli's *Quality Ratings by Experts* (Guinard et al, 1999) paper is a good candidate. This study covers 71 commercially available beers, and a panel of 17 people who were selected for their backgrounds in beer and brewing industries. QDA was used to structure the descriptive analysis of the beer. Training took place over a span of about one month (~25 hours) and was done in groups. Testing took place almost every day for 4 months, taking a long time because of the huge number of stimuli and the repetition of each beer four times. Additionally, assessor fatigue was minimized by rating slowly with small amounts, and dealing with only ten beers per session. Interpretation of results was done using PCA. Additionally, one interesting means of intererpreting the results was the slope analysis where preference data was plotted against quality ratings, giving some mean idea of how each sensory attribute affects average preference. (Guinard et al, 1999) This type of result is especially useful when seeking to optimize sensory characteristics. Another valuable method for interpreting results is internal or external preference mapping, where consumer preference vectors are mapped onto the same sensory space as the products, and needs and wants of a market can be readily identified. (Guinard et al, 2001)

2.5.2 Research using Individual Vocabulary

Flash profile methods have been widely used as a means of individual vocabulary profiling. Its advantages come from its quickness to implement and the lack of time-consuming attribute training sessions. Flash profile method was used to explore fruit yoghurt, and was performed in parallel with conventional (consensus) profiling. This does well to illustrate the differences in time required for results. The conventional profiling panel composed of 10 judges were previously trained over a period of 6 months to a year and half, and received an additional 25 to 80 hours of training for evaluating fruit flavor. The 10 flash profile method judges were all experienced sensory profilers and were not specifically trained in evaluating fruit aromas. The flash profile panel used only four sessions (between 30 and 75 minutes); the first session was used for attribute generation, the second session was used for refinement of attributes, and third and fourth sessions for evaluation. This structure is quite similar to the testing schedule for this thesis. The results achieved by flash profile method were nearly identical to those from conventional profiling, but were achieved in a fraction of the time. (Delarue et al, 2004)

Flash profile method has also been applied to sound research, with examples such as Lorho's *Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction* (Lorho, 2005). This study is similar to Delarue's flash profiling process, but is quite rigorous in its selection of assessors. Assessors are selected using discrimination tests and personal interviews where they were asked to describe perceptual differences in a pair of stimuli. Well performing assessors from the discrimination test and those who were desciptively skilled were selected. Attribute generation, refinement, and testing was done over five sessions of approximately one hour each. Results were interpreted using GPA and PCA, and also a dendogram produced by agglomerative cluster analysis. (Lorho, 2005)

2.5.3 Sensory Profiling Applied to Mechanics

While sensory profiling has been widely used in food science for decades, it has been slow to spread beyond that field. There are some notable cases of its adoption in other fields, including Lorho's aforementioned stereo headphone reproduction study. There are some additional examples of mechanical applications using sensory profiling. Among the most compelling is Lageat, Czellar, and Laurent's study of cigarette lighter opening sound (Lageat et al, 2003). They sought to study the sound produced by cigarette lighters and how this affected perceptions of luxury. For reasons related to repeatability and to ease test administration, recordings of lighter sounds were made for each lighter in the study. Twelve judges were used; all judges had no specific training or experience in sensory evaluation. Judges created descriptors for each sound and ranked the sound according to the selected attribute. Then in group sessions, attributes were eliminated because they failed to describe the sensory variations. Following this attribute elimination, low and high anchors were created for the seven selected sound descriptors using Cool Edit Pro. The study also gathered untrained consumers ratings for determining the luxury of each sound. Using PCA, correlation between attributes, products, and consumers perceptions of luxuries could easily be examined (Lageat et al, 2003). It should be noted that this combination of assessor evaluations and hedonic data from untrained consumers is one the keystones of successful sensory profiling. The elegance of this method comes from the fact that experts provide the characterization, and the consumers provide the preference data. The assessor's evaluations are not clouded preference considerations, and the consumers are not asked to describe the stimuli at all.

Perhaps the research that is most similar to this thesis is a study of brake feel in automobiles by Renault Research Division and ENSIA (Dairou et al, 2003). The goal of this study was to improve braking comfort and safety by exploring its contributing factors. This was achieved using a braking by wire system and an active pedal feel emulator. Vocabulary for describing brake feel was arrived at using flash profile methods and a five person panel. The sensory profiling itself used an eight person panel and QDA to explore 12 different braking laws. ANOVA and PCA were used to interpret the results. (Dairou et al, 2003) In terms of area of modality, brake feel is highly analogous to the slide
phone mechanics. Braking feel is perceived primarily as tactile and acceleration related sensations, and slide phone mechanics are primarily tactile with some complimentary auditory sensations.

3. EXPERIMENTAL PROCEDURE

3.1 Sensory Panel Composition and Selection

The sensory panel for this study was comprised of 13 assessors – 8 men and 5 women. The majority of the assessors were naïve assessors, with one selected assessor with previous sensory profiling experience relating to sound. Of the 12 assessors, six are employed in the mobile phone industry in some capacity pertaining to mechanics or mechanical quality and two of the assessors have little or no previous work experience relating to mobile phones. The selected assessor was intentionally sought out so as to allow some rough comparison between naïve and selected assessors' quality of attributes. Differences were found to be negligible.

3.2 Experiment Environment

All sensory evaluation was administered in an auditorily isolated listening chamber (ambient noise level ~25 dBA) for convenience and to minimize assessor distractions. The testing chamber is well lit, and assessors sit at a table and are presented with 8 mobile phones with bistable slide mechanics. Also in the room is a flat 19 inch LCD monitor, wireless mouse and keyboard, and paper and pencil to allow the assessor to take notes. Phones are presented in a power off state - with no alterations made to obscure logos, brands, or model numbers. Phones were referred to by letters A through H, and rest on two A4 sheets with markings to indicate the phone's proper position and associated letter. Assessors were free to rearrange the physical layout of the testing setup to their taste.

Assessor side testing computer is a PC running Windows XP and using Exceed PC X-server software to display the testing interface. Backend testing PC is a computer with a Linux variant operating system and Guinea Pig 3 testing software. Guinea Pig 3 (Hynninen et al, 1999) was selected as test UI because of previous experiences, ease of modification, and availability in the testing lab. GP3 was originally designed as subjective audio testing software, but extends well to meet the needs of broader fields of user testing as well. One of the few downsides of using GP3 for physical product sensory evaluation is the requirement that the test subjects use the keyboard and mouse while simultaneously manipulating samples with their hands. This requires the subjects to take their hands away from the keyboard and mouse, interrupting work flow and increasing cognitive load.

3.3 Phones in Study

Eight bistable slide phones were used in the study. A slide phone is a mobile phone with two halves which slide relative to each other. This movement is spring assisted, meaning that the mechanism initially resists movement and then snaps into place once a certain point has been passed, hence the name bistable. The interplay between sliding friction, spring force, spring bias, and sliding speed are factors which have significant impacts on perceived quality and pleasantness. The phones used in this study were selected to represent typical variation in these characteristics. The phones in the study are shown in Figure 4.

Figure 4. Stimuli used in sensory profiling trial

3.4 Testing Process

The testing took place over four sessions of approximately one hour and fifteen minutes each. Assessors performed a given set of tasks per session, so session times vary significantly according to the speed of the assessor.

3.4.1 Sessions One and Two

In the first session, assessors began by performing a simple preference test where they were asked to rate the mechanical feel and sound of the eight phones used in the study. Assessors were specifically instructed only to consider mechanical attributes relating to the slide, not to non slide related ergonomics, styling, nor any other factors. The preference test was intended both to give data for internal and external preference mapping and also to familiarize the assessors with all the phones used in the study. The user interface for the preference test is shown in Figure 5.

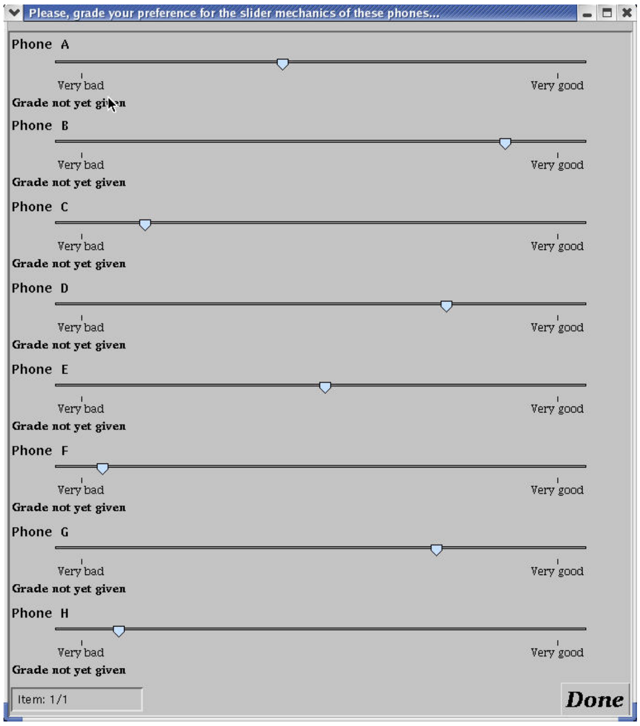


Figure 5. GuineaPig 3 user interface for preference tests.

In addition to gathering preference test data from each of the 13 assessors, 28 other individuals took the preference test to increase the size of the dataset.

Following the preference test, assessors were instructucted to create opposite descriptive terms to describe differences in feel for a randomly selected pair of mobile phones. The UI for this task is shown in Figure 6.

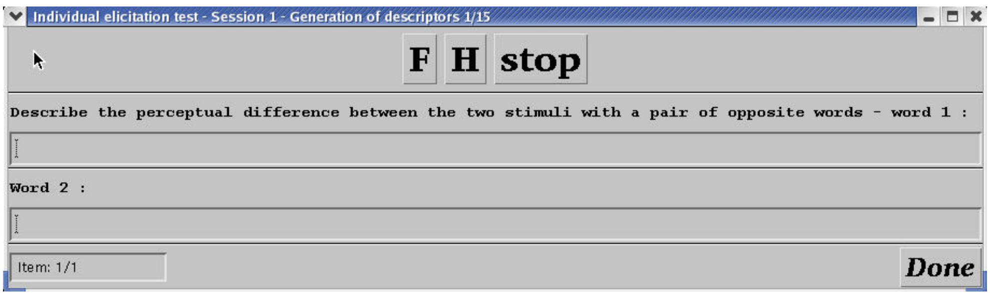


Figure 6. GuineaPig 3 user interface used for dyad stimuli comparison

Assessors were exposed to fifteen such randomly selected pairs. In the first round, two phones were compared. This two stimuli comparison, dyad comparison, asks assessors to find differences between two stimuli and describe those differences. Some research suggests that dyad comparison may give more clear opposite pairs than other methods (Epting et al, 1971). Another method, triad comparison, poses the question "In what way are two of these stimuli similar, and how is the third one different?" Eight such triad phone comparisons were made at the start of the 2nd session. Assessor comments showed no clear difference in preference between dyad or triad repertory grid techniques. The triad repertory grid UI is shown in Figure 7.

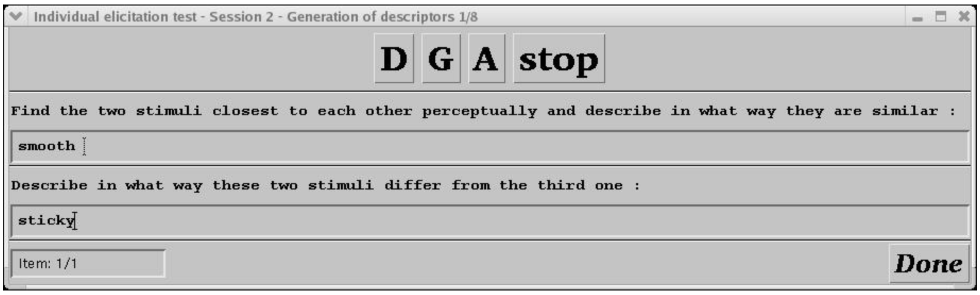


Figure 7. GuineaPig 3 user interface used for triad stimuli comparison

Then, after completing the dyad or triad comparisons, assessors were instructed to generate attributes using their list of opposite terms and find a low and high anchor for each. For example, an opposite set like 'sticky' and 'smooth' might yield an attribute like 'smoothness', with low and high anchors of 'not very smooth' and 'very smooth' respectively. The UI for this task is shown in Figure 8.

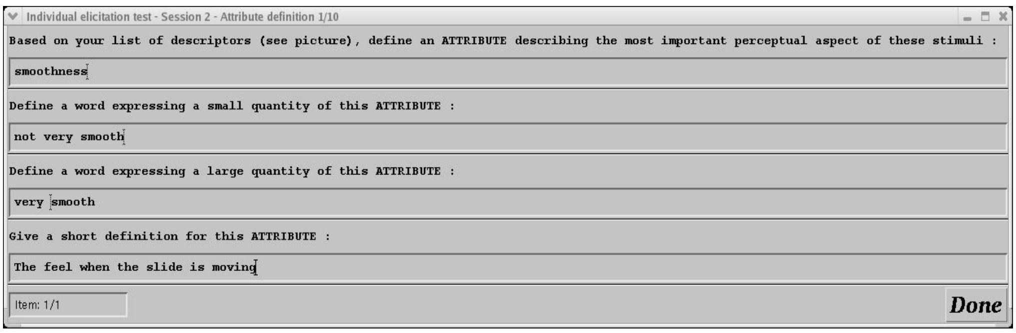


Figure 8. GuineaPig 3 user interface used attribute definition

For each attribute, assessors were instructed to evaluate four of the phones according to that attribute. Since the first session was a training session, only four of the eight phones were evaluated to allow for quicker results. In all subsequent sessions, all eight phones were used in the testing process. The GP3 window for trials with all eight stimuli is shown in Figure 9.

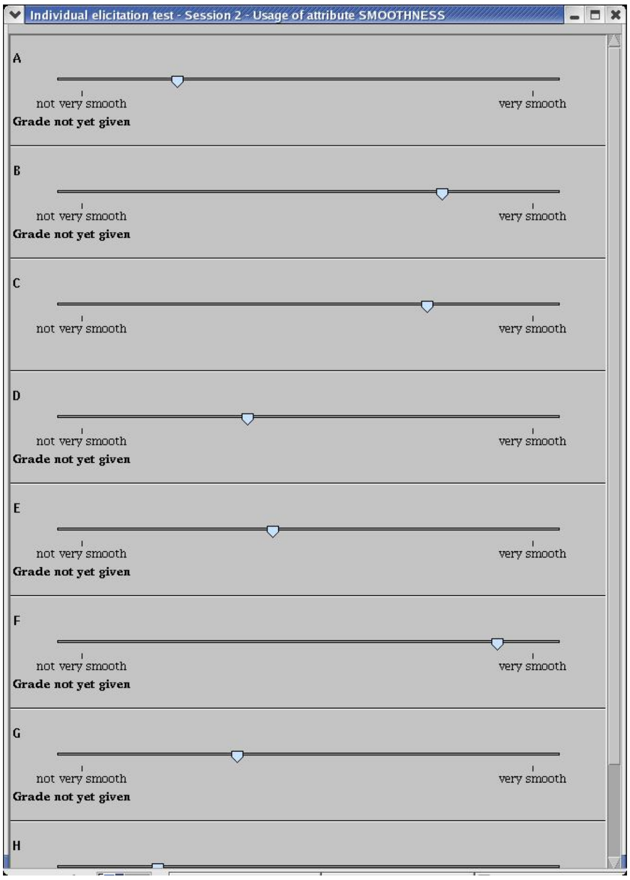


Figure 9. GuineaPig 3 user interface used for dyad stimuli comparison

3.4.2 Sessions three and four

The third and fourth sessions contained no elicitation, or generation of opposite words based on comparisons, but rather assessors were asked to consider the previously generated attributes and their applicability to the set of phones. Assessors were instructed to be as accurate as possible during third and fourth rounds since these results would be used as a basis for final results.

3.4.3 Overall Process Diagram

The entire process diagram is shown in Figure 10. It should be noted that while up to 15 windows for attribute definition and use were included in the testing process, assessors were instructed that fifteen was a maximum and smaller numbers of attributes were acceptable as well. In other words, attribute quality rather than quantity was emphasized as being important.

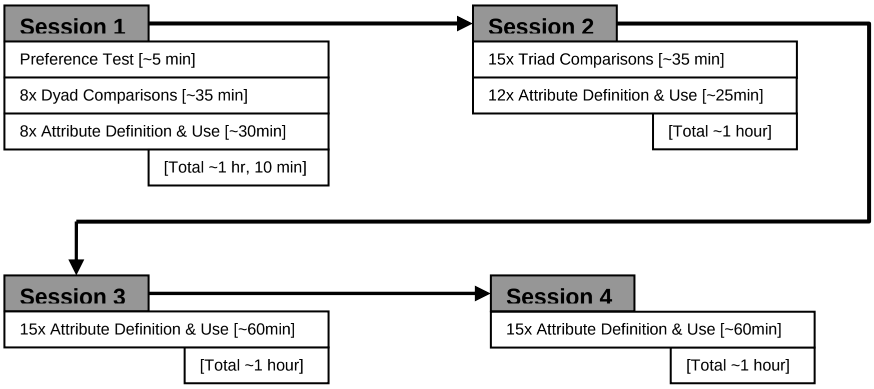


Figure 10. Process diagram detailing sessions one through four and the tasks performed within each session.

4. RESULTS

4.1 Agglomerative Hierarchical Clustering

Early in the analysis of this dataset, there was a need for a means to look at the individual attributes, or the unique words being used by each assessor to describe the stimuli. It is convenient to have some simple and graphical way of studying at these attributes. To facilitate this, agglomerative hierarchical clustering was applied to this dataset to allow some visualization of similarity between attributes. AHC is commonly used in conjunction with the individual vocabulary methods to allow some grouping of the different attributes produced by the assessors. When clustering attributes together which have more similarity, those attributes with low dissimilarity will theoretically be closer and might presumably have similar meanings. In this instance, AHC was implemented within XLSTAT. The resulting plot is shown in Figure 11.

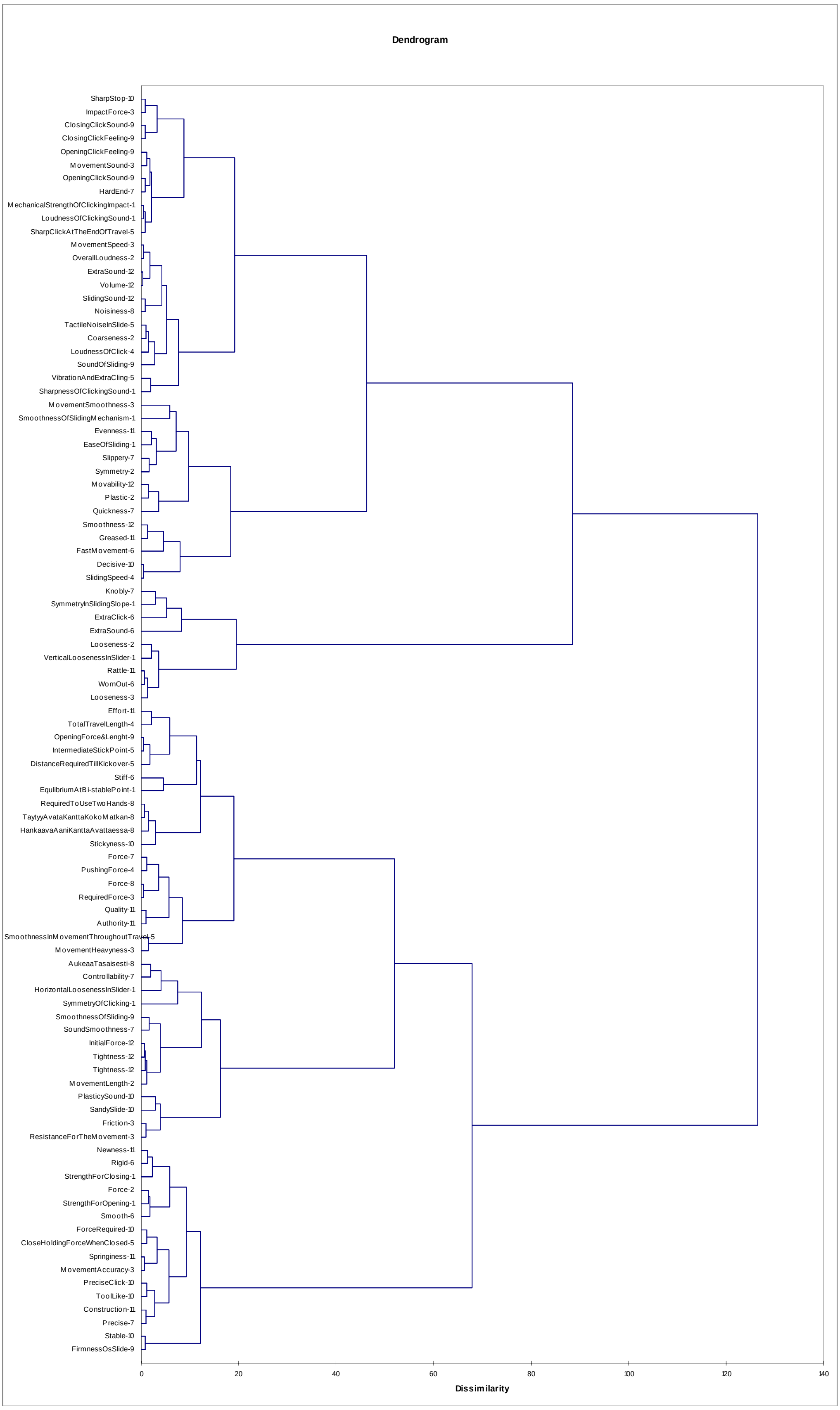


Figure 11. Agglomerative Hierarchical Clustering plot showing attributes and their dissimilarity. More similar attributes have links which have lower dissimilarity scores.

Clearly, the plot from the AHC shows some very interesting groupings. For example, four different assessors have selected force as an attribute and are shown with very tight clustering. This tight clustering indicates that there are distinct similarities in the way that those four separate assessors apply their individual ‘force’ attributes. Also tightly clustered in this group are ‘authority’, ‘movement heaviness’, and ‘smoothness throughout movement’. The attribute ‘quality’ is also tightly clustered therein, but since this may be a hedonic attribute, meaning that it deals with preference and does not necessarily describe or characterize it, should be treated with caution. Several other similarly convincing clusters can be found, and in general we can preliminarily conclude that the assessors have focused on several main perceptual directions and show some degree of consensus.

4.2 Principal Component Analysis correlation Loadings

After applying GPA to attain a group average configuration, PCA allows us to reduce the dimensionality of the data and to visualize correlations. The model explained 82,4% of the variation in the first three principal components (PC1: 35,8%, PC2: 28,8%, PC3: 17,8%). The plots in Figure 12 show the correlation loading plot and PCA scores of products with 95% confidence intervals over PC1 and PC2. This figure shows some solid groupings of attributes similar to those seen the agglomerative hierarchical clustering plot. The four perceptual aspects highlighted in the graph are drawn in by hand during interpretation (force, rubbing, stickiness, click, and looseness), and show that there is an inverse correlation between force and associations of looseness, since the vectors are roughly opposite. Also, the force and stickiness aspects are orthogonal, implying that there is no correlation between the stickiness of movement and the force of movement. Finally, since stickiness and strong click aspects are in nearly opposite directions, this implies a negative correlation – meaning that generally those phones with strong click were perceived as lacking in stickiness of the slide. Whether this is because stickiness inhibits strong click by limiting speed is a question which has yet to be answered.

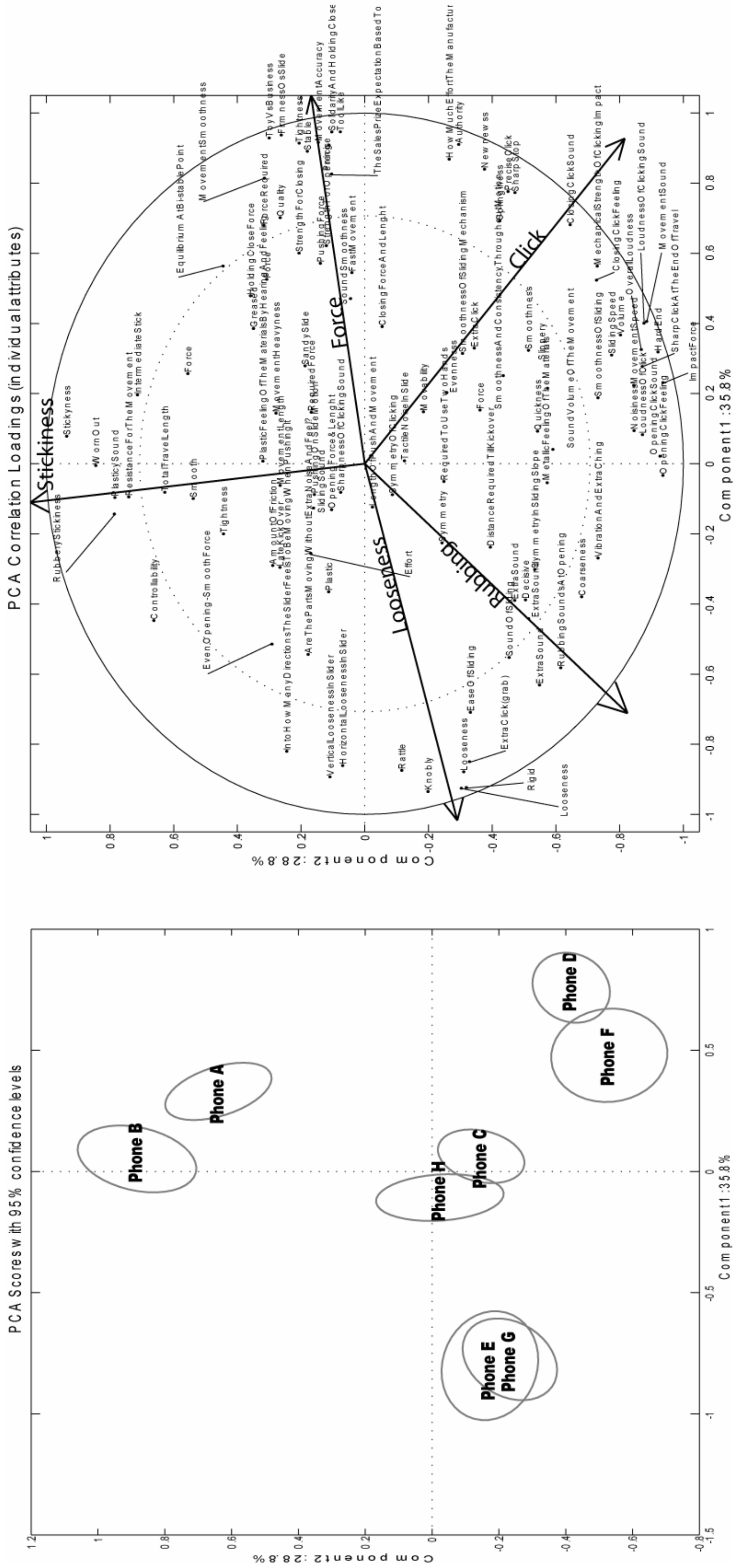


Figure 12. Plots showing PCA scores of products with 95% confidence elipses, and PCA correlation loadings with attributes mapped over the same principal component space for PC1 and PC2.

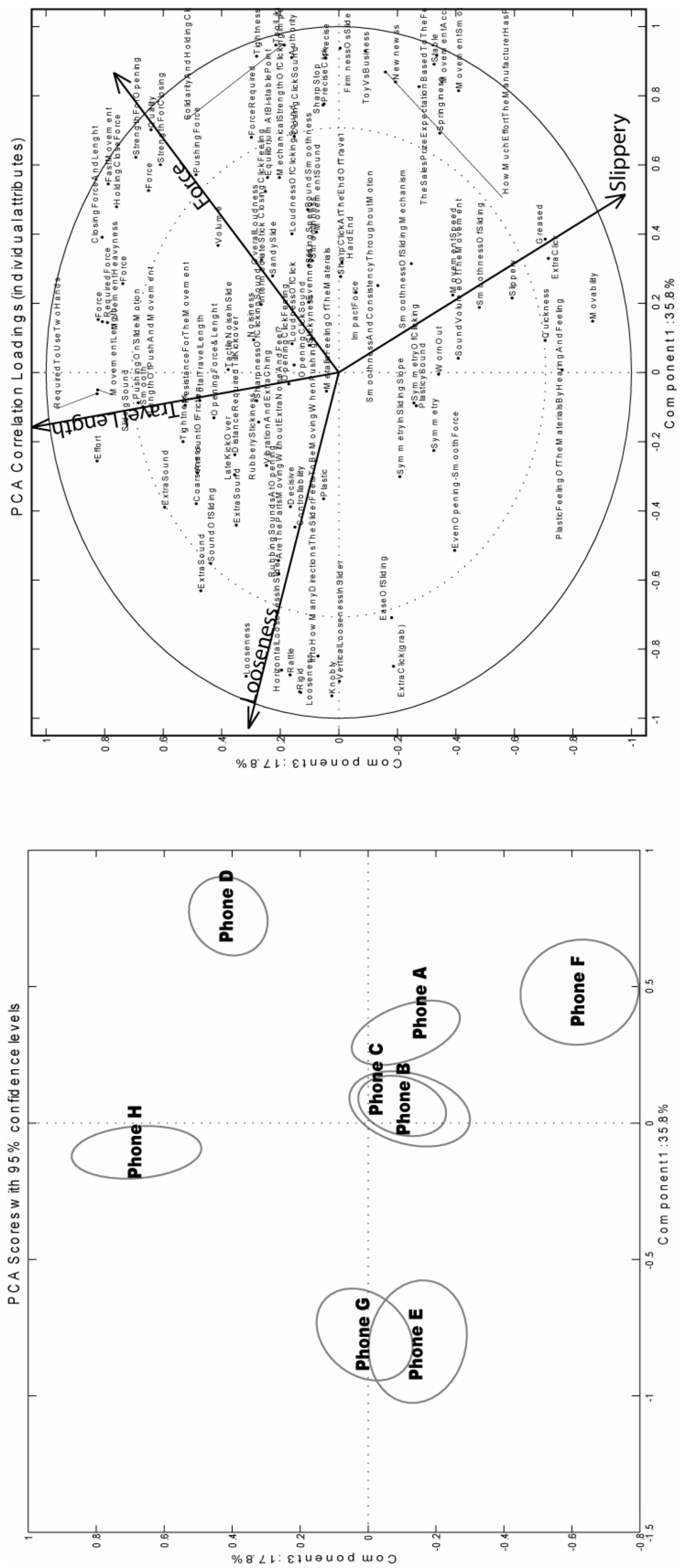


Figure 13. Biplot showing PCA scores of products with 95% confidence elipses, and PCA correlation loadings with attributes mapped over the same principal component space for PC1 and PC3.

Shown in Figure 13 is the plots over PC1 and PC3. The perceptual aspects highlighted in Figure 13 include ‘looseness’, ‘length of travel’, ‘force, and ‘slippery’. Also included in Figure 13 are some excellent correlations between the highlighted perceptual aspects and several mobile phones. For example, the Phone D seems be loaded heavily in the direction of the ‘force’ perceptual direction. Included in this perceptual direction are attributes like ‘quality’, ‘strength for closing’, ‘force’, ‘strength for opening’, ‘pushing force’, ‘tightness’, ‘tool like’, ‘force required’, etc… This implies that given the choice, the assessors who identified those attributes would readily associate them with the Phone D if given the opportunity. Other clear correlations between products and perceptual directions include the heavy loading of the Phone F on the ‘slippery’ perceptual aspect. Some of the attributes which contributed to the placement of slippery in that direction include ‘movability’, ‘greased’, ‘quickness’, ‘plastic feeling’, ‘slippery’, ‘smoothness of sliding’, ‘movement smoothness’, and ‘movement speed’. Also in that rough cluster is extra click, which clearly is referring to something else, and may have come from human error or some other related factors.

While the PC1/PC3 and PC1/PC2 plots shown in Figures 12 and 13 represent the majority of perceptual aspects, there are still some perceptual aspects which are not well represented on this combination of principal components. Some other information concerning the interrelation between ‘plastic feel’, ‘smoothness’, and ‘stickiness’ is presented on the PC2/PC3 plots, and is shown in Appendix B.

4.3 Internal Preference Mapping

Finally, looking at assessors opinions of the phones alongside the preference test data, we can relatively easily visualize information on which phones were preferred. The internal preference map data is presented in figure 14. In this case, PCA was used to visualize the results – so the graphs are interpreted in a similar manner. Again, the numbered vectors indicate one individual's preference. The product names are shown as named points. Vectors and product that are further out from the center are more strongly correlated. Proximity in terms of rotation about the circumference indicates similarity, points or vectors that are orthogonal have no correlation and those that are in opposite directions have a negative correlation, or strong dissimilarity. Clearly, in this chart we see the largest cluster of preferences seem to to generally prefer the mechanical feel of the Phone D, Phone F, and Phone C. A rather indistinct and small cluster of consumers seem to prefer the feel of the Phone B and Phone A. Other, non-clustered preference vectors are going off in other directions. The main import of this chart is that the largest cluster of individuals prefer the feel of the the Phone D, Phone F, and Phone C and those same individuals strongly dislike the Phone G and Phone E. .

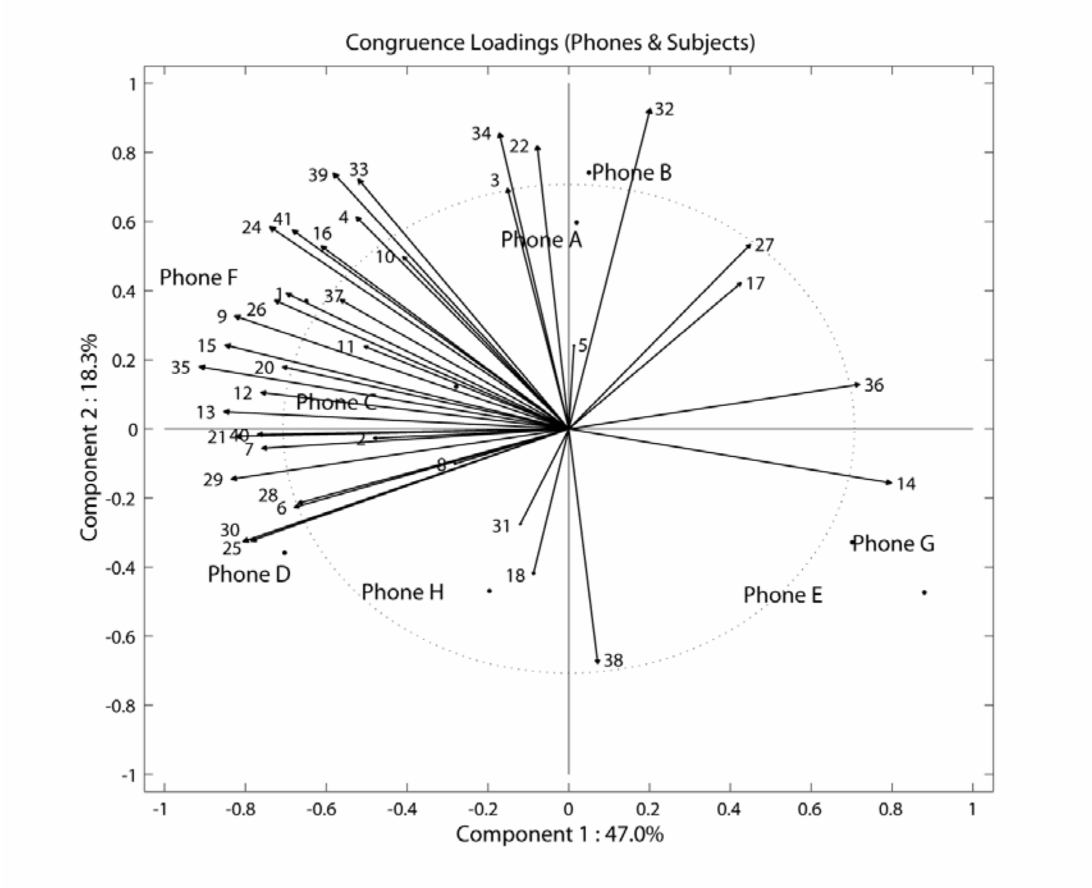


Figure 14 - Internal preference map. Vectors indicate individual preferences, and products are shown as named points. Dashed and solid circles show 50% and 100% explained variance.

4.4 External Preference Mapping

As discussed previously, external prefence is a combination of preference data and perceptual data (individually generated attributes). Now given the two correlated matrices, we need some way to relate them. nPLS, or multilinear partial least square regression, is used to relate the data from the two matrices: the attribute matrix [X] and the preference data matrix [Y]. A short description of nPLS is available from KVL, the Danish Royal Veterinary and Agriculture University (Bro, 2006). In our case, nPLS was computed with the PLS Toolbox (Eigenvector Research, 2006), and three factors were used. The results are as follows:

Percent Variation Captured by N-PLS Model

LV	X-Block	Y-Block
1	30.40	38.44
2	54.86	56.56
3	73.53	67.30

Then, the results are visualized using PCA plots, with the main difference being that the PCA plot shows factors for both the attribute matrix [X] and the preference data matrix [Y] in each principle component. The PCA plot over the first two factors is shown in Figure 15.

Considering Figure 15, generally the same trends that were visible in the PC1/PC2 plot based on attribute ratings alone in Figure 12, although some shifting of products and attributes has occurred as a result of the nPLS regression. For example, in both Figure 12 and Figure 15, the Phone E and Phone G are very tightly grouped. One difference is the increased distance between Phone F and Phone D which is evident in Figure 15. Presumably, this occurs because changes resulting from the nPLS resgression.

When considering the rightmost plot in Figure 15, we see that the largest significant cluser of consumers generally prefer the Phone F and to a lesser extent the Phone D. These two devices are both heavily loaded in the ‘force’ and ‘click’ perceptual directions. The Phone D is highly loaded in force perceptual direction, and is well correlated with attributes like ‘sharp click’, ‘loudness of click’, ‘sliding speed’, ‘impact’, ‘smoothness’, etc… Also, it should be noted that this device is located at the edge of the preference cluster, indicating that this device is somewhat polarizing in its response in the preference study.

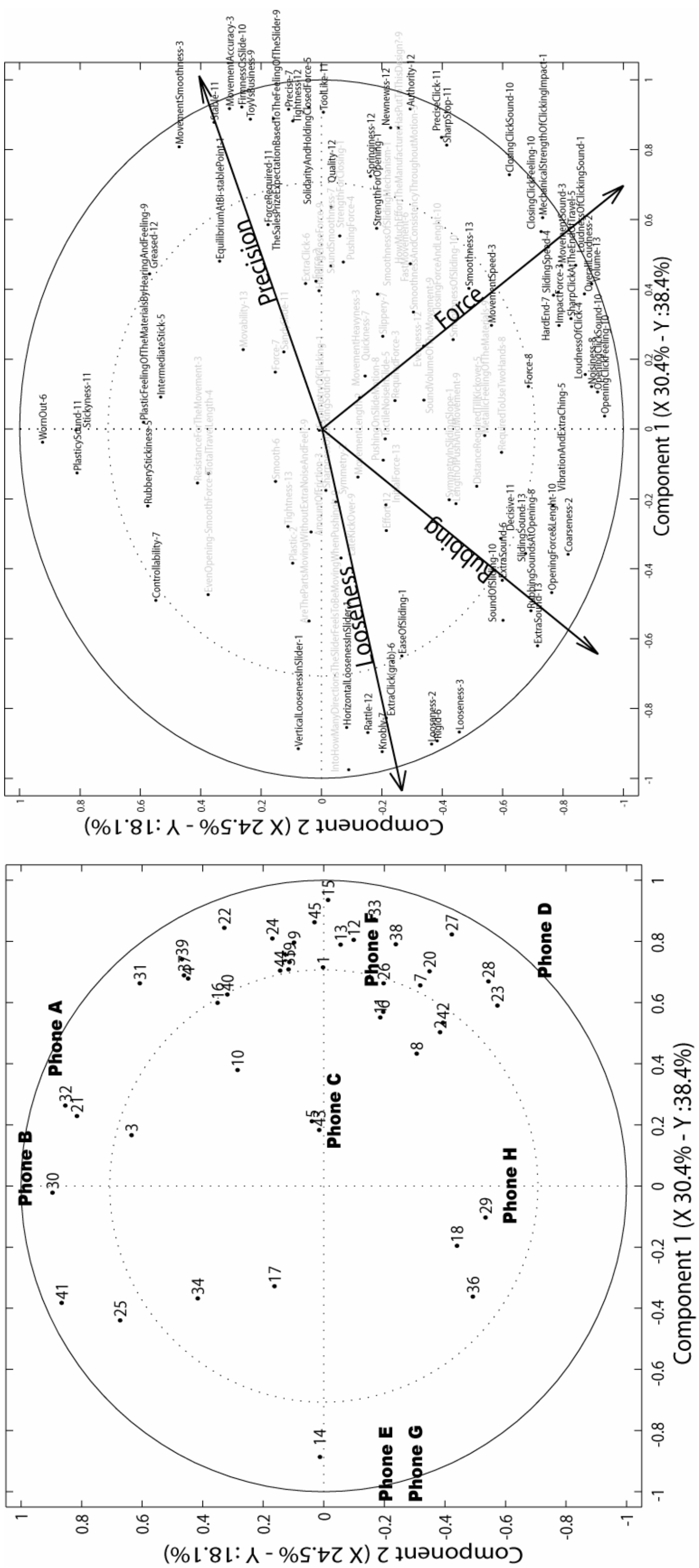


Figure 15. PCA analysis of the first two nPLS factors. Numbers represent consumer preference scores, and the names represent products. Inner and outer circles indicate 50% and 100% explained variance.

4.4.1 Attribute Listings

Based on the information from the PCA Correlation loading plots and the agglomerative hierarchical clustering plots, some key perceptual aspects emerge. The main perceptual aspects are as follows:

- Force
  - Force Required [PC1/PC2] [PC3/PC1]
  - CloseHoldingForce
  - StrengthForClosing
  - StrengthForOpening
  - Authority
  - Pushing Force

- Looseness [PC3/PC1][PC1/PC2]
  - Rattle
  - Looseness
  - Looseness
  - WornOut
  - VerticalLoosenessInSlider

- Smoothness [PC3/PC2]
  - Slippery
  - Quickess
  - MovementSmoothness
  - Greased
  - SmoothnessOfSliding
  - Movability
  - SmoothnessofSlidingMechanism
  - FastMovement

- Total Travel Length [PC3/PC1]
  - Movement Length
  - RequiredToUseTwoHands
  - TotalTravelLength
  - HaveToPushEntireMovement
  - DistanceRequiredForKickover

Click [PC1/PC2]

- OverallLoudness
- ClosingClickSound
- SharpStop
- ClosingClickFeeling
- SharpClickAtEndofTravel
- MechanicalStrengthofClickingImpact

Rubbing [PC1/PC2]

- Rubbing Sound at Opening
- Coarseness
- Vibration and extra noise
- Extra Sound
- Extra Sound (other assessor)
- Metallic feeling of materials

Movement Smoothness [PC2/PC3]

- MovementSound
- Noisiness
- SlidingSound
- Coarseness
- TactileNoiseInSlide
- Volume

This attribute set comes from a group of assessors with largely no previous experience in sensory profiling. As such, the quality and breadth of attributes and their descriptive powers were greater than what was initially expected. Obviously, the existance of some hedonic attributes (attributes relating to preference) is a problem, but on the whole the attribute set is very satisfactory.

4.4.2 Attribute Type Breakdown

What follows is a breakdown of the attributes created by the various assessors. The attributes created during sensory profiling can be divided into three main categories depending on their modalities. Attributes were divided into tactile, auditory, and hedonic modalities. Obviously tactile attributes relate to the force and feelings imparted to the user by the slide mechanism, and auditory attributes relate to the sound created by the slide mechanism. Hedonic attributes, or attributes relating to preference – are not so useful when describing the perceptual sensory characteristics, and should generally be discarded. There is some leeway for overlap between auditory, tactile, and hedonic modalities, since a slide that sounds coarse and sandy might also impart a tactile impression of being coarse and sandy, but for the most part attributes were clearly either auditory or tactile. A complete listing of the percentages of tactile, auditory, and hedonic attributes is presented in Figure 15.

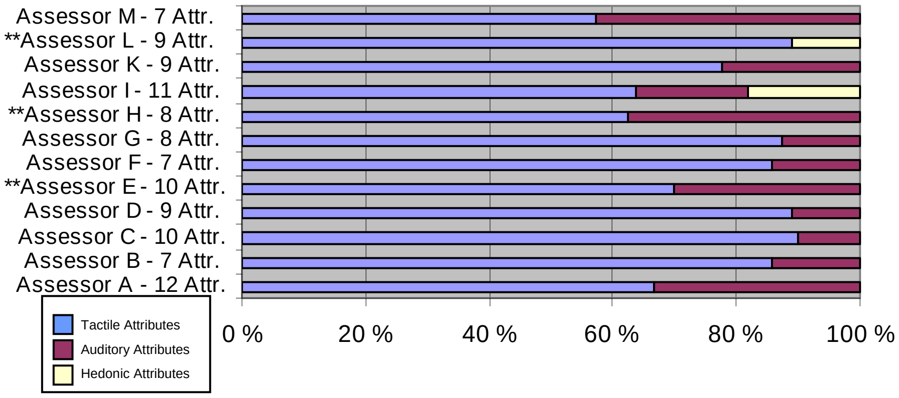


*Figure 15. Chart showing the percentages of attributes falling under tactile, auditory, or hedonic modalities. Assessors creating one or more attributes in his/her native language are indicated with **.*

Note that in Figure 15, assessors creating attributes in his or her own native language are indicated with two asterisks preceding the assessor ID. It was originally hypothesized that assessors creating attributes in a non-native language would experience more difficulty with the task and be more likely to resort to hedonic attributes. This does not seem to be the case, though. Another view of the attributes generated in this study is presented in Appendix A. Therein, the attributes are grouped according to interpreted meaning, and we can relatively easily extract some rough measures of how often certain attributes were selected by different assessors.

5. DISCUSSION

The attribute generation process was quite satisfactory, especially given the limited experience of the panel. Especially notable was the speed with which the predominantly naïve assessors were able to grasp the attribute definition and attribute use tasks and generate attributes. After roughly two and a half hours of product familiarization, dyad and triad repertory grid comparison, and attribute use and generation sessions the assessors had arrived at their final attributes. Assessors were given the opportunity to repeat the repertory grid and attribute definition phases from the second session to allow for the generation of additional attributes, but none accepted. Presumably, this indicates that the assessors were satisfied with their attributes and that they felt that they had exhausted the realms of possibility for new attributes.

One interestesting result was the number of defect attributes developed by assessors. The ‘looseness’ attribute, referring to mechanical looseness and wobble present when the slide was stationary, is an example of one such defect attribute. It was originally hypothesized that defect attributes would be prevalent, and this seems to have been generally correct. Other examples include ‘coarseness’ or ‘rubbing’, ‘stickiness’, and of course ‘plastic feel’. Note that the ‘plastic feel’ mentioned here refers to the feel of movement and clicking rather than any material feel felt on the surface of the phone. Previous research suggests that defect attributes were better predictors of quality than descriptive attributes, and it was expected that the same trend would be evident here (Guinard et al, 1999). Some other attributes, such as ‘stickiness’, ‘sharp click’, and ‘tightness’ may be defect attributes – and may indeed prove to be good predictors of preference and quality, but more study is needed.

One problem that was encountered during the repertory grid, attribute definition, and attribute use phases was large differences in task motivation and engagement between assessors. Perhaps the inclusion of some discrimination task and initial interviews to gauge interest in participation might rectify this situation. In addition, there was little appreciable difference in attribute quality between assessors with a technical background and those with working in unrelated fields.

5.1 Lingual hiccups in the process

One of the clear advantages of free choice profiling and flash profile method is the freedom for the panelists to create attributes without the necessity of communicating those attributes to others. In contrast, during consensus vocabulary generation, if a subject found a relevant attribute – not only would he or she have to precisely define it in a language that all assessors could understand, and in addition he or she should cite examples of low and high anchors for that attribute. This freedom of language with respect to attributes meant that potentially all assessors in this sensory testing test could define attributes in his or her own native language. As indicated in Figure 15, we see that in this sensory profiling test only three of the twelve final assessors created attributes in their own native language. Presumably, this was a result of the debriefings at the end of the each of the four sessions. During these debriefings, assessors were asked to tell a bit about their attributes in English with the test administrator, and this may have discouraged assessors from using their own native languages. One might speculate that if all assessors had chosen to create attributes in their own native language, the descriptiveness and breadth of attributes may have been better. This should be explored further.

5.2 Possible Improvements

First off, it should be noted this sensory evaluation trial employed a panel in which twelve of the thirteen assessors had no previous sensory profiling, and the test administrator had no previous experience moderatating sensory testing. Both of these factors contributed to minor mistakes along the process. Among these mistakes are the inclusion of several hedonic attributes, missing data for one assessor due to problems with testing setup and software, and consumer preference data which ideally would have been taken from a much larger and more randomly selected population. Simply rectifying these glaring mistakes would have greatly improved the quality of the results.

As mentioned previously, some problems arose with assessor motivation and aptitude. One way to deal with this could have been to pre-screen assessors. Ideally, this pre-screening should first involve some type of discrimination task where panel members are evaluated according to some preset criteria. There are four main qualities present in good sensory panels:

- Repeatability – whether or not the same stimuli generates the same level response time and time again.

- Agreement – whether a single assessor gives the same response as the accepted response (can be taken to be a panel mean)

- Discrimination – Whether an assessor or panel can distinguish between several stimuli for a given attribute or attributes

- Multivariate sensory information – whether attributes are redundant, ie redundantly dealing with the same perceptual aspects.

If these criteria are evaluated prior to the main sensory profiling task, assessors which perform poorly might be eliminated, giving a better final result (Zacharov et al, 2006) Additionally, an interview process where potential assessors are asked to describe stimuli might be one possible method for gauging descriptive abilities and task engagement.

Finally, by organizing group discussions after the completion of the four rounds of individual profiling, the assessors might discuss their individual attributes and gain a deeper understanding of the perceptual aspects involved. This type of moderated discussion is very similar to what occurs in the early stages of consensus methods like QDA, but would presumably be less time consuming and less demanding of the assessors. Obviously, following this, one or more additional attribute generation and use rounds should be repeated in order to reflect these changes.

6. CONCLUSIONS

This thesis presented a novel method for evaluating mechanical perceived quality. The use of sensory profiling as a means of approaching the problem of classifying and understanding mechanical perceived quality is a new and novel application. The initial hypothesis was that repertory grid method and the flash profile method would be flexible and versatile enough to be successfully applied to evaluating bistable slide phone mechanics. This hypothesis proved to be correct, and most assessors developed from having no previous experience with sensory profiling to comfortably and confidently evaluating the perceived mechanical quality of the given stimuli.

When considering the resulting plots showing PCA correlation loadings and PCA scores for the products over the same principal component space, it is evident that we can draw some preliminary conclusions about what attributes are present and prominent in which phones. Additionally, the consideration of PCA correlation loadings alone and the agglomerative hierarchical clustering plots show a satisfactory degree of consensus between assessors and their individually developed attributes. From this, we can surmise that, despite the fact that individual vocabulary methods have been employed, we have attained some reasonable degree of consensus. Obviously, this consensus could be greatly improved by more rigorous selection of assessors, some means of assessor feedback and training, and perhaps group discussions with all assessors and an impartial moderator.

Furthermore, the internal and external preference maps builds upon the information presented in PCA plots and AHC plots and allow improvements to be made to the highlighted perceptual attributes. Simply put, it would be relatively straightforward to adjust some selected attributes to better fit consumer preferences. The combination of attributes, products, and preference presented in external preference maps is absolutely invaluable in striving to make perceived mechanical quality optimizations.

Three different analysis techniques were considered in this thesis – agglomerative hierarchical clustering, preference mapping, and principal component analysis. Agglomerative hierarchical clustering analysis proved itself to be acceptable as a tool for making preliminary evaluations of whether or not some consensus was achieved between the different assessors. Additionally, AHC is valuable in that it provides a quick snapshot of how different attributes are related and how similar or dissimilar they are. The usefulness of AHC, as a preliminary tool, is quite acceptable. However, in order to extract detailed results about the relationships between attributes and products, some more complex methods are needed. Principal component analysis provides a means of visualizing how different attributes relate to each other, as well as how the products fit in relation to those selected attributes. In this sense, PCA is an invaluable tool for quickly and concisely gaining a deeper understanding of the phenomenon at play.

Finally, in order to visualize the relationship between preference and attributes or preference and products, internal and external preference mapping are invaluable. Simply put, PCA alone is nearly useless if one is striving to make informed changes to a product to better satisfy consumers. Internal and external preference mapping bridges the gap between simply describing the attributes of products and understanding why people prefer those products. Furthermore, internal and external preference mapping allow us to explore the relationship between product attributes and people's affinities for those products.

7. REFERENCES

Addinsoft, XLSTAT, version 2006.06, Paris (2006).

Arnold, G. M., Williams, A. A. 1986 The use of generalised Procrustes techniques in sensory analysis. *Statistical Procedures in Food Research*.

Berg, J., Rumsey, F. (2000) “Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods.” 109th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 48, p. 1106

Borell, K., (1994) “Repertory Grid. En kritisk introduction.” Report, Mid Sweden University. 1994, 21.

Bro, R. “Multilinear PLS” Webpage. http://www.models.kvl.dk/users/rasmus/presentations/Npls_sugar/npls.htm (Accessed May, 2006)

Catell, R.B., Hurley, J.R., (1962) “The Procrustes Program: Producing Direct Rotation to Test a Hypothesized Factor Structure.” Behavioural Science, 7, 258-262.

Dairou, V., Priez, A., Sieffermann, J.M., Danzart, M. (2003) “An Original Method to Predict Brake Feel: A Combination of Design of Experiments and Sensory Science.” Society of Automotive Engineers, Volume 1761, pp 1-8.

Delarue, J., Sieffermann, J.M., (2004) “Sensory mapping using Flash Profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products.” Food Quality and Preference. Volume 15, 2004, 383-392.

Eigenvector Research Corporation, PLS Toolbox, Version 3.5. Washington, USA (2006)

Epting, F.R., Schuman, D.I., Nickeson, K.J. (1971) “An evaluation of elicitation procedures for personal constructs.” British Journal of Psychology, 62: 513-17

Guinard, J.X., Marty, C, (1996) “Acceptability of fat-modified foods in a population of older adults: contrast between sensory preference and purchase intent.” Food Quality and Preference, Volume 7, Number 1, 1996, pp. 21-28(8).

Guinard, J.X., Uotani, B., Schlich, P. (2001)“Internal and external mapping of preferences for commercial lager beers: comparison of hedonic ratings by consumers blind versus with knowledge of brand and price“, Food Quality and Prefence, Volume 12, pp. 243-255.

Guinard, J.X., Yip, D., Elba, C. Cubero, E., Mazzucchelli, R. (1999) “Quality Ratings by Experts, and Relation with Descriptive Analysis Ratings: A Case Study with Beer.” Food Quality and Preference, Volume 10, pp. 59-67.

Hynninen, J., Zacharov, N., “GuineaPig – a generic subjective test system for multichannel audio,” 106th Convention, Munich, Germany, ISO 8586, “Sensory Analysis – General guidance for the selection, training and monitoring of assessors”, Genéve, 1993.

ISO 11035, “Sensory Analysis – Identification and selection of descriptors for establishing a sensory profile by a multidimensional approach”, Genéve, 1994

Jackson, E. “A User’s Guide to Principle Components”, New York: John Wiley & Sons, 1991.

Kelly, G. (1955) “The Psychology of Personal Constructs”. Norton, New York.

Kleef, E. Trijp, H.C., Luning, P. (2005) “Internal versus external preference analysis: An exploratory study on end-user evaluation.” Food Quality and Preference, May 3 2005.
Kunert, J., Qannari, E. “A simple alternative to generalized procrustes analysis: Application to sensory profiling data,” Journal of sensory studies, 1999 - Food & Nutrition Press

Lageat, T., Czellar, S., Laurent, G. (2003) ”Engineering Hedonic Attributes to Generate Perceptions of Luxury: Consumer Perception of an Everyday Sound,“ Marketing Letters, Volume 14, Issue 2, July 2003, Page 97

Lorho, G. (2005) “Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction,” Audio Engineering Society, 119th Convention.

Narain, C., Paterson, A., Reid, E. (2003). “Free choice and conventional profiling of commercial black filter coffees to explore consumer perceptions of character”, Food Quality and Preference, 15, 31 – 41.

OMahony, M. (1986) “Sensory Evaluation of Food”, New York. Marcel Dekker, Inc.

Procrustes, Wikipedia. “Procrustes” Webpage. http://en.wikipedia.org/wiki/Procrustes (Accessed March, 2006)

Smulian J.C.; Sigman R.K.; Tuorila H. Andersson A.; Martikainen A.; Salovaara H. (1998) “Effect of product formula, information and consumer characteristics on the acceptance of a new snack food.” Food

Quality and Preference, Volume 9, Number 5, September 1998, pp. 313-320(8)

Stone, H., Sidel, J.L., Oliver, S., Woolsey, A., Singleton, R.C., (1974) “Sensory evaluation by quantitative descriptive analysis”. Food Tech. 28 (11), 24-34.

Zacharov, N., Lorho, G. (2006) “What are the requirements of a listening panel for evaluating spatial audio quality?” Spatial Audio & Sensory Evaluation Techniquest, Guildford, UK. 2006 April 6-7.

Zamora, M.C., Guirao, M. (2004) “Performance Comparison Between Trained Assessors and Wine Experts using Specific Sensory Attributes.” Journal of Sensory Studies, Volume 19, Pages 530-545.

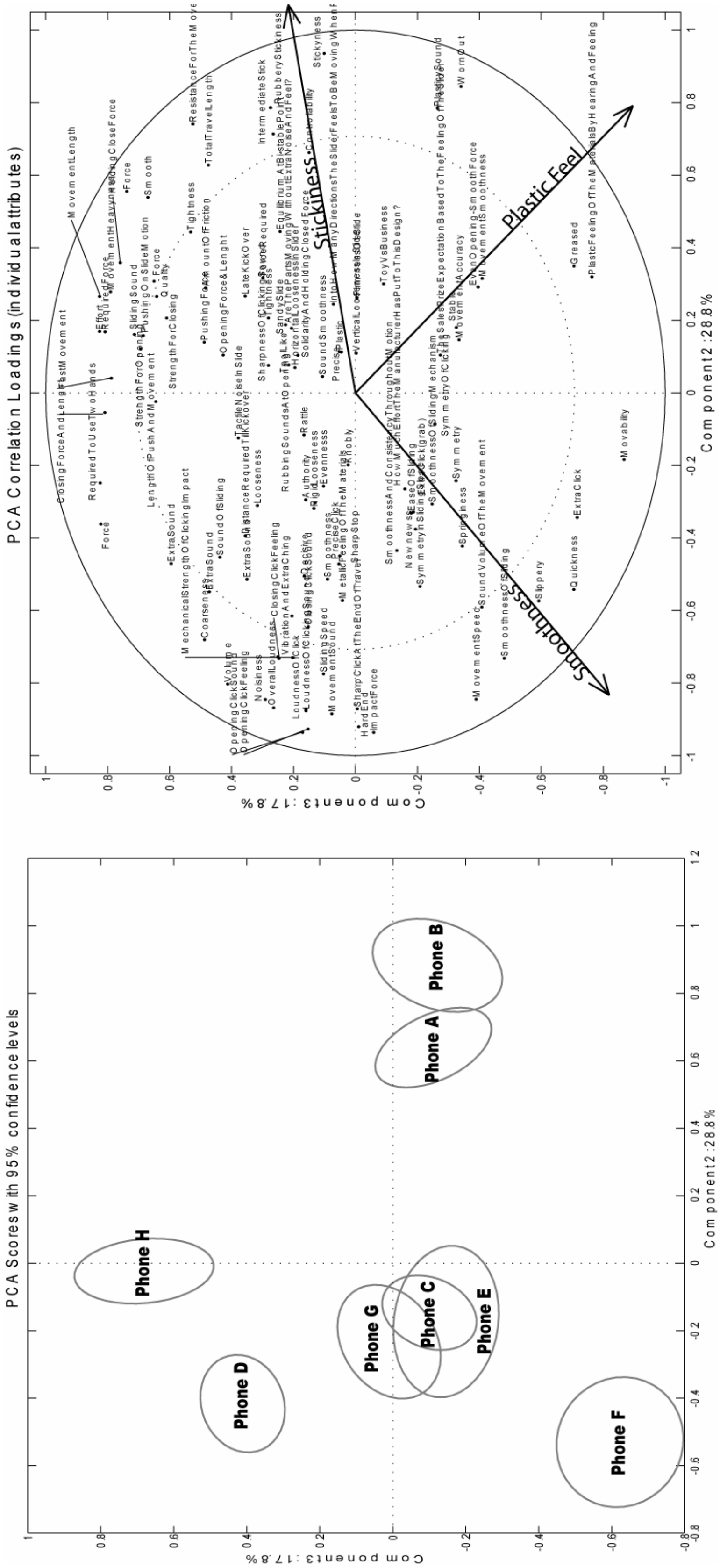
8. APPENDIX

8.1 Appedix A

Category	Attribute name	Word anchors	Assessor	Comments
Force	OpeningForce	small / large	1,9,12	
	ClosingForce	small / large	1,5	
	Symmetry	symmetrical / asymmetrical	1,2	
	StrengthOfImpact	weak / strong	1,3,7,10	
	General Force		2,3,4,6,7,8,10,11	
Sliding	EaseOfSliding	easy / difficult	1,3,12	
	EquilibriumAtBistablePoint	not present / present	1,5,10	Refers to stick in intermediate position
	MovementHeaviness	light / heavy	3,4	
	Friction	high / low	3,4	
Looseness	VerticalLoosnessinSlider	tight / loose	1	
	HorizontalLooseness	tight / loose	1	
	GeneralLooseness	precise / loose	2,3,9,11,12	
	Rigidity	rigid / not rigid	6,7,10,11	
Tactility	Smoothness	smooth / rough	1,4,5,6,7,11,12	
	Coarseness	smooth / coarse	2,5,10,12	
	Plastic	not plastic / plastic	2,7,9,10	
	RubberyFeel	present / not present	5	
	MetallicFeel	present / not present	9	
Auditory	LoudnessOfClick	soft / loud	1,2,4,8,12	
	SymmetryOfClick	symmetrical / asymmetrical	1	
	SharpnessOfClick	short / long	1,5,10	
	MovementSound	loud / soft	3,8,9	
	ExtraSounds	present / not present	5,6,12	
	ExtraClick	present / not present	6	
	Sound smoothness	smooth / scratchy	7,8,9	
Length	MovementLength	short / long	2,4,8	
	PushingLength	short / long	5,8,9	The length needed to push before spring assist takes over
Speed	Movement speed	slow / fast	3,4,6,7	
Unknown	Movement Accuracy	innacurate / accurate	3,4,5	
	Worn out	new / loose	6	
	Controllability	ballistic / controlled	7,8	
	EvennessOfMovement	even / not even	8,11	
	ToyOrTool	Toy / tool	9,10,11	
	PriceExpectation	low / high	9	
	EffortInDesign	low / high	9	

Appendix A. A breakdown of the different categories of attributes produced by the assessors and their frequency of use, shown in the 'Assessor' category.

8.2 Appendix B



Appendix B. Biplot showing PCA scores of products with 95% confidence elipses, and PCA correlation loadings with attributes mapped over the same principal component space for PC2 and PC3.