UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Doctoral Thesis

Statistical Methods in Kansei Engineering Studies

Presented by Lluís Marco Almagro

Supervised by Xavier Tort-Martorell Llabrés

Barcelona, October 2011

To my parents
To Núria and Oriol

Abstract

This PhD thesis deals with Kansei Engineering (KE), a technique for translating emotions elicited by products into technical parameters, and statistical methods that can benefit the discipline. The basic purpose of KE is discovering in which way some properties of a product convey certain emotions in its users. It is a quantitative method, and data is typically collected using questionnaires. Conclusions are reached when analyzing the collected data, normally using some kind of regression analysis.

Kansei Engineering can be placed under the more general area of research of emotional design. The thesis starts justifying the importance of emotional design. As the range of techniques used under the name of Kansei Engineering is rather vast and not very clear, the thesis develops a detailed definition of KE that serves the purpose of delimiting its scope. A model for conducting KE studies is then suggested. The model includes spanning the semantic space – the whole range of emotions the product can elicit – and the space of properties – the technical variables that can be modified in the design phase. After the data collection, the synthesis phase links both spaces; that is, discovers how several properties of the product elicit certain emotions. Each step of the model is explained in detail using a KE study specially performed for this thesis: the fruit juice experiment. The initial model is progressively improved during the thesis and data from the experiment is reanalyzed using the new proposals.

Many practical concerns arise when looking at the above mentioned model for KE studies (among many others, how many participants are used and how the data collection session is conducted). An extensive literature review is done with the aim of answering these and other questions. The most common applications of KE are also depicted, together with comments on particular interesting ideas from several papers. The literature review also serves to list which are the most common tools used in the synthesis phase.

The central part of the thesis focuses precisely in tools for the synthesis phase. Statistical tools such as quantification theory type I and ordinal logistic regression are studied in detail, and several improvements are suggested. In particular, a new graphical way to represent results from an ordinal logistic regression is proposed. An automatic learning technique, rough sets, is introduced and a discussion is included on its adequacy for KE studies. Several sets of simulated data are used to assess the behavior of the suggested statistical techniques, leading to some useful recommendations.

No matter the analysis tools used in the synthesis phase, conclusions are likely to be flawed when the design matrix is not appropriate. A method to evaluate the suitability of design matrices used in KE studies is proposed, based on the use of two new indicators: an orthogonality index and a confusion index. The commonly forgotten role of interactions in KE studies is studied and a method to include an interaction in KE studies is suggested, together with a way to represent it graphically.

Finally, the untreated topic of variability in KE studies is tackled in the last part of the thesis. A method (based in cluster analysis) for finding segments among subjects according to their emotional responses and a way to rank subjects based on their coherence when rating products (using an intraclass correlation coefficient) are proposed. As many users of Kansei Engineering are not specialists in the interpretation of the numerical output from statistical techniques, visual representations for these two new proposals are included to aid understanding.

Acknowledgements

Now that this thesis is finished, it is a pleasure to thank the many people who made it possible.

I want to first thank my supervisor Xavier Tort-Martorell. He is very good in detecting promising areas of research, and he correctly thought Kansei Engineering was an appropriate topic for me. He has always been available to help me with new and exciting ideas, brilliant contributions and, above all, with his friendship. I am especially thankful for his arrangement that allowed me having one semester without any lectures in 2009: that was a great drive for this thesis.

Pere Grima has always given me sound advice on both my professional and personal live in a relaxed and friendly athmosphere. I am very grateful for transmiting me the joy of learning – and teaching – statistics, and for his trust on me from the very beginning of my academic career.

I would like to thank all my colleagues of the Department of Statistics for providing the best possible working athmosphere. Many of them helped me at different points of this work. I have learnt a lot on statistical modeling thanks to Josep Ginebra, Lourdes Rodero, Ignasi Solé and Moises Valls; I even remember a whole Friday afternoon working with Lourdes and Ignasi on an issue related with QT1: thank you. Lourdes Pozueta was also a source of knowledge on design of experiments. In my early days in the Department, I attended the excellent lectures by the late Albert Prat: my memories go to him from these lines.

I had the honour of attending one of Josep's courses on advance statistical modeling in the beginning of my doctorate. Josep has the unique hability of explaining complex topics in a way that is both understandable and enthusiastic: he is the best statistics teacher one can have. I only regret being too immature at that time to fully take advantage of the lectures.

I bothered Josep Anton Sánchez and Xavi Puig many times asking for help with R programming. Jan Graffelman helped me with some concerns related with multivariate analysis. Oriol Camps was the responsible of several inspiring conversations about the idea of quality. Àlex Riba gave me some advice on producing the thesis. I am very grateful to all of them.

A special thank you goes for Pia Margarit, who is always there to efficiently help in the often laborious administrative world, and who has the ability to solve difficulties even before they arise.

It was a pleasure to share the path of writing a PhD thesis with wonderful people like – in more or less order of "appareance" with their PhD – Roberto Behar, Daniel Martínez, Arturo De Zan, Guillermo de Leon, Lesly Acosta, Martí Font and Sara Fontdecaba. Thank you for all the helpful discussions we had.

René Álvarez defended his PhD thesis on Kansei Engineering in December 2009. I am grateful to him for explaining me his proposals in that thesis (specially his procedure for detecting outliers in KE studies).

I am indebted to several students who did research with me in topics related with Kansei Engineering. Irene García Subirats worked with the idea of using simulations for testing several tools used in the synthesis phase. Ana Gómez and Elisabeth Peralta prepared the fruit juice experiment that is used in many chapters of this thesis. Clara Iniesta and Marta Ocaña partially started the research that led to the literature review in this work.

I wish to specially thank Simon Schütte for his generousity in sharing his knowledge on Kansei Engineering and for his friendship. The course Simon gave on KE in our university in April 2007 after our invitation created an exceptional momentum very important for the development of this thesis.

Very special thanks to Margarita Vergara and Shirley Coleman, reviewers of this doctoral thesis. They have both analysed my work really meticulously, being able to detect mistakes only discovered with a very carefull reading. That requires time – a lot – and effort: thank you very much for your generousity.

I am happy to also express my gratitude to Tony Greenfield, who read this thesis and gave me confidence in my skills with English writing.

Lastly, I wish to thank my entire extended family for providing a loving environment for me. A special big thank you for my parents and for Núria and Oriol: to them I dedicate this thesis.

Contents

List of Figures

List of Tables

1 Introduction

This first chapter describes the motivation for studying the topic of Kansei Engineering focusing on statistical methods. The structure of the thesis is also sketched.

1.1. Motivation and Research Objectives

Users of products and services – all of us – are becoming more and more demanding. In this beginning of the XXI century, we do not only want products that work well and satisfy our needs, we also want products that we like. When customers are questioned on what they want, a list of needs normally referring to functionality is obtained. Designers and engineers can translate this voice of the customer into technical parameters, so that the product fulfills those needs. However, customers do not usually explain their emotional needs, probably because they are not aware of having them or are unable to tell which they are. Even when those emotional needs are discovered, it is not obvious which technical properties of the product will elicit those desired emotions.

Kansei Engineering (KE) is a method for incorporating emotions in the product development phase. The main purpose is discovering which technical parameters of a product elicit the chosen emotions. The method was first proposed by Mitsuo Nagamachi. Kansei Engineering has an enormous pontentiality and the underneath idea is brilliant. But two main difficulties arise when reading the available literature on the topic:

- Many different methodologies used in product design, some of them qualitative and other quantitave, appear under the name Kansei Engineering. In fact, it is not clear what a Kansei Engineering study really is.
- Some statistical tools used in Kansei Engineering are not the most adequate; even when the techniques used are appropriate, they are sometimes wrongly applied. In fact, the discipline could benefit from a more extensive use of statistical methods.

This thesis attempts to overcome these two difficulties. The main objectives are the following:

1. Proposing a definition of Kansei Engineering that delimits its scope.
2. Justifying the importance of Kansei Engineering, linking it with quality.
3. Devising a model for conducting Kansei Engineering studies that makes emphasis in data collection and analysis (a quantitative approach).
4. Revealing how Kansei Engineering studies are usually performed on the basis of an exhaustive literature review.
5. Describing the most used statistical techniques for Kansei Engineering, enhancing them if possible.
6. Suggesting new statistical methods appropriate for Kansei Engineering.
7. Addressing variability in Kansei Engineering studies in a way that allows the extraction of richer information.
8. Making suggestions on how to present results from KE studies in a visual way, easily understandable by non-statisticians.

The thesis uses both simulated and real data as a basis for conducting and presenting the research. Real data comes from a Kansei Engineering study called the fruit juice experiment. This KE study was done in an academic environment exclusively for this work. Data from this experiment is analyzed and reanalyzed in several parts of this thesis with progressively higher levels of complexity.

Statistics plays an irreplaceable role in my view of Kansei Engineering. In fact, statistics are always essential when dealing with numerical data (or what is the same, when dealing with applying the scientific method). And applying statistical methods without specialized software is nowadays inconceivable. I have basically used `R` (R Development Core Team 2010) as statistical software in this work. Each `R` package used is cited in its first appearance in the text. Other software packages employed have been MINITAB (versions 14, 15 and 16) and SPSS (version 17).

1.2. Organization of the thesis

Figure 1.1 shows a diagram summarizing the structure of this thesis.

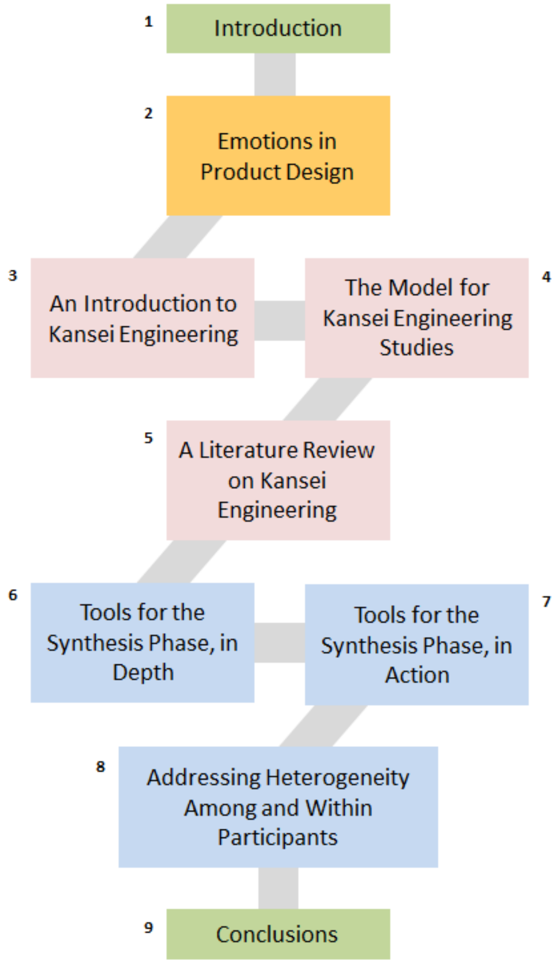


Figure 1.1. Organization of the thesis, in chapters.

After this first introductory chapter, Chapter 2 justifies the importance of the topic, linking it with product design and quality. Chapter 3 introduces Kansei Engineering as a quantitative technique for product design, and discusses how to measure emotions. Chapter 4 describes a model for conducting Kansei Engineering studies using the fruit juice experiment as an example. Chapter 5 contains an exhaustive literature review on the topic. This chapter answers many questions on how KE studies are usually conducted, and depicts the most common used techniques. Chapters 4 and 5 together constitute an excellent portray of Kansei Engineering studies. Chapter 6 takes the most commonly used techniques identified in the previous chapter and studies them in detail. Several enhacements are also suggested. Chapter 7 uses simulations for discovering how the different techniques compete, and which the preferred ones are. This chapter also makes some sound proposals of steps to be incorporated in the model for conducting KE studies. Chapter 8 addresses variability in Kansei Engineering. This chapter introduces and develops several innovative ideas that are later incorporated in the model. Finally, Chapter 9 summarizes this work, presents the final complete model for KE studies and suggests further research topics on Kansei Engineering.

Table 1.1 maps the chapters with the objectives of the thesis.

Chapter	Objectives
2	2
3	1
4	3
5	4
6	5, 6, 8
7	6, 8
8	6, 7, 8
9	3, 8

Table 1.1. Mapping of chapters and objectives in the thesis.

2 Emotions in Product Design

This chapter discusses the importance emotions have in our daily lives, in general, and in products design, in particular. Emotional design is also linked with quality.

2.1. The role of emotions in our lives

During many years, there has been a tendency in western cultures to oppose cognition against emotion. Cognition was supposed to be rational and it was thus encouraged, whereas emotions were to be set aside. Emotions remained as part of our animal origins, and had no place in our sophisticated and highly scientific-oriented society. Especially when making decisions, it was important to stay composed and rational to avoid mistakes.

But research in the last decades has shown that these ideas regarding emotions were wrong. Emotions are inseparable from cognition and play an important role in our daily lives. Neurologist Antonio Damasio describes in his 1994 book "Descartes' Error" (Damasio 1994) cases of patients that had damage in the prefrontal cortex of the brain. Although neither their reasoning capabilities, nor their motor control, nor their language skills nor their ability to learn were impaired, their lives changed completely. Conventional IQ tests did not detect any abnormalities, but relatives and friends reported that they were different from before. They became aloof in their relationships with others, and had difficulties in deciding on simple things, such as the restaurant to go to for dinner or which dessert to choose.

Damasio explains this inability to choose using his somatic marker hypothesis (Damasio, Everitt & Bishop 1996)[1]. Sometimes, decisions must be made between different alternatives, with high degrees of uncertainty. In such cases (and in some others), each stimulus induces an affective state, either positive or negative, that guides (or at least biases) our decision. This guidance appears at an unconscious level, before we have "rationally" made our decision.

Some somatic markers are with us since our birth (they are "preinstalled" and come from thousands of years of evolution); others are recorded as we experience life. When we store a memory of an action, an associated emotion is also stored. So not only do we consciously remember previous experiences in our lives when we come across a similar situation, emotions are also sparked beneath our conscious awareness and guide us.

The Iowa gambling test is a seminal experiment designed by Antonio Damasio to prove the somatic marker hypothesis. Players have to pick cards from four different decks. Each card has an economic reward or an economic penalty. There are two favorable decks and two disadvantageous decks. The aim of the game is making as much profit as possible (or losing as little money as possible) in the 100 allowed turns. Normal people take cards from all the decks, but after some draws (20 or 30) they start biasing to the favorable decks. Emotionally impaired people (mainly with damages to the prefrontal cortex) are unable to make this decision, and continue to take cards from the disadvantageous decks until the end.

But what is really remarkable is that, when measuring skin conductance in normal players, a reaction is detected when choosing cards from the disadvantageous decks even before the player consciously disregards these decks. No variation in skin conductance is found in emotionally impaired players.

So emotions are there to guide us. When an object falls on us, an emotion of fear is triggered, and we immediately move to avoid this falling object. This is done even before we feel the emotion of fear. Emotions are biological "intelligence" already prepared to make an organism do the best it can without having to think too much.

[1] The explanation about the somatic marker hypothesis and the Iowa gambling test comes from the Damasio's book "Descartes' error", but also from the lecture dictated by Antonio Damasio in the Aspen Institute on 7th April 2009, and available online at fora.tv (http://fora.tv/2009/07/04/Antonio_Damasio_This_Time_With_Feeling, accessed September 2009)

Thinking a lot in those situations could mean wasting time, which could have a high price.

However, emotions are there for the good and for the bad. Many times emotions lead to correct decisions, but this is not always the case. Dan Ariely, a professor of behavioral economics, explains in his 2008 book "Predictably Irrational" (Ariely 2008) many real experiments where decisions are less than optimal. Behavioral economics tries to explain why, sometimes, economic decisions are not made under the rules of classical economic theory. To do this, psychology is integrated into classical economic theory, and emotions play an important role. The field of behavioral economics has increased in popularity over the last years (especially after Daniel Kahneman, one of the most prominent names in the field, was awarded the 2002 Nobel Memorial Prize in Economics).

One of the discussions by Dan Ariely refers to the irresistible attraction we all have to free stuff. In one of his experiments, he offered people in a mall in Boston a choice between a free $10 Amazon gift certificate and a $20 gift certificate for 7 dollars (Ariely 2008, p. 58). In fast answers, most of the people took the $10 gift certificate (a profit of $10 instead of the $13 profit from the other option). This "irrational" response probably comes from the fact that the profits from free goods are clear (you get what is offered without losing anything), whereas it is not that clear what you gain when you also lose some money. Anyway, the emotion of excitement is triggered promptly when we hear the word *free*, seconds before we can start thinking about the choices (and seconds is an eternity in the realm of the brain, where milliseconds are the common unit of time measurement).

What I find especially relevant from the ideas of Dan Ariely is that these "irrational" behaviors are predictable. We make the same mistakes over and over, especially if we are not aware of the possibility of making errors. I think this idea of our irrationalities being predictable link well with the somatic marker hypothesis, and, what is more important for this dissertation, with Kansei Engineering's aim of connecting physical properties of products with the emotions they elicit.

Kansei Engineering, as we will see, tries to connect physical properties of products with the emotions they elicit. Of course, this would only be possible if the same emotions are consistently (predictably) triggered when a person interacts with a product with certain properties. It is clear that the emotion triggering mechanism described by the somatic marker hypothesis is universal (disregarding people with brain damages, particularly in the prefrontal cortex). However, our experience and common sense tells us that emotions evoked by products are personal. Looking at the same watch, a 75 year old man might find it modern whereas a 16 year old teen might find it classical. However, this fact is overcome in Kansei Engineering by defining a target group for which the study is valid. The same emotions are supposed to be provoked by the designed product in the wide range of population that belongs to the target group.

According to Dan Ariely, we are "predictably irrational" because a wide group of people (although not everybody), systematically reacts by choosing the free $10 gift certificate. Likewise, a wide group of people (hopefully the ones in the target group) are predictable because they experience the same emotion when exposed to a product with certain properties (and thanks to Kansei Engineering, we know which properties evoke which emotions)[2].

Therefore, theories from Antonio Damasio and Dan Ariely promisingly support a key assumption in Kansei Engineering: certain product properties will elicit expected emotions in users.

An aside on emotion and feeling, sensation and perception

Disciplines deeply involved in the study of the mind (neuroscience, cognitive psychology, philosophy of mind) try to use words such as *emotion*, *feeling*, *sensation* and *perception* with precision.

According to Antonio Damasio, *emotion* is a collection of automated actions that are aimed at a particular effect that is important for the regulation of life. In case of a threat, we immediately freeze in place or run away from the source of threat. We take on a posture of fear, heart rate goes up, blood pressure increases, and hypothalamus spreads cortisol into the entire organism. Emotion is, by definition, nonconscious.

Because we have a mind, we have the potential of having a perceptual assumption on what is happening to us. When we have a *feeling* of fear, what that means is that our mind is representing what has changed in our organism; we have a number of thoughts and a number of cognitive strategies that we can engage. Feelings come after emotion, and now in the conscious level. Many species distant from humans have emotions – snakes, for example. But it is hard to believe that snakes have feelings.

---

[2] In fact, perhaps things are not that simple… What happens if not everybody in the target group experiences the same emotions? Although the target group is homogenous according to criteria (socio-economic position, age, gender), are people also homogenous in their emotional response? The answer could perfectly be no. A solution to this fact, based on the idea of segmenting population according to their emotional reaction to a product, is proposed in Chapter 8 of this dissertation.

Perception is defined as the process of registering sensory stimuli as meaningful experience[3]. Usually they are distinguished from *sensations* stating that sensations are simple sensory experiences, whereas perceptions are complex constructions of simple elements joined through association. Perceptions are also more subject to the influence of learning. With these definitions, our perception of an emotion would be a feeling.

In the area known as emotional design (or affective design), and also in the more specific field of Kansei Engineering, these words are generally used, in my opinion, in a much more relaxed way. The motto "incorporating emotions into products", usually stated as the purpose of emotional design, is in fact very ambitious, if we take into account the definition of emotion given by Antonio Damasio. I think that when emotions are taken into consideration in product design (especially in Kansei Engineering), we are really thinking about giving users some perceptions in a non-rational way. For example, we design a car that gives the perception of being powerful, even before the user drives it, or knows the objective fact of the horsepower the car has.

Some authors emphasize how psychology defines and explains emotions. Pieter Desmet, in his PhD dissertation "Defining Emotions" (Desmet 2002), presents different traditions in the research of emotion, contemplates how to differentiate emotions, and questions whether product emotions are "real" emotions (that is, the same emotions we experience towards important events in our lives[4]). In this case, this preoccupation in precise definitions is justified: his work focuses on selecting which emotions (such as desire, satisfaction, boredom…) are enough for describing the full emotional reaction elicited by a product and how to measure them.

But in the literature on Kansei Engineering, this deep discussion on what emotions are is not common. The elicited "emotions" in Kansei Engineering are defined by the so-called Kansei words[5], which usually are adjectives describing how the product is perceived. Although the words emotions and feelings are also used, they in fact refer to perceptions almost all the time.

I will attribute this work to the "relaxed" use of words related with emotions, and I will use the words emotions and perceptions as basically synonymous.

---

[3] According to Encyclopædia Britannica (http://www.britannica.com/EBchecked/topic/451015/perception, retrieved in September 2009)

[4] The main conclusion of the first part of Pieter Desmet's dissertation is that product emotions are not a special type of emotions, but are as "real" as, and have the same qualities as the emotions we experience in our daily lives.

[5] The use of Kansei words will be further explained in section 2.4.

2.2. Why integrating emotions into product design is valuable

We have just seen in the previous section that emotions play an important role in our lives. As we interact with products on a daily basis – either making buying decisions or using them – it seems logical to incorporate emotions, in some way, in the design and improvement of products. Here, the concept of product mainly refers to physically manufactured products (a mobile phone, a car, a bottle of water...), but also to what is commonly known as services (medical care, internet access...)[6].

For many years, designers were not very interested in emotions. The focus was more on making usable products. Donald Norman, a professor of cognitive science and usability guru, describes in his 1988 book "The Design of Everyday Things" (Norman 1998) the psychology behind what he considers good and bad design. He details problems that bad design can cause when a product is not easy to use, and thus proposes several design principles to make usable products. In his book, Norman uses the term "user-centered design" to define a design based on the needs, wants and limitations of the user, not paying attention to what he considers secondary issues, such as aesthetics.

But making products that work well and fulfill user expectations is not enough. When analyzing the products we normally use, we realize that we love some gadgets that are far from perfect, but that we just like them. Or we have a deep appreciation for a product because of the person that gave it to us as a present, or because it reminds us of the good times we had using it. These emotional aspects attached to products cannot be disregarded.

Donald Norman also noticed that, and in his 2003 book "Emotional Design" (Norman 2004) he states:

> *When I wrote "The Design of Everyday Things", my intention was not to denigrate aesthetics or emotion. I simply wanted to elevate usability to its proper place in the design world, alongside beauty and function. I thought that the topic of aesthetics was well-covered elsewhere, so I neglected it. The result has been the well-deserved criticism from designers: "If we were to follow Norman's prescription, our designs would be all usable – but they would also be ugly".*
>
> *Usable but ugly. That's a pretty harsh judgment. Alas, the critique is valid. Usable designs are not necessarily enjoyable to use.*

---

[6] However, many explanations in this chapter are hard to translate to services.

Many of us would agree that beauty in products is a desired property by itself. But, does this sole fact make it worth the effort of integrating emotions into products design? Probably, yes. But there are two practical reasons that also justify the endeavor: attractive products sell better and work better.

An aside on the distinction between user and purchaser

In the thesis by Simon Schütte (2005, p.22), an interesting distinction is made between user and purchaser. Purchaser is the person that buys a product, whereas user is the person that actually interacts with the product. Many times both buyer and user are the same person: this is usually the case with products bought for personal use. But products for use in the working place are commonly bought by a professionally employed purchaser. Sometimes, also products for personal use are not bought by the final user, as happens with gifts.

2.2.1. Attractive products sell better

There is evidence that in today's saturated market, there are many products that have similar functionality and price. Looking at the basic mobile phones in Figure 2.1, which of the four 39 € phones is better?

Figure 2.1. Basic mobile phones taken from The Phone House catalogue in September 2009

As the price is exactly the same, a rational way to answer the above question is comparing its technical specifications. However, looking at Table 2.1, we realize that this technical comparison is not very helpful. Of course, there are some differences among the models. Both Nokia models have radio, whereas the LG and Samsung do not have it. But if we are not interested in having radio in the mobile phone, all other characteristics, considered together, are similar. Perhaps one mobile slightly excels in comparison with the others in a given feature, but it is worse in another.

Table 2.1. Specification comparison among four basic mobile phones

		LG GB102	Nokia 1650	Nokia 1661	Samsung E1120
General	2G Network	GSM 900 / 1800	GSM 900 / 1800	GSM 900 / 1800 GSM 850 / 1900	GSM 900 / 1800
Size	Dimensions	103 x 45 x 14 mm	104 x 44 x 18 mm	108 x 45 x 13.8 mm	105.6 x 43 x 15.1 mm
	Weight	70 g	80 g	82 g	66,4 g
Display	Type	CSTN, 65K colors	CSTN, 65K colors	TFT, 65K colors	CSTN, 65K colors
	Size	128 x 128 pixels 1.5 inches	128 x 160 pixels 1.8 inches	128 x 160 pixels 1.8 inches	128 x 128 pixels 1.52 inches
Sound	Alert Types	Vibration Polyphonic ringtones	Vibration Polyphonic ringtones MP3 ringtones	Vibration Polyphonic ringtones	Vibration Polyphonic ringtones MP3 ringtones
	Speaker phone	Yes	Yes	Yes	Yes
Memory	Call records	10 dialed, 10 received, 10 missed calls	20 dialed, 20 received, 20 missed calls	20 dialed, 20 received, 20 missed calls	20 dialed, 20 received, 20 missed calls
Features	Messaging	SMS	SMS, EMS, Instant Messaging	SMS	SMS, EMS, MMS
	Radio	No	Yes	Yes	No
	Games	Yes	Yes	Yes	Yes
Battery	Stand-by	Up to 400 h	Up to 420 h	Up to 475 h	Up to 540 h
	Talking time	Up to 6 h 30 min	Up to 8 h	Up to 4 h 10 min	Up to 8 h 30 min

In this situation of equivalent functionality and price in many products, so common in today's marketplace, emotional impact can be crucial. A user in a shop probably chooses the most appealing mobile phone, the one that he or she prefers (although it is difficult to say why it is preferred[7]).

This positive reaction to attractive products in the market can be explained by the theory of elicited emotions described by Donald Norman (2004, pp.63-98). Norman describes three components in his framework: visceral, behavioral and reflective. Visceral level refers primarily to the initial impact, to its appearance. Behavioral level is about performance when using the product (function, understandability, usability, and physical feel). Reflective level is about the meaning of a product, the personal remembrances the product evokes, how it makes one feel, the image it portrays, the message it tells others about the owner's taste.

---

[7] Even though the buyer is not able to describe why he or she likes a specific gadget, it is possible to design the product to be attractive and to elicit the desired emotions. This is precisely the aim of emotional design!

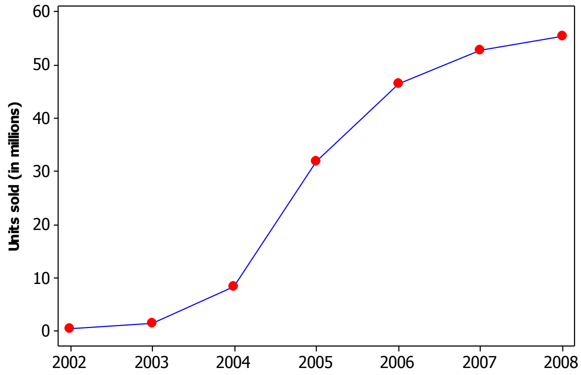


Figure 2.2. iPod units sold per year, in millions (data from Wikipedia.org, retrieved in September 2009)

Although a product elicits reactions on the three levels, the visceral level (our first reaction at an unconscious stage) is central for a product being selected from the shelf. Furthermore, people are quite similar all over the world on this visceral level (behavioral and reflective level depend much more on education and experiences, thus is less robust to cultural differences) (Norman 2004, p.39) *"Beauty may be a driving force to become the owner of a product"* (Hassenzahl 2004).

A typical example of "emotional product" that has been a blockbuster is Apple's iPod. Figure 2.2 shows its tremendous rise in sales, especially from 2004 to 2006. In January 2007 the iPod market share reached 72.7%[8]. Its success stems from its appearance and interface; more than from its functionality (many other mp3 players have the same or even better functionality).

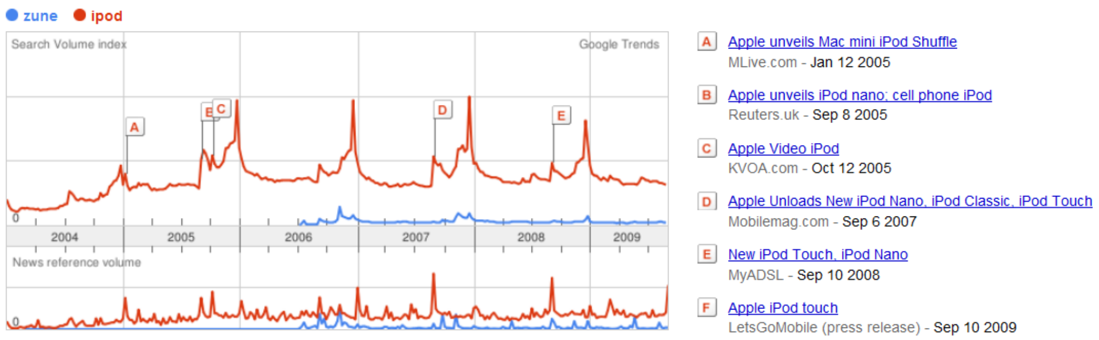


Figure 2.3. Search volume index for "zune" and "ipod" using Google Trends

Attempts to compete with iPod had not been very successful until 2009. Microsoft released Zune in 2006 addressing the same market as iPod. But neither its sales nor

---

[8] According to Wikipedia.org ("IPod," *Wikipedia, The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=IPod&oldid=313500009, accessed September 2009)

the media coverage it received were very promising. Figure 2.3 shows the search volume index for “zune” and “ipod” using Google: interest for Apple iPod has always been much greater than for Microsoft Zune.

Another classical example of emotional design is the roadster Mazda MX-5 (also known as Mazda Miata). The first Mazda MX-5 appeared in 1989 (Figure 2.4). Mitsuo Nagamachi, founder of Kansei Engineering, personally collaborated in the initial design of Mazda MX-5. The target market of Mazda MX-5 was young adults. Pictures were taken and videos recorded of young drivers driving cars in order to discover their behavior. Based on the analysis of this material, each behavior was described on a card, and all cards were grouped in a tree structure[9]. The concept of “unification of driver and machine” was identified as the zero-level concept (Nagamachi 2002). From here, the concept was broken down into subconcepts, until these subconcepts were translated into physical properties in the car. Every generation of Mazda MX-5 has followed this “oneness of horse and rider" credo.

Figure 2.4. The first generation of Mazda MX-5 (photo from Wikipedia.org, retrieved in September 2009)

The result has been a compact and light roadster that has gained the favor of the market. From 1989 to 2009, more than 850,000 units have been sold, making it the best-selling sports car ever[10].

[9] Some ideas for describing this short history of the conception of Mazda MX-5 design have been extracted from http://www.qfdi.org/lifestyle_qfd_and_kanseiengineering_miata.htm, accessed September 2009

[10] According to the Guiness Book of World Records.

2.2.2. Emotional products work better

During many years, designers were not only uninterested in beauty in products, but also suspicious. “If it is pretty, it won’t work” summarizes one of the common prejudices (Hassenzahl 2004). However, this statement is not necessarily true.

In 1995, Japanese researchers Masaaki Kurosu and Kaori Kashimura conducted an experiment to explore the relationship between the perception of the ease of use of an automatic teller machine (ATM) before actually using it (they called this “apparent usability”) and the appearance of the interface (Kurosu, Kashimura 1995). To their surprise, a high correlation was found between the aesthetics of the interface and the apparent usability. Of course, what is really important is the real usability of a product (called “inherent usability” by Kurosu and Kashimura), more than the apparent usability. However, first impressions count, and some studies state that users are more likely to express dissatisfaction with an interface after long term use if it was initially perceived as difficult to use (Hiltz, Johnson 1990)

Noam Tractinsky, an Israeli professor of Information Systems, was surprised with the results and in 2000 decided to replicate the experiment in Israel (Tractinsky 1997). The purpose of his study was to demonstrate that the correlations found by his Japanese colleagues were culturally specific, and that correlation between aesthetics in the interface and apparent usability would be less important in Israel than in Japan. On the contrary, those correlations were even higher[11].

Tractinsky went a step further. He repeated the ATM experiment, but now measuring not only the initial perceptions on interface aesthetics and apparent usability, but also inherent usability (that is, the ease of use when actually interacting with the ATM). Inherent usability was measured with objective magnitudes such as error rates and time to complete a task. Beautiful interfaces were not only rated as easy to use before touching them, but were also more usable (less errors, faster, easier to learn) when actually interacting with them. Probably, the conclusions were so inspiring that Tractinsky decided to entitle the paper with these results with the unambiguous claim “What is beautiful is usable” (Tractinsky, Katz & Ikar 2000).

But, why do attractive products work better? In his book about emotional design, Donald Norman (2004) exposes the following mechanism:

- Attractive products make people feel good, leading to a positive mood.

---

[11] More details about this story of reproducing the experiment with the ATM and the conclusions that were derived can be found in Chapter 1 of the influential book by Donald Norman “Emotional Design” (Norman 2004).

- People in a good mood, who are relaxed and happy, become more creative and imaginative.
- This creativity makes the finding of solutions to encountered problems with a product easier, so tasks performance is improved.

Problems when using beautiful products are compensated thanks to the positive mood these beautiful products arouse.

The fact that people's mood influences performance on many tasks seems quite proven. Some psychologists state that positive mood improves creative problem solving, working memory and consolidation of long-term memory due to increased brain dopamine levels (Isen, Ashby & Turken 1999).

The fact that products rated as attractive before being used induce a positive mood, and that this positive mood remains when using the product, even if faced with problems, is more debatable. In fact, some authors state that when we have problems with a product, we may prefer to stop interaction with it than compensate the problems (Schifferstein, Hekkert 2007, pp. 287-302)

Some critiques to methodological aspects in experiments such as the ATM test by Noam Tractinsky have also been reported, questioning the high correlation between perceived beauty and inherent usability (Hassenzahl 2004). Even if these high correlations really existed, the mechanism proposed by Donald Norman to justify why beautiful products work better has been accused of confounding correlation with causality.

At any rate, the assertion that attractive products work better is so appealing to researchers that undoubtedly more studies will be conducted to determine not only its veracity but also to clarify the mechanism that makes it possible. What many studies definitely show is that beauty is a good (usually the best) predictor of a product's overall impression or general user satisfaction (Hassenzahl 2004). And user satisfaction has a lot to do with quality: I will go deeper on this issue in the next section.

2.3. Emotional design: the last step in the quality revolution

During the 20th century, production changed from a product-out to a market-in paradigm, and the growing importance of quality is what was responsible. In fact, what is understood by quality has changed together with the transformation experienced by manufacturing companies. The Japanese management guru Shoji

Shiba explains this transformation in his book "Four practical revolutions in management" (Shiba, Walden & Graham 2001, pp. 3-17).

Professor Shiba starts his review of the history of quality with the following statement:

> *Customer satisfaction and quality can be thought of as two different names for the same thing. When customers are asked to define what quality means to them, in general what they mean is what it takes to satisfy them.*

He traces the evolution of quality describing four levels of quality, called the four fitnesses. Each level has been achieved once the previous ones have been consolidated. These four fitnesses are fitness to standard, fitness to use, fitness of cost and fitness to latent requirements:

- Fitness to standard evaluates if a product built as designed passes the standard. In other words, quality is defined as having products that do what the designers intend them to do; products that, simply, work. This was assured by inspection or – in order to avoid checking products when they were already built – monitoring process and product variables with statistical process control. Inspection has the problem of increasing costs and that workers see inspectors as enemies. But the worst difficulty is that market needs are not considered.
- Fitness to use means to assure satisfaction of market needs. Products must be useful to people, they have to fulfill their expectations. Of course, this requires listening to customers; and market research techniques have been developed to get the "voice of the customer".
  The competitive advantage a company has when reaching this level lasts for a while, but sooner or later other companies will create equally good products fulfilling the same needs, and perhaps cheaper. Time to go to the next level.
- Fitness of cost means having high quality, but low cost. This can only be achieved by systematically collecting data and analyzing it in order to improve processes. In fact, successful quality initiatives such as Six Sigma – which makes an intensive use of statistics – focuses on projects aimed at increasing quality and reducing costs.
- Fitness to latent requirements means meeting customer needs before customers are aware of those needs. With this "attractive quality" a company may achieve a monopoly for a short time, and ask for a higher price.

Although the four fitnesses have been historically obtained one by one, nowadays companies often have to implement all four fitnesses simultaneously (but perhaps not for all the products). In any case, the later fitnesses are not better than the earlier fitnesses: all of them are important, but the later ones are built on top of the earlier ones. When this order is broken, the product will fail. For example, a low price watch may meet the fitness of cost quality, but it is worthless if unable to give time correctly, meeting the fitness of standard.

Quality function deployment (QFD) can be a useful tool for entering in the fitness to latent requirements level of quality. QFD, and especially its simplified form of house of quality, is a commonly used tool in order to translate customer needs (those vaguely expressed by customers with words) into technical parameters to be met by a new product design (expressed by engineers with numbers and specifications). In the house of quality, the left part contains the customer requirements (the "whats") and the top part the technical characteristics (the "hows"). The interrelationship matrix, in the middle, connects customer requirements with technical characteristics. Usually, symbols are used indicating a strong relationship, a medium relationship, or a weak relationship. Correlations between technical parameters are stated in the roof of the house.

Besides considering customers' requirements that can be explained explicitly with words, why not also add some emotional needs to the list not expressed by customers (perhaps even not imagined)? Figure 2.5 shows this idea. I think tools from emotional design like Kansei Engineering try to achieve just this: detecting emotional needs and translating them into technical parameters, so that products fulfill these latent emotional requirements. Emotional design belongs to this last step in the quality revolution: the fitness to latent requirements.

When thinking about consumer needs, a hierarchy can be defined, as proposed by designer Patrick Jordan (2002). The hierarchy of consumer needs described by Jordan comprises the following three levels:

1. Functionality: A product is useless if it does not contain the functions necessary to perform the tasks for which it is intended. If a product does not work, users will be dissatisfied, regardless of any other consideration. Designers must understand how the product will be used in order to provide the appropriate functionality.
2. Usability: Once people are used to functionality, they also want these products to be easy to use. Functionality is a prerequisite for usability. Many products are now in this stage, and usability has received wide attention by designers during the last decades.

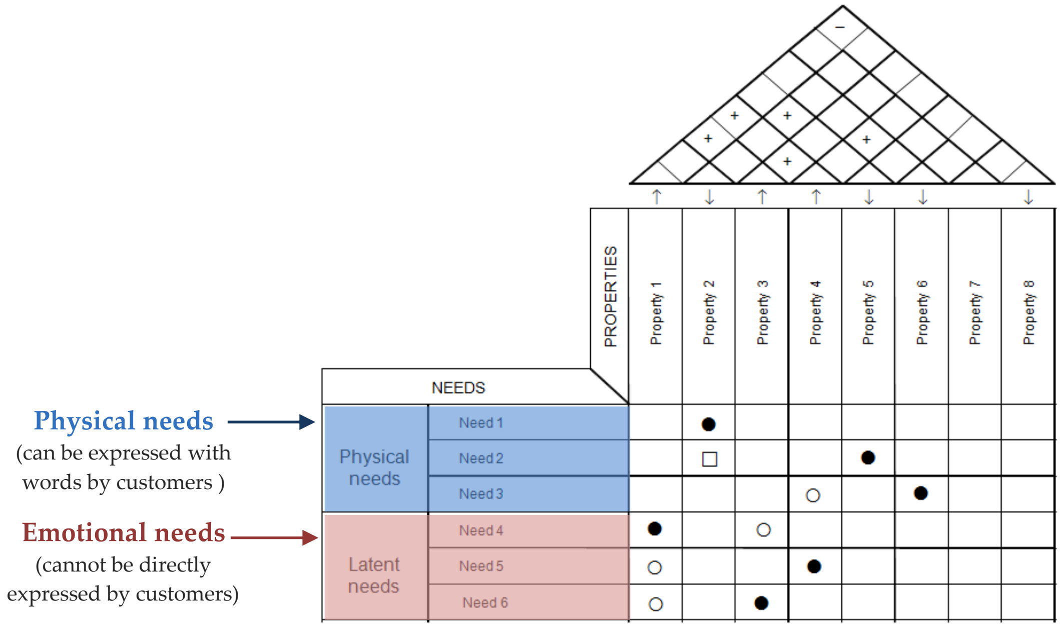


Figure 2.5. Part of a house of quality, where the emotional needs have been added below the physical needs

3. Attractiveness: Having become used to usable products, people will want something more: beautiful products that fulfill not only our needs of functionality, but also our emotional needs.

The three levels proposed by Jordan are somehow parallel to the hierarchy of needs proposed by psychologist Abraham Maslow (1943). In Maslow's hierarchy of needs the lowest level is associated with physiological needs, while the uppermost level is associated with self-actualization needs. The higher needs in Maslow's hierarchy come into focus when the lower needs in the pyramid are met[12]. Similarly, in our consumer needs pyramid (Figure 2.6), when people get used to having something, they start looking for something more.

This hierarchy of consumer needs can be linked with the model for customer satisfaction proposed by Noriaki Kano in the 1980's (Kano et al. 1984). In the Kano model, the horizontal axis displays the degree of achievement and the vertical axis the customer satisfaction. Certain properties in a product must be there (security issues, for example), so people are not particularly satisfied when present and usually are not reported when asked. On the contrary, customers are aware of one-dimensional quality, as many times this relates to non standard properties, so it is

[12] More information on this topic can be found in the entry "Maslow's hierarchy of needs" in Wikipedia ("Maslow's hierarchy of needs." *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=Maslow%27s_hierarchy_of_needs&oldid=316622771, accessed September 2009)

usually expressed as desirable. Attractive quality is not expressed by people because it is not expected, but it is well received when present.

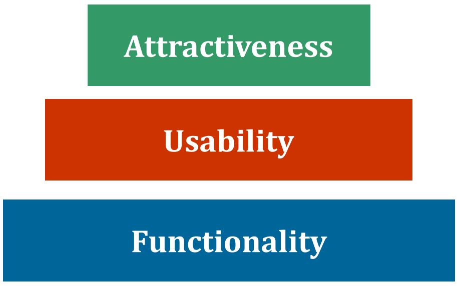


Figure 2.6. A hierarchy of consumer needs, as proposed by Jordan (2002)

Product properties have a life cycle following this model. In the beginning, when newly introduced, a property is perceived as attractive. When people get used to it it turns into one-dimensional quality, and finally, the property becomes so common and widespread that it becomes must-be quality. A very typical example of this quality life cycle is the remote control for a television. In the 1970's, this device was considered to have attractive quality, as it was something new and unexpected. In the 1980's it became more and more common, so it moved to one-dimensional quality. Since the 1990's, all televisions have been sold with remote controls, so this device is now must-be quality.

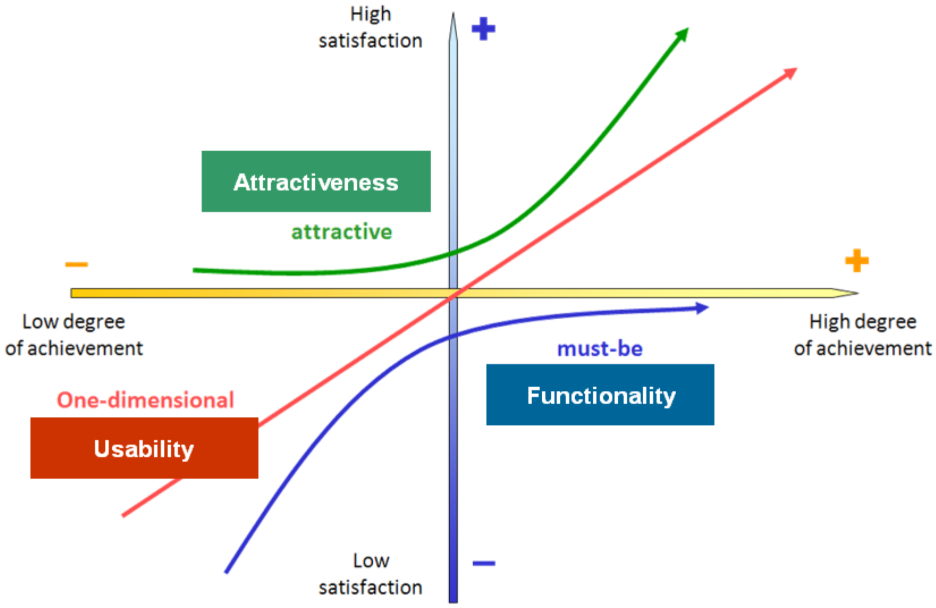


Figure 2.7. Establishing a link between the Kano model and the hierarchy of consumer needs

In general, at our current state of development, functionality is must-be quality: everybody expects a product to work as promised. But some products are easier to use than others, and people recognize these products as having "better design" than others. Usability is, thus, one-dimensional quality. Having attractive products that evoke pleasant emotions is still attractive quality. As attractive quality can give a competitive advantage to companies, they should focus on designing attractive products (of course, keeping the previous quality levels). Figure 2.7 shows this relationship between the Kano model and the hierarchy of consumer needs.

This life cycle is simply ascending the pyramid of consumer needs shown in Figure 2.6, and can be followed for every new product in the market, or even for new properties in existing products. What will happen when products have followed these three steps of functionality, usability and attractiveness en masse? Probably some other characteristic will occupy the highest step, for example, social corporate responsibility[13].

The example of Microsoft Windows: from simply working to an emotional product

The evolution of Microsoft Windows operative system is a good example of this path from functionality, to usability and attractiveness.[14]

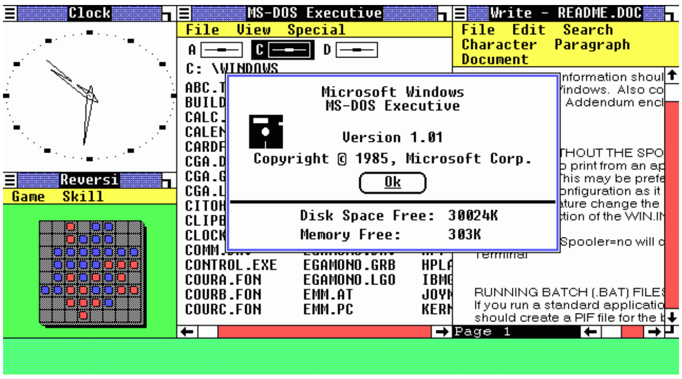


Figure 2.8. A screenshot from Windows 1.01 (1985). It essentially had a clock, a calendar, some games, a notepad and a text editor

First versions of Windows (both Windows 1.0 and 2.0) were very rudimentary (Figure 2.8). But it worked well in the tasks the engineers had planned; it was adjusted to the standards: operations done with the calculator were correct, the clock

[13] More and more people are now concerned with environmental issues and working conditions, and will be reluctant to buy products from companies that do not share these concerns, regardless of the quality of the product.

[14] Information for this extremely short history of Microsoft Windows has been largely retrieved from Wikipedia ("Microsoft Windows," *Wikipedia, The Free Encyclopedia,* http://en.wikipedia.org/w/index.php?title=Microsoft_Windows&oldid=305810722, accessed August 2009) and the article by Michael Miller in http://www.informit.com/articles/article.aspx?p=1358665, accessed August 2009

showed the time correctly, it was possible to write a text using a simple text editor, etc. These first two versions of Windows were not successful. There was not a lot of demand for a graphical user interface when applications being used (basically in companies, not yet in the home those years) were text-based.

Probably Windows 3.0, released in 1990, and Windows 3.1 (Figure 2.9), released in 1992, were a milestone in the history of this product. For the first time Windows started to give users what they wanted: interaction with the computer was made simpler, it was possible to use more than one program at a time, the interface was nicer… It was on a new level of quality that fulfilled consumer expectations. Not surprisingly, Windows 3.0 was the first commercially successful version of the operating system. It worked well (good functionality) and was relatively easy to use (good usability).

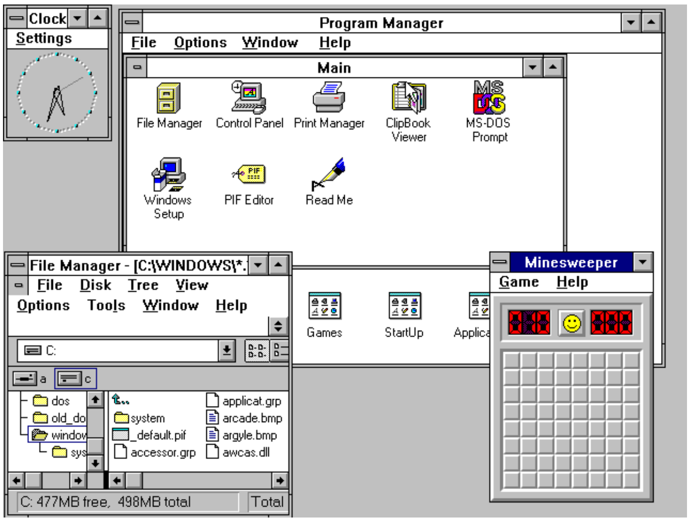


Figure 2.9. A typical desktop in Windows 3.1 (1992). User interface was much clearer and it had multitasking

Listening to users, Windows improved in subsequent versions. Windows 95 was a major change in user interface and functionality. It introduced the taskbar and the Start button, making it easier to launch applications and change among them. It also made possible the use of long file names, a much demanded utility (previous versions of Windows only accepted 8 character long file names). Although there were some minor releases in between, Windows XP was the next big landmark in the history of Windows. Again, there were improvements in performance (it was faster) and in the user interface (it was a better-looking version). Also, more functionality was added, like the possibility of fast changes between user sessions.

When Windows Vista was to appear in 2007, many people felt they had no need for a new version of Windows. Both functionality and usability levels had been correctly achieved, so they were happy using Windows XP. In fact, Windows Vista is not so different from Windows XP in terms of functionality (Figure 2.10). It seems that not many more features can be added, all daily work could be done without big difficulties with Windows XP. In such a mature product as Microsoft Windows, what else can be done to make it attractive to consumers?

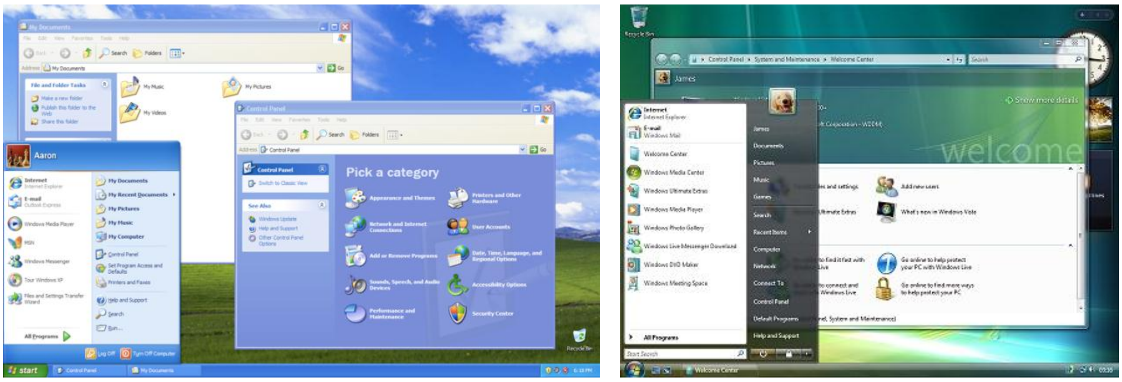


Figure 2.10. Left, desktop in Windows XP (2001). Right, desktop in Windows Vista (2007)

There were definitely some improvements in user interface and some added functionality in Windows Vista. But this version of Windows focused on the emotional aspects of using the operative system, in creating a "wow" experience (thus achieving the next level: attractiveness). Advertising for Windows Vista emphasized this idea of the great emotional experience of using the operating system (Figure 2.11). TV commercials for Windows Vista did not show people using computers, but watching TV with astonished expressions when man first landed on moon, reviewing the entire route after an exhausting but satisfying jogging session… All "wow" experiences!

Figure 2.11. The "Wow" campaign in the 2007 release of Windows Vista

Only when a product works well in the sense that it follows the standards, fulfills the necessities that users can express with words and is easy to use, does the importance of the emotional facet become more obvious. This is clear to Microsoft design engineers when they state as their mission: "We seek to create products that people will love" (Figure 2.12).

Figure 2.12. Emotional design as a key characteristic of products from Microsoft

The extended description of this mission reads[15]:

> *One of our greatest challenges is imbuing our products with "spirit" and "magic." This is easier said than done. Designers at Microsoft work side by side with the world's best software developers and usability experts to attain that elusive yet worthwhile goal. We take what can be a very complex problem and provide product design solutions based on a thorough understanding of the user's abilities, hopes, and expectations. Through our design, we strive to creative products of simplicity that you will also find intuitive to use. We aspire to creative beautiful experiences that not only look amazing, but also make our customers feel amazing about themselves.*
>
> *Like love, great design requires no explanation.*

If Windows Vista had this high quality (in all its levels), why is there a general perception that Vista was a failure and that many users maligned it? Probably because it does not run well in old computers as it requires a quite powerful system and because some security features are too intrusive and annoy users. From the discussion in Section 2.2 one can state that when a product is emotionally attractive, we are more prone to forgive its faults. But of course, this has a limit. If the previous levels of quality are broken, and the product does not work well, or not as we expect, the emotional track is lost. This also relates to price. Perhaps a product is absolutely

[15] This information about design at Microsoft was retrieved in August 2009 from www.microsoft.com/design.

awesome, but it can be a failure if it is too expensive (although people will pay more for it if it touches them emotionally…)

I have not talked about price in this discussion about Microsoft Windows. But the price element is quite particular in Windows, and probably in the whole software market. Price in Microsoft Windows operative system has always been reasonable. But its overwhelming international expansion has been favored by the widespread habit of software piracy in some countries, and by pre-installation in many computers when they are purchased.

To summarize, the first versions of Windows tried to simply work. As the operative system evolved, user necessities and expectations were more and more important, and functionality was added to fulfill them, together with a concern in keeping procedures easy for users (usability). The last step was adding this emotional touch, trying to create a "wow" experience. This path works well when each quality level is built upon the previous ones, without breaking them.

2.4. From emotional products to emotional companies, from the present to the future

We have seen up to this point the importance of incorporating emotions in the design of new products, and how emotional design can be considered attractive quality in many products, but will gradually become must-be quality, following the usual life cycle in the Kano model.

This dissertation basically focuses on products' appearance. However, other aspects beyond the scope of this work must be considered in order to be an "emotional company": not only products must be emotional, but also the relation the company has with its clients through branding, commercials and all other marketing issues must be aligned with its "emotional values".

On the other hand, thanks to information and communication technologies and flexible production schemes, customer co-creation and mass customization is now possible and will presumably become more popular in the near future. This gives the opportunity to personalize the emotional experience elicited by products (thus having to link design properties and emotional responses at an individual level).

The following years will also bring an increase in automation in current products, and the emergence of new robotic devices. Automation and robots evoke even higher emotional responses than more static products.

I will discuss all of these ideas with more detail in the following subsections.

2.4.1. Design and marketing: the winning combination

When trying to evoke certain perceptions from users, not only the design of the product, but also the marketing used to promote it plays an important role. Both design and marketing should be complimentary. Once the desired message to be conveyed is clear, both the design and the marketing must communicate this message. Designer Patrick Jordan states[16]:

> *If you have decided that the product should be, say, 'elegant, retro and masculine' make sure that this is reflected in both the design decisions and any marketing material. Any inconsistency between design and marketing sends a confused message to the consumer and gives the impression of a lack of integrity in the product.*

Both design and marketing can contribute to differentiation from competition and to strengthen the brand identity of a company. Looking again at iPod, Apple uses the same typography and aesthetics in its website, its commercials and in the product itself.

I think a good example of this connection between product design and marketing comes from the Swedish company Volvo, and specifically from the comparison of its cars and the commercials made to promote them 25 years ago and today.[17]

Figure 2.13. A 1990 Volvo 740 GTL, showing its squared appearance

Volvo Cars is an automobile manufacturer founded in 1927 in the city of Göteborg. In 1998 it was acquired by Ford. During many years Volvo cars had a reputation of solidity and reliability. Marketing strategies from Volvo stressed safety both in

---

[16] In an interview available at http://www.design-emotion.com/2006/12/04/getting-emotional-with-pat-jordan, retrieved in September 2009

[17] Simon Schütte offered a course on Kansei Engineering in Barcelona in April 2007, and he cited Volvo as an example of company "becoming emotional". This example is built on that idea.

magazine advertisements and in television commercials (TV commercials from the 1980s and 1990s often showed Volvos crashing, with minimal consequences). When safety was not the main topic, the advertisement focused on performance and durability. Many Volvo models were compared to tractors and were considered to be heavy. They earned the distinction of “bricks”, a visualization of its boxy aspect (see Figure 2.13).

Figure 2.14. The Volvo C30 release campaign, asking opinions about the car

But things started to change around the year 2000 with new Volvo models being more stylish. This alteration in style had its counterpart in Volvo advertisements. The release campaign for the C30, in December 2006, encouraged people to give their opinion (whether positive or negative) about the new car using a website full of video and images (Figure 2.14). People were asked to give their opinion based on their feelings, an idea that was summarized with the statement “Don’t read, just look”.

Figure 2.15. The Volvo XC60 release campaign, in 2009

Volvo has deepened this emotional approach to its models since the release of Volvo C30 in 2006. The strategy has proven successful with the S80 sedan, the XC90 sport wagon (Figure 2.15) and the C70 convertible. In fact, consumers reacted so positively to the advertisements that Volvo decided to make the campaign its mainstream advertising. Instead of cars crashing, the TV commercials now show vehicles at rest or peacefully cruising, with people staring at them or even touching them.

Two magazine advertisements from Volvo, 20 years apart, give a clear example of this shift in focus from features and performance to emotional experiences (Figure 2.16).

Figure 2.16. Two very different Volvo advertisements: from 1989 (on the left) and 2009 (on the right)

The old advertisement reads:

> *Unlike other cars, Volvo's record on safety is well documented. More than three decades ago, Volvo introduced a concept virtually unheard of in the American automotive industry. Safety. And year after year, regardless of trends, Volvo has continued to build cars that people feel secure driving.*
>
> *Recently, other manufacturers have begun to discuss safety when selling their cars. But who are you going to trust? A car company that's just started to promote safety, or the one that's defined it for years?*

Collate the statement with the text in the new advertisement:

> *There's more to life than a Volvo. There's running off for a weekend, with no phone reception. Running into an old friend and rolling back the years. Running into your*

ex and running right past. Running late for the gym (hurrah). Running late for tea and biscuits (boo). There's running all over town to find an anniversary gift, the day after your anniversary. And there's not running into the car ahead of you, in your XC60. That's why you drive one.

Of course, the classic Volvo "bricks" still persuade some people. Although these models were probably not designed with emotion in mind, they also evoke feelings (nostalgia, for example); different from the ones elicited by the new models. Their fan groups are on websites such as turbobricks.com or swedishbricks.net.

As Volvo cars try to be more viscerally attractive now than 20 years ago, the strategy used to market them has also changed. It seems that the current most successful products follow this integrated approach of aligned design and marketing. It also seems that, as Volvo has done, changing the focus from product performance and functionality to people's experiences with the product is a good strategy. In fact, what is more important, the product *per se* or people's lives with the product?

2.4.2. The emotional power of brands

Without doubts, the brand printed on a product has an effect on the perception the user has about that product. Brands have a real value (they can be financial assets), as the way brands are explicitly valued in mergers and acquisitions prove.

In 1975, Pepsi started a campaign in the United States called the Pepsi Challenge. It is a blind taste test of Coke and Pepsi. The test came to the conclusion that Pepsi was preferred over Coke. The question that arises is: why Coke outsells Pepsi almost everywhere in United States if Americans prefer Pepsi's taste? Neuroscientist Read Montague repeated the classic test in a magnetic resonance imaging machine in 2004 (Montague 2006). Sixty seven people had the blind taste test of Coke and Pepsi, and basically half of them chose Pepsi. But when the test was repeated knowing which brand they were drinking, three out of four people preferred Coke. The magnetic resonance unit showed activation in memory regions of the brain, not revealed in the blind test. The brand has a value in the brain, regardless of the content of the can.

Surely Pepsi is aware of the above results, and perhaps for this reason they try to change the emotional perception consumers have of their brand with their communication strategies. In 2009, Pepsi changed its logo and packaging simplifying it and converting its white strip into a kind of smile (Figure 2.17).

Figure 2.17. Old Pepsi logo, until the beginning of 2009 (on the left) and new Pepsi logo, from 2009 (on the right)

When one is presented with a set of products, not only the appearance but also the brand has an effect on the choice. Recall the four mobile phones shown in Figure 2.1. As we have seen, they are similar in functionality and the price is the same, so probably its visual or tactile appearance will determine our choice. But also our previous experiences with mobile phones from a specific brand and simply the reputation that brand has for us are important. Imagine we think Samsung is a good brand: this could bias our decision to prefer the mobile phone from this brand.

This influence brands have on users' perception of a product is produced by the mental ideas the user has about the values of this brand. But at the same time, this mental image of the brand is built after having experiences with products from that brand. So the brand image works at the top of the pyramid of necessities presented in Section 2.3. After systematically experiencing products from a specific brand that work well, are easy to use and attractive, we are more eager to welcome a new product from that brand. Of course, other issues affect our mental images of brands, such as social responsibility of corporations and the collective perception of the brand (easily discovered nowadays thanks to user review websites such as ciao[18])

Emotions elicited by brands is the topic of the recently coined *emotional branding* field of study. The effect of brands is not considered in this dissertation, although some of the techniques described could be adapted to decide on the best word for a new brand or the colors and shapes its logo should have. When presenting sets of products to be rated by users in a Kansei Engineering study, no brands should be visible to avoid confusion between the effect of properties in the product and the brand effect. This means not only hiding the brand, but also not using products that are easily identified with a brand even without showing the brand's name (for example, the classic Coca-Cola contour bottle).

---

[18] ciao has several websites adapted to different countries: Spain (www.ciao.es), United Kingdom (www.ciao.co.uk), Germany (www.ciao.de), France (www.ciao.fr), Italy (www.ciao.it), , Sweden (www.ciao.se) and The Netherlands (www.ciao-shopping.nl)

2.4.3. Incorporating customers in the production process

In 1987, author Stan Davis coined the term mass customization in his book "Future Perfect" (Davis 1987). The idea behind mass customization is using flexible computer-aided manufacturing systems to produce products personalized at an individual level. In this way, costs can be relatively low because of mass production, but flexibility is gained thanks to individual customization. These personal products can be design, to a certain level, by the customer, in a process called customer co-creation. The widespread use of internet makes customer co-creation feasible.

The sportswear company Nike combines customer co-creation and mass customization in NIKEiD (www.nikeid.com), an interactive website where users can personalize shoes changing materials and colors (Figure 2.18). It is possible to start from a proposed model or from a blank model. The shoe is updated in real time as changes are made, and shoes can be viewed from several angles.

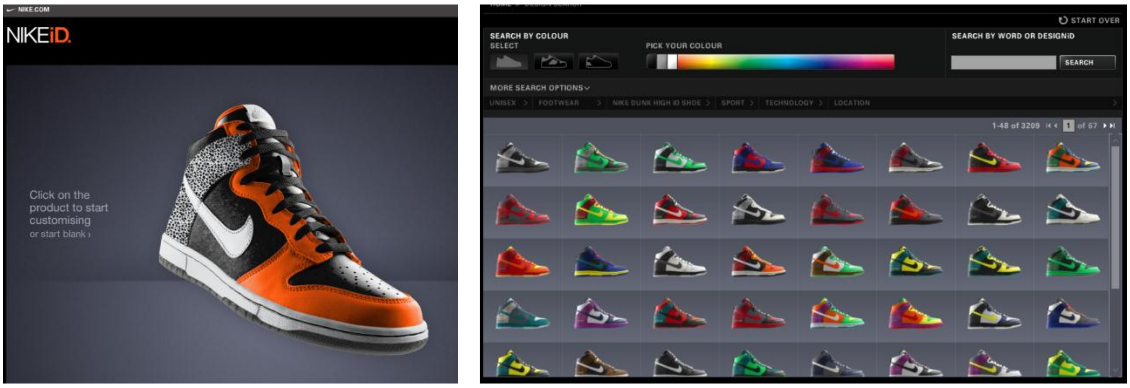


Figure 2.18. NIKEiD allows people to design their own shoe, either starting from a proposed model or beginning with a blank model

The technology corporation Dell also allows the customization of its computers from its website. The user chooses processor, memory, hard drive capacity, etc. But Dell has gone one step further. After having trouble and decreasing sales, in February 2007, Dell launched the website www.ideastorm.com. In this website, users can post suggestions and the community votes, so that the most popular ideas rise to the top. One of the first ideas was to sell computers with preinstalled Linux instead of Windows. This suggestion was followed by Dell for several models.

I think this tendency of having people more and more involved not only in the product, but also in the values of the company producing the product, will increase. The Internet revolution makes this involvement possible. Users are literally "manufacturing" some products, as with the collaborative web-based encyclopaedia Wikipedia[19].

In emotional design, and specifically in Kansei Engineering, we make the assumption that groups of people (our target group) will react positively to certain products. In fact, this assumption is not necessarily needed when talking about mass customization. Each person can have their personalized product, eliciting his or her desired emotions. Knowing about this person, and trying to discover how design elements relate to his or her personality and perceptions will be as important – if not more so – than in mass production of products. This knowledge will allow us to make personal recommendations and proposals, similar to the book recommendations already done by Amazon.

Multinational manufacturing companies headquartered in the western world are systematically moving their production plants to countries with lower production costs. This trend is also observed in services companies (call centers, software coding, etc). In fact, even research centers are beginning to be moved to these "cheaper" countries. I think the added value of personalization by considering consumers' emotional responses can be an antidote against industrial delocalization for western economies.

2.4.4. The present is emotional, the future will be even more!

The last decades have seen an emergence of new products based on technology, for example, walkmans in the 1980s and mobile phones in the 1990s. Nobody seemed to have a necessity for listening to music while walking or jogging, but walkmans in the beginning and mp3 players nowadays are a big success. The boom of mobile phones at the end of the last century was also quite unexpected, and common people did not seem to need them, but in many western countries today we can count more mobile phones than inhabitants. The importance of innovation becomes clear when we think about these essential products of today that were inconceivable just yesterday. For certain, the decades to come will bring many new products that are difficult to imagine now.

Following the pyramid structure of consumer needs in Figure 2.6, when a new product appears in the market, efforts are invested in endowing functionality to the product and assuring it works correctly. But soon usability issues demand attention, and also design to make it attractive. Although historically the pyramid has been ascended step by step (this has happened with "old" products, like cars), all three

[19] Incidentally, according to some studies, accuracy in Wikipedia is similar to that of more traditional encyclopaedias, like Encyclopædia Britannica (Giles 2005)

necessities (functionality, usability and attractiveness) are usually considered simultaneously in new products that aim to be successful.

The characteristics of many new products, with a high dependence on newly developed technologies, makes the concern about emotional design even more evident, if possible. Consider the example of the domestic vacuum cleaner Roomba, made and sold by iRobot (Figure 2.19).

Roomba is an autonomous robotic vacuum cleaner able to navigate a living space and its obstacles while vacuuming the floor. Roomba was first introduced in 2002, and several versions of the product have appeared since then. As of January 2008, over 2.5 million units have been sold[20]. While automation has been in factories for some time, I think Roomba is one of the first products that clearly introduced automation in our homes. This tendency will surely increase in the next years in all kind of products (cars are another good example of products with growing levels of automation).

Figure 2.19. The first generation of roomba robots, introduced in 2002.

As Donald Norman states in his book "The design of future things" (Norman 2007), automation needs huge communication between user and machine. We all have in mind the annoying experience of products that try to do everything by themselves, not allowing the user to intervene: when computers try to be too intelligent they can mess up everything. Great efforts still need to be made to polish how robots understand our human needs and act in response.

The presence of more and more robots in our lives, performing activities only done by humans until now, will increase the amount of emotional reaction we have when interacting with them. Ethnographic studies show that a robotic vacuum like Roomba generates significant changes in families using it, while a stick vacuum does not (Forlizzi 2007). When first used in a home, many families sit and watch Roomba clean for half an one to an hour. Small children and pets react with curiosity (Figure 2.19).

[20] According to irobot website (www.irobot.com), accessed August 2009.

Roomba changes daily life because not only does it maintain~~s~~ the domestic space cleaner, but it also urges family members to be tidier (for example, picking up small things from the floor to prevent a malfunction in Roomba). And what is more surprising, many people give names or nicknames to Roomba, assign it a gender (mainly thinking it is male) and impose personalities (reporting it is brave, clever, etc.). It is also common to feel guilty or worried about Roomba when it has to work a lot because of a very dirty room.

Figure 2.20. A cat (on the left) and a baby (on the right) "interact" with a Roomba robot

These human reactions with a robot are not exclusive to Roomba. It seems that people tend to interact with robots and real pets similarly. Children with the robotic pet AIBO from Sony treat it with respect and try not to hurt it, whereas this behavior is not observed when interacting with stuffed animals. Also soldiers develop an emotional relationship with bomb diffusing robots (which are usually given a name), expressing grief when lost in an explosion.

Probably this emotional relationship with robots comes from the appreciation we develop when a machine serves its intended purpose well. But its external appearance is also important. In any case, the years to come will see an increase in the number of robotic products designed to help us in our daily lives, and added automation to many existing devices. As machinery becomes more "similar" to human behavior, emotional reactions increase and must be taken into account when designed.

3 An Introduction to Kansei Engineering

This chapter locates Kansei Engineering within several techniques used in emotional design; a short story and a definition of Kansei Engineering are provided and a model for conducting KE studies is proposed. Finally, the topic of measuring perceptions in emotional design, in general, and in Kansei Engineering, in particular, is addressed.

3.1. A short summary of research methods used in product design

Designers usually rely on their intuition, creativity and experience for creating a new product. But they also use different qualitative and quantitative methods to collect information on how products are perceived and used. Almost all of these methods come from the field of user-centered design, a project approach that focuses on fulfilling the needs and requirements of users, and are well known by usability experts. Some popular user-centered design methods are[21]:

[21] Besides the book "Designing Pleasurable Products" by Patrick Jordan (2002), the articles from the website www.webcredible.co.uk/user-friendly-resources/web-usability (accessed October 2009) have been used as reference for this description of techniques.

- **Focus groups**: It is a technique originally developed (and commonly used) in market research. A group of potential or current users of a product are invited to share their thoughts, feelings and ideas on the product. There is a discussion leader and several participants (usually something between four and eight). The leader has an agenda of issues to be covered in the discussion. However, the agenda is not rigid, as topics that participants propose are probably those of more concern.
- **Participatory design**: In this technique, a group of participants are usually together with designers in order to discuss issues relating to a product's design. The main difference with focus groups is that participants are actively involved in the design in a "hands-on" way.
- **Interviews**: Here the investigator generally speaks to one participant at a time, posing a series of questions about the product. The participant's point of view can be explored in detail and any unclear statements can be identified and quickly addressed.
- **Experience diaries**: Notebooks are issued to users so that they can make notes of their experiences with the product over a period of time. Participants are asked, for example, to fill in a page of the diary every week, stating what they like and dislike, how they feel about the product, problems aroused when using it, etc. With experience diaries it is possible to monitor how the interaction with the product changes over time.
- **Usability testing**: Usability testing is, as its name indicates, basically appropriate for searching if a product is easy to use. It follows the first rule of usability according to usability guru Jacob Nielsen: "pay attention to what users do, not what they say"[22]. Normally, a person will be asked to perform a series of tasks while a moderator takes notes. The session is video recorded and users can be asked to follow the think-aloud protocol which asks them to verbalise what they are doing and why they are doing it.
- **Field observations**: It involves watching people in the environment in which they would normally experience a product. The investigator can observe how the user interacts with the product, the difficulties that the participant has and the apparent mood of the person when using it. It is important to ensure that the investigator's presence has a minimal influence on user's behavior.
- **Questionnaires**: These are lists of questions, normally with a limited choice of answers. Sometimes some open-ended questions are also added. It is quite common to indicate the level of agreement with a statement having to mark a box with labels "strongly disagree", "disagree", "not sure", "agree" and

[22] This quote by Jacob Nielsen is extracted from http://www.useit.com/alertbox/20010805.html, accessed April 2011.

"strongly agree", or some other similar rating. Statistical methods can be used to analyse the results.

Designer Christian Rohrer (2009) proposes two dimensions for characterizing user-centered design methods: qualitative-quantitative approach and behavioral-attitudinal data source (behavioral can be described as what people do whereas attitudinal is what people say). Figure 3.1 is an adaptation of this proposal. The graded filling in circles (not in the original proposal) gives information about the difficulty of conducting the technique in terms of time and resources needed.

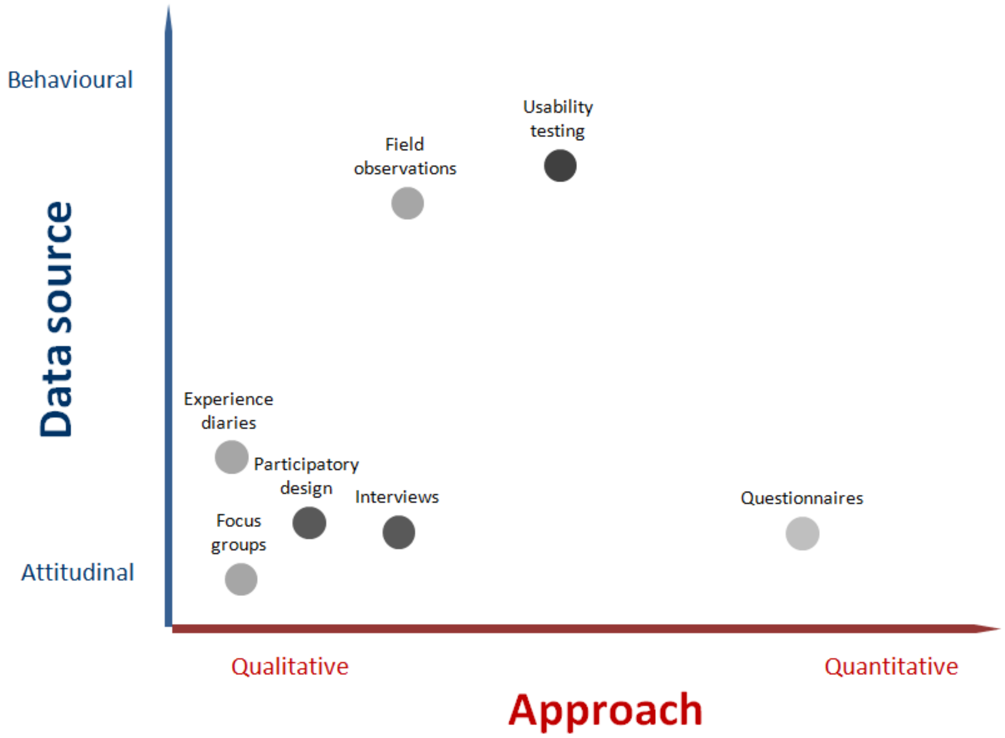


Figure 3.1. Several methods used in product design located in qualitative-quantitative and the behavioral-attitudinal dimensions, adapted from Rohrer (2009)

As portrayed by Figure 3.1, questionnaires' outcomes are quantitative. Usability testing can also sometimes give some quantitative data (for example, time needed to perform tasks). Field observations are less likely to provide quantitative data. I have located interviews not strictly on the left part of the approach axis as at times some kind of rating or ranking is performed (this is in fact the use of a questionnaire together with the interview). The other techniques are clearly qualitative.

Usability testing and field observations analyze what users do with the product. All other techniques are based on what people report.

Usability testing is costly because of the required equipment and the time needed. Although interviews do not need sophisticated devices to gather data, it is a very time-consuming procedure. Other techniques are easier. Questionnaires are especially straightforward, as once designed and prepared they can be given to many participants.

Many of these methods used in user-centered design can be adapted to emotional design. Patrick Jordan describes these methods in Chapter 4 of his book "Designing Pleasurable Products" (Jordan 2002, pp.136-204). However, as emotions elicited by products are difficult to measure directly, techniques such as usability testing (it should be called elicited emotions testing or something similar) and field observations cannot be applied easily. Thus emotional design relies basically on self-reported techniques, such as focus groups, interviews, experience diaries and questionnaires. Of course, it is arguable if this reporting of emotions using words is a rationalization of emotions, and this implies a loss of validity. However, no easy alternative is available nowadays[23].

The most commonly used techniques in emotional design are focus groups and interviews, in the qualitative approach, and questionnaires in the quantitative approach. From Chapter 4 to 8 of this dissertation I will focus the discussion on the quantitative approach and more specifically on Kansei Engineering (which uses questionnaires for collecting data). However, this does not mean the important role qualitative approaches play in emotional design is not recognised. Both quantitative and qualitative approaches are needed: I will delve deeper into this relationship in Section 8.5.

3.2. Kansei Engineering: the quantitative approach to emotional design

The last section included a review of common methods used in user-centered design, and a discussion on how these methods can be adapted to emotional design. In this section I will focus the work on the field of Kansei Engineering, a quantitative approach to emotional design.

In my view, there are three clear advantages in using quantitative approaches to discover the emotions elicited by products instead of qualitative approaches:

---

[23] In fact, some devices can be used to "directly" measure emotions, and I will discuss this in Section 3.3. For sure, the near future will bring new possibilities of doing this emotional measurement: I sketch these imagined possibilities in Section 9.4**Error! Reference source not found.**.

- Results from qualitative approaches depend a lot on the person leading the focus group or performing the interview. It is easy to talk too much and guide users' opinions, or talk too little and not discover interesting issues. Without doubt, the interviewer's experience has an effect and can bias output.
- Qualitative approaches, especially interviews, require a lot of time. So usually only a few interviews are performed, and conclusions are derived from asking a small amount of people. Moreover, one acts quite blindly when selecting people for the interviews (usually a target group of product consumers is defined, but there is no security that this group is homogeneous from an emotional point of view).
- It can be difficult to obtain product design guidelines due to the fact that users are not typically thinking in a designer's paradigm. In addition, some users say what they think they are supposed to say, or what they think the interviewer wants to hear.

As we will soon see, a quantitative approach like Kansei Engineering overcomes these difficulties (although it has some others).

3.2.1. Focusing the field study for this dissertation

One problem for the promotion of emotional design techniques is the lack of a common terminology. Simon Schütte (2005) lists several names used in this area of knowledge: affective design, affective engineering, pleasure with products, and some others. I use only two terms in this dissertation: emotional design and Kansei Engineering.

I designate as emotional design all methods used to incorporate emotions into product design (understanding emotions in the flexible way I explained in Section 2.1). All discussions in Chapter 2 related to emotional design, in this broad sense.

I place Kansei Engineering under the umbrella of emotional design (Figure 3.2). Kansei Engineering is, then, one of the available methods in emotional design (with connections and even overlapping with other methods). In my view, three characteristics make Kansei Engineering different from other emotional design methodologies:

- In Kansei Engineering, the aim is to connect physical properties of a product with the elicited emotions. Imagine we are studying watches. We want to discover, say, that if the face of a watch is rectangular, the watch is perceived as being elegant. Or that analogue displays are perceived as classical, whereas digital displays are perceived as modern. In this way, we not only get information about the specific watches used in the study, but also discover general rules valid even for watches not yet created.

- In Kansei Engineering, there is an attempt to describe the whole range of emotions a product can convey. We are not modelling a unique response (such as the elements that make people prefer a watch over the others), but several responses (such as the elements that provoke that people perceive a watch as being modern, and elegant, and reliable, and so on). There are as many responses as necessary concepts to cover the whole range of expected emotions.
- Kansei Engineering is based on collecting quantitative data (usually ratings made by users). Usually statistical methods or automatic learning techniques are used to link the physical properties to the elicited perceptions.

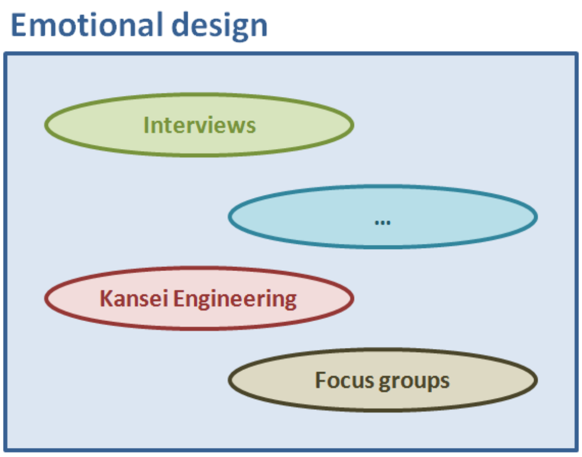


Figure 3.2. A schematic diagram that locates Kansei Engineering under the more generic field of emotional design

I am aware that Kansei Engineering is not understood in this precise way everywhere in available literature. The term Kansei Engineering is sometimes used referring to other tools that I place under the broader term emotional design. In some papers, it even refers to usability techniques such as field observations.

I think the three characteristics of Kansei Engineering stated above bring clarity to the topic and are operationally useful as they link well with the proposed model for Kansei Engineering studies explained in Section 3.2.4.

3.2.2. What is Kansei: etymology and idea

Many times, Japanese kanji signs are difficult to translate to other languages, as they are deeply connected with Japanese culture and world view. The Japanese term Kansei (感性) consists of two different kanji signs (Schütte 2005):

- 感 (Kan) can be translated as feeling, sensation or emotion.

- 性 (Sei) derivates from two other signs: 心 (translated as heart, mind, soul) and 生 (translated as be alive, dynamic).

All together, 感性 (Kansei) means sensitivity or sensibility, involving an active participation of the subject.

Based on a Kansei definition by Mitsuo Nagamachi, Simon Schütte (2005, p.36) proposes the following explanation of Kansei:

> *Kansei is an individual's subjective impression from a certain artifact, environment or situation using all the senses of sight, hearing, feeling, smell, taste and the sense of balance as well as recognition.*

Upon looking at the picture on the left in Figure 3.3, at first sight one would probably feel a sensation of peace, calmness and quietness. On the contrary, the picture on the right in Figure 3.3 possibly elicits a sensation of nervousness and stress. Although it is feasible to argue why each picture gives those impressions, we feel calm or nervous before thinking about the reasons, or even without thinking at all. These feelings are the elicited Kansei by the pictures.

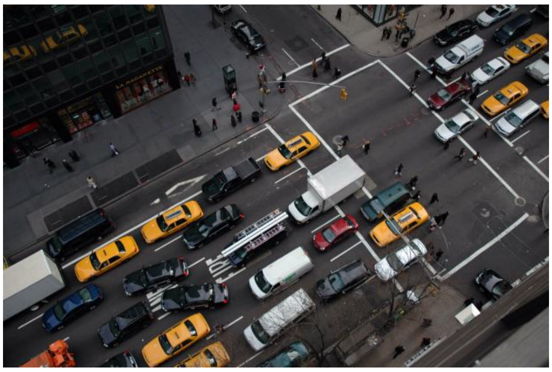



Figure 3.3. Landscape in Yellowknife, Canada (left) and street in New York (right)[24]

Keep in mind that the Kansei elicited by a situation or a product is a holistic experience, as it comprises all feelings experienced by an individual. In Kansei Engineering, we try to "capture the Kansei" conveyed by a product, and that means all the feelings the person experiences. But as Kansei Engineering is a quantitative methodology (data must be collected on a series of responses) it is necessarily reductionist. This reductionism of Kansei Engineering is unavoidable, but at least practitioners must be aware of this fact.

[24] Photo on the left taken from http://www.sxc.hu/photo/1223460; photo on the right taken from http://www.sxc.hu/photo/348661 (both accessed December 2009).

Although the Kansei elicited is personal (different persons could have different feelings when looking at the same picture), Kansei Engineering assumes that a large group of people will share the same Kansei on a product[25].

The Kansei elicited by a product can depend a lot on the context where the product is exhibited (Álvarez 2009). This fact has to be taken into account when collecting data for Kansei Engineering studies[26].

3.2.3. What is Kansei Engineering: history and types

In the 1970's, visionary and pioneering researcher Mitsuo Nagamachi (Figure 3.4) developed a methodology for translating consumers' feelings into design elements. He had a background in psychology and medicine and was working at that time in Hiroshima University's Faculty of Engineering (Childs et al. 2003).

Figure 3.4. Professor Mitsuo Nagamachi, founder of Kansei Engineering[27]

The term Kansei Engineering was first used in 1986 by Kenichi Yamamoto, then the president of Mazda Motor (Schütte 2005). Professor Nagamachi soon adopted this term. In the 1980's and the 1990's he created systems – which usually implied the use of computers for collecting data and statistical methods for analyzing it – for capturing users' Kansei and connecting it with product properties. Professor Nagamachi and his team collaborated with companies from many sectors during those years: automotive (Mazda, Nissan), apparel (Wacoal, Goldwin), electronic home products (Sanyo, Sharp), office machines (Fuji, Canon), and cosmetics (Shiseido), among others (Childs et al. 2003). In 1995, a paper with Mitsuo Nagamachi as sole author was published in the journal "Applied Ergonomics". I

[25] Chapter 8 of this work delves deeper in the homogeneity of the elicited Kansei among people, and how to segment people according to this Kansei.
[26] More will be discussed about this fact in Section 4.5
[27] This photo was taken in the Second European Kansei Engineering Conference, developed under the umbrella of the QMOD Conference in Helsingborg, in August 2008.

consider this paper – "Kansei Engineering - a New Ergonomic Consumer-Oriented Technology for Product Development" (Nagamachi 1995) – as a seminal paper where Professor Nagamachi presents his proposal to the scientific community in the world. In this paper, he states that Kansei Engineering was founded 30 years ago. But surely the ideas he explains in the article were developed during years of practical experimentation with the procedure. Besides explaining the idea behind Kansei Engineering, its utility for industrial design is justified and two case studies from Mazda and Shiseido are presented.

A boom in the interest of the topic of incorporating emotions into the design of products was produced during the last years of the 20th century and the years just after 2000. The publication of the book "Emotional Design" in 2004 by Donald Norman (2004), a respected guru in the field of industrial design, and the tremendous success of emotional products such as Apple's iPod (first launched in 2001), might have reinforced this interest.

Kansei Engineering has increased its growth and spread during the last years. Different groups in Western research centers and universities have started projects in this area. In 2002, researcher Simon Schütte from Linköpings Universitet, with a background in engineering, proposed a model for developing Kansei Engineering studies in his licentiate thesis. The model was further developed after a stay in Japan working with founder Mitsuo Nagamachi and his colleagues (professors Ishihara and Nishino among others). The model, together with a description of the idea behind Kansei Engineering, was explained in a 2004 paper co-authored by Simon Schütte, Jörgen Eklund and Jan Axelsson from Linköpings Universitet, together with founder Mitsuo Nagamachi (Schütte et al. 2004). In my opinion, this model systematizes the procedure and is a milestone in the effort of translating the ideas of Professor Nagamachi to the Western World.

Kansei Engineering is nowadays an exciting area of research with plenty of opportunities for practical application. Clearly a multidisciplinary field, scholars with different backgrounds can provide ideas and implement useful methods:

- Engineers, designers, experts in ergonomics, usability and human-computer interaction, due to its deep links with industrial design.
- Neurologists, psychologists and experts in life sciences in general, because it deals with emotions and how we process them.
- Sociologists and economists, as emotional products have an impact in society both at a personal and social level.
- Statisticians, due to the need for collecting and analyzing data for formulating conclusions.

Professor Nagamachi opened an enormous area of research with his brilliant and innovative ideas some decades ago! As Kansei Engineering started to be applied more and more often, Mitsuo Nagamachi proposed a classification for different types of studies. The classification has been updated as new types were introduced. Table 3.1 shows a summary of all these types.

Table 3.1. Classification of Kansei Engineering studies, adapted from Schütte et al. (2004)

Type I Category classification	Using a laddering procedure, a general Kansei is divided into several subconcepts, and finally each atomic concept is related with a physical property (in a similar fashion to QFD). No statistical methods are used.
Type II Kansei Engineering System	Computer databases are generated containing Kansei words (the words describing perceptions) and statistical methods are used to connect Kansei words with physical properties of products.
Type III Hybrid Kansei Engineering System	Similar to Type II, but besides going forward (from Kansei to properties design), it allows going backward (from properties design to Kansei)
Type IV Kansei Engineering	A mathematical model is built to predict the emotional response of consumers.
Type V Virtual Kansei Engineering	Virtual reality is integrated with Kansei Engineering for presenting the prototypes to evaluate.
Type VI Collaborative Kansei Engineering designing	Group work is used to conduct Kansei Engineering studies with software and databases over the internet.

I think the above classification might be a little weird, and I find it somewhat difficult to understand. Probably, the classification is done having a wider perspective of what Kansei Engineering is, and identifying what for me is something specific – Kansei Engineering – with the broader field of emotional design (Section 3.2.1). However, in almost all types the general characteristics of Kansei Engineering listed in Section 3.2.1 are present: the aim of connecting the Kansei with physical properties, the attempt of covering the whole range of possible elicited emotions and the use of quantitative data (only KE type I can perhaps be done without collecting data).

From now on, I will link Kansei Engineering with studies done following the model explained in the next section. This model could be located under type IV of the above classification (although type II and type III could also apply to some parts of the model).

3.2.4. A proposed model for Kansei Engineering studies

As I stated above, I consider the paper "Concepts, methods and tools in Kansei Engineering" (Schütte et al. 2004), published in 2004 in the journal "Theoretical Issues in Ergonomics Science” to be a key paper in the effort to translate the ideas and procedures of Kansei Engineering to the Western world. This paper describes the steps needed for conducting a Kansei Engineering study. This proposed model systematizes the phases that comprise a KE study, so it brings clarity to the procedure. Moreover, this organization facilitates research in specific areas of the method (as the procedure is then divided into clear phases).

The proposal, as described in the paper, is the following:

> *The basic idea is to describe – based on an earlier chosen domain – the idea behind the product from two different perspectives:*
> *(1) The semantic description; and*
> *(2) The description of product properties.*
> *These two descriptions span a vector space each. Subsequently, these spaces are merged with each other in the synthesis phase, indicating which of the product properties evokes which semantic impact.*

The steps in this original model are shown in the flowchart in Figure 3.5. A brief description of the steps is offered here:

Choosing the product domain: The product domain is defined. This means not only choosing the product that will be the protagonist of our study, but also the target group for the product.

For example, the product domain could be analogue watches to be used by middle aged women.

Spanning the semantic space: Words that emotionally describe the product (called Kansei words) are collected from different sources. The initial set of words is reduced either using affinity diagrams or multivariate techniques such as cluster analysis. The output from this step is a list with all the Kansei words that will be used in the study.

Examples of Kansei words for the watches could be modern, elegant, comfortable...

Spanning the space of product properties: Design attributes from the studied product are collected. For each attribute, several possible values are considered. Attributes that designers think can have an effect on the emotional response are prioritized.

Attributes in the example of the watches could be face color (white or brown) or face shape (rectangular or round).

A set of products (either real working products or prototypes) is prepared to be shown to participants in the study. The output from this step is the collection of products for the KE study (these products must differ in the properties just selected).

In the example of the watches, four watches could be used: a rectangular watch with a white face, a rectangular watch with a brown face, a round watch with a white face and a round watch with a brown face.

Synthesis: Using statistical methods or automatic learning techniques, a link between product attributes (space of properties) and Kansei words (semantic space) is established. For every Kansei word, product properties are found that affect the Kansei word.

In the watches example, one conclusion could be that watches with a rectangular face are perceived as elegant.

Test of validity: Factor analysis is used to locate the Kansei words on the first principal components. This gives an idea of which words are perceived similarly, and can be the basis for some confirmatory experiments.

Model building: After being validated, several final models are proposed that relate the product properties with each Kansei word.

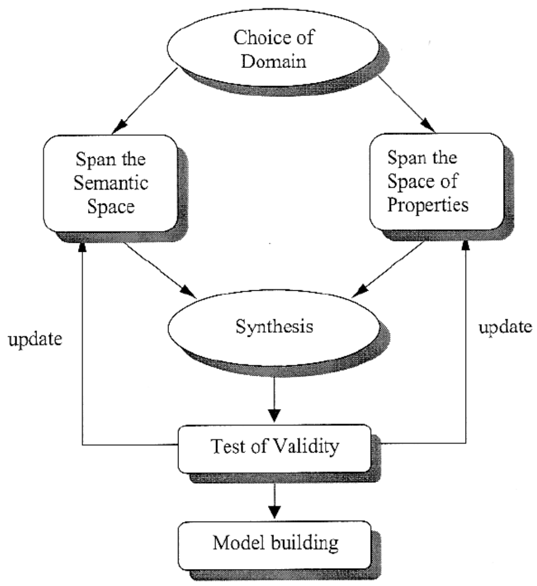


Figure 3.5. The original model for conducting Kansei Engineering studies, directly reproduced from the paper "Concepts, methods and tools in Kansei Engineering" (Schütte 2005)

Based on this model for conducting KE studies, a new proposal is made that incorporates two differences from the original model:

- A new phase is added just before the synthesis phase: "data collection".

- The last step “model building” is renamed “presentation of results”.

First modification of the model: addition of a “data collection” phase

Of course, at some point in a KE study, participants are asked to rate each product on all the Kansei words[28]. For example, one participant might give a rating of 6 (from an available scale from 1 to 7) to the rectangular watch with a white face for the Kansei word elegant.

In the original model, this data collection is the last step in the space of properties phase or the first step in the synthesis phase. As both products (selected in the space of properties phase) and Kansei words (selected in the semantic space) are used for the data collection, it is preferable not to include data collection in the spanning of the space of properties phase.

Instead of having data collection as the first step in the synthesis phase, I think it is better to have it as a distinct phase to emphasize two key facts:

- A Kansei Engineering study needs data, and before data can be analyzed in the synthesis step, it must be collected. The quality of this data is important to achieve reliable results.
- Statistics have a role not only in the synthesis phase, when data is analyzed, but also in the data collection phase. In fact, the way data is collected can determine the analyses that can be done later.

The existence of a data collection phase also provides a framework for discussing some topics central to Kansei Engineering: amount of participants in the study, how participants must be selected, where and how the data collection is performed, motivation of participants, etc.

Second modification of the model: renaming the “model building” phase

When the synthesis phase is finished, several models relating each response – that is, each Kansei word – with product properties are obtained. These models are thus basically built at the end of the synthesis phase. Validating the models – for example with some confirmatory experiments – is very reasonable, and the test of validity phase is devoted to this.

Once the models are validated, the last task is “publicizing” the models: thus the proposal of renaming this last phase to presentation of results. In fact, this renaming

[28] As we will see, rating Kansei words is not the only possibility for collecting data in a KE study (although it is by far the most frequent one). Ranking products is an alternative.

also has the purpose of pondering the importance of communicating the results from a Kansei Engineering study in a way that is easy to understand and interpret by non-statisticians.

The presentation of results phase must provide tools for showing conclusions from the KE study in a visually appealing way, basically using graphs that everybody can understand. Chapter 8 of this dissertation includes proposals for this phase.

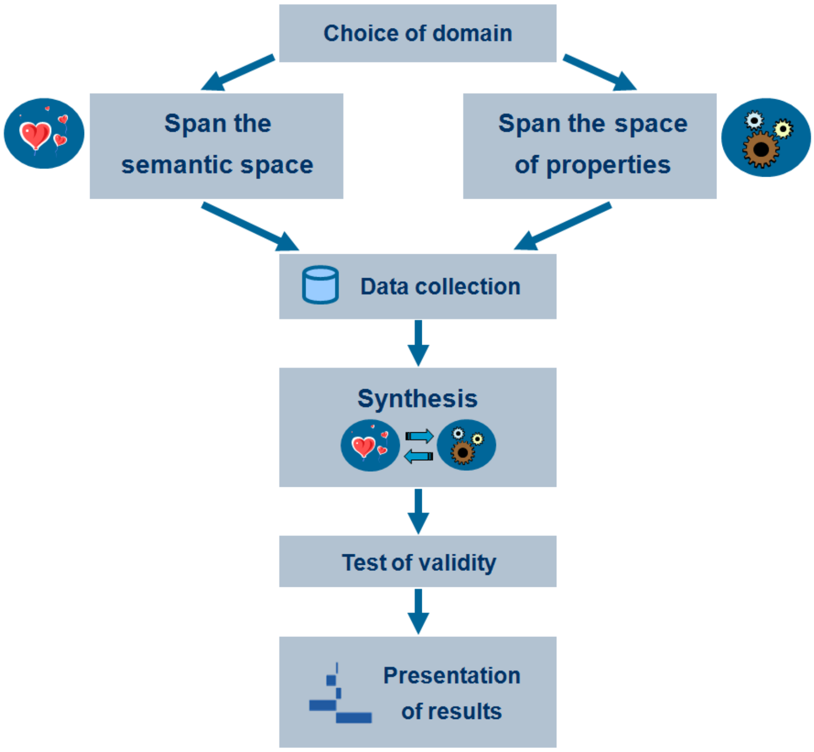


Figure 3.6. A modified model for developing Kansei Engineering studies, adapted from the original proposal by Simon Schütte (2005)

The final model for conducting Kansei Engineering studies that will be used in this work, and that incorporates both modifications of the original model, is shown in Figure 3.6. All the phases in a Kansei Engineering study briefly outlined above will be developed in detail in Chapter 4 using an example.

3.3. Methods for measuring the Kansei

As Kansei Engineering is a quantitative technique, it relies on numerical data. Many times collecting data to describe a phenomenon is difficult. However, data collection is crucial, as data is our raw material in a quantitative technique: if data is incorrect, conclusions will be erroneous.

In social sciences, the terms validity and reliability are used as requisites for a good measurement system. Validity is related with really measuring what we want to measure and having the closest approximation to the truth. Reliability is associated with the consistency of the measurement (the degree to which an instrument gives the same measurement every time it is used to measure the same)[29]. Thinking in statistical terms, validity is having unbiased estimations of reality, and reliability is having consistent estimators. The analogy of a dartboard is sometimes used to illustrate these ideas (Figure 3.7).

Kansei Engineering deals with emotions elicited by products. And measuring emotions seems like a difficult task! But of course we want to have measurements as valid and reliable as possible. How can this be achieved?

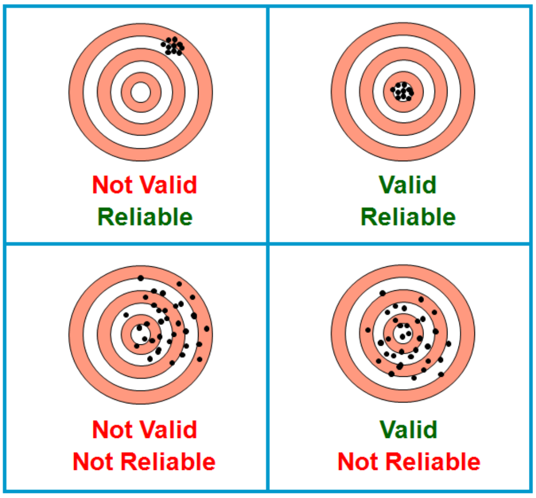


Figure 3.7. The dartboard example for illustrating the need of a valid and reliable measurement system

A practical classification of emotions can be useful at this point. A distinction can be made between two types of emotions depending on how much cognitive processing is required before the emotion is constituted (Poels, Dewitte 2006)[30]:

- Lower-order emotions are emotions that occur automatically[31]. They are spontaneous and uncontrollable reactions.
- Higher-order emotions depend on deeper cognitive processing of the situation.

Some basic emotions, such as fear or anger, are located between these two extremes. For example, if we are standing in front of a lion, we feel fear coming from a lower-

---

[29] Section 8.4 of this dissertation is devoted to reliability in Kansei Engineering studies.

[30] Besides the stated references, information for this section has been also taken from http://www.socialresearchmethods.net/kb/relandval.php (accessed November 2009).

[31] These lower-order emotions are the only ones that could be labeled as emotions according to Antonio Damasio's definitions we saw in Section 2.1.

order emotion. But fear can also be felt if there is a general economic meltdown and a person develops fear of becoming unemployed because has to pay a mortgage and feed two children. This second kind of fear is felt after cognitive reflections on the situation. Figure 3.8 summarizes this distinction.

There is currently no way to "measure an emotion" in such a direct, valid and reliable way as we measure our weight on a scale. Two approaches exist that allow an indirect measurement of emotions:

1. Capturing physiological body reactions when we experience particular emotions. For example, measuring heart rate, skin conductance, eye pupil dilation, face expressions, face temperature, stroke volume or blood pressure changes.
2. Trusting self-reporting of emotions. For example, asking a person to describe with words what he or she feels or to rate his or her degree of excitement on a scale from 1 to 100.

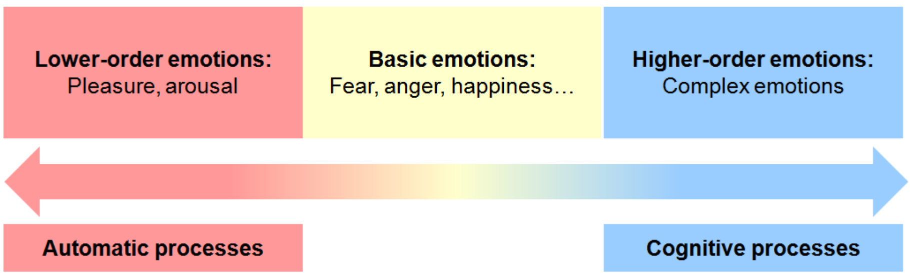


Figure 3.8. A practical distinction between lower-order and higher-order emotions, adapted from Poels & Dewitte (2006)

3.3.1. Capturing the Kansei through physiological body reactions

Capturing the Kansei through physiological body reactions has the advantage that unconscious reactions are measured, so in theory validity is guaranteed and lying is more difficult. Lower-order emotions can be better captured with this kind of measurement. But there are several problems:

- The necessary equipment is expensive. And even if it is not that costly, it usually needs dedicated installations where it can be used[32]. This is the case of cameras for recording interviews: camcorders are relatively cheap nowadays, but a proper use requires complex infrastructure.

[32] These kinds of installations are sometimes called living labs.

- The required equipment is very often too intrusive. When dealing with human beings, the best measurements are obtained when people are not aware that they are being measured (otherwise they are proven to change their behavior even unconsciously). Having a camera two centimetres from one of your eyes in order to detect your pupil dilating is so unnatural that perhaps your pupil decides not to dilate in that uncomfortable situation.
- It is possible to measure that an emotion fires, but there is no way to know which emotion or to distinguish between emotions. For example, if we measure the electrical conductance of the skin, we can detect an activation of the autonomic nervous system due to an increase in sweat secretion. However, we cannot know what produced this activation, neither if it was due to a positive or a negative feeling.
- It is not clear that reactions to slight changes in product properties are large enough to be detected with current equipment. Some studies reveal changes in relatively easy to measure magnitudes such as heart rate when exposing a person to extreme images (peaceful landscape, nudity, intense violence...) (Sakata et al. 2007). But these reactions are not always triggered with small changes in product parameters (such as showing a watch with a round face instead of a rectangular face).

Although measuring emotions using physical bodily reactions is a promising area of research it still has a limited applicability, in spite that some procedures, like eye tracking, are already being used successfully.

Figure 3.9. Eye tracking heatmap of a website[33]

[33] Screenshot by Michael Sauers, in flickr.com. Available from http://www.flickr.com/photos/travelinlibrarian/193383382, accessed November 2009.

Eye tracking is a methodology especially used in website design. In eye tracking, the point of gaze is captured, so it is possible to obtain heat maps that show how much users looked at different parts of a web page (Figure 3.9). Areas where users looked the most are colored red; the yellow areas indicate fewer fixations, followed by the least-viewed green areas.

3.3.2. Self-reporting the Kansei

Self-report measures have been extensively used for measuring emotions. Self-report measures register the respondent's subjective feeling by rating, ranking or choosing words or images. The validity of the measurement as a means of capturing the Kansei is somewhat doubtful, as emotions captured in this way are necessarily "rationalized" before giving the rate.

It is clear that self-report measures are not adequate to quantify lower-order emotions, but they are good enough for higher-order emotions, where cognition also plays a role. In an attempt to capture not "too rationalized" emotions, participants are asked to answer fast, giving their "first impression", and without revising previous answers.

The good thing is that all the problems listed above for physiological reaction measurements have now been transformed into advantages:

- No sophisticated equipment is needed. Although computers are often used to guide the data collection, paper forms can also be used and can sometimes be more appropriate.
- Ratings can be done in a relaxed atmosphere, and no unnatural devices are required during the measurement. In fact, filling out a form either on paper or on a screen is quite a familiar task nowadays for everybody.
- As emotions are described either with words or images, it is possible to distinguish among different emotions.
- It is easier to capture differences in perception due to changes in product properties, because participants are aware of the scale range and unconsciously try to establish differences among products shown.

There are two types of self-report methods for measuring subjective feelings: verbal self-report and visual self-report.

In verbal self-report, participants in the study are asked to express their emotions verbally by answering open-ended questions or by rating several words expressing emotions using semantic differential or Likert scales[34]. In visual self-report, the rating is done not on words, but on cartoon-like figures representing different emotions. Although rating is the most common procedure, it is also possible to rank products on a word, or choose among several products on a word.

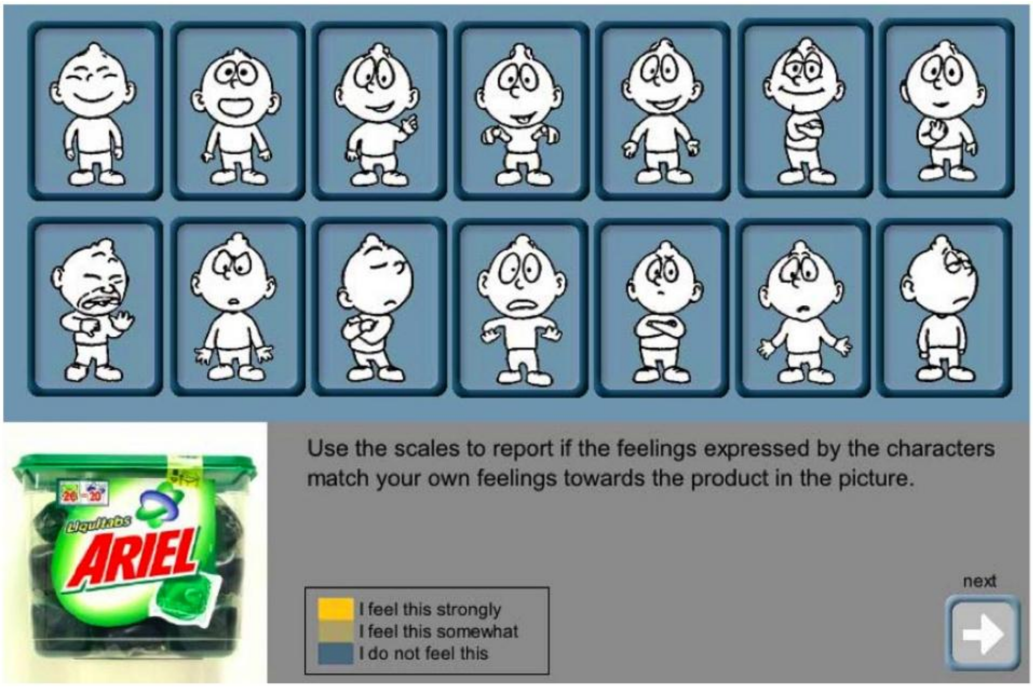


Figure 3.10. A screenshot from a PrEmo study

An example of a visual self-report instrument is PrEmo, developed by Pieter Desmet (2002). PrEmo includes 14 animations of 1 or 2 seconds, with sound. Each animation represents an emotion. There are 7 positive emotions (desire, pleasant surprise, inspiration, amusement, admiration, satisfaction and fascination) and 7 negative emotions (indignation, contempt, disgust, unpleasant surprise, dissatisfaction, disappointment and boredom). A product is shown in the lower left corner of the screen. Participants have to indicate how strongly they feel each emotion rating each puppet in a three point scale (I feel this strongly, I feel this somewhat, I do not feel this). Figure 3.10 shows a screenshot of PrEmo.

Using a visual tool for self-reporting emotions has the advantage that it can be used cross-culturally as nothing has to be verbalized (although some cartoons expressing emotions could be interpreted differently depending on the culture). On the contrary, in verbal self-reporting of emotions, the problem of language arises. Some words can

[34] Section 3.4.1 will explain the differences between semantic differential and Likert scales for rating kansei words.

have an unclear meaning to some people and if the study has to be done in different countries, it can be hard (if not impossible) to find exact translations for all words.

Figure 3.11. An example of ratings by a person on two watches for 3 Kansei words

However, verbal self-report is very popular due to its familiarity and versatility. Rating statements is common nowadays due to surveys, so rating words expressing emotions looks familiar at first sight. The greatest advantage is that any word of interest can be used (with no need to debate if it is an emotion or not).

Consider the example shown in Figure 3.11. Two watches get ratings (on a scale from 1 to 7) for 3 different words that describe perceptions elicited by the watches: luxurious, comfortable and masculine. Figure 3.11 has the rating of one person, another participant in the study would almost surely give different ratings.

Forms for collecting data in KE studies are similar to the verbal self-report form in Figure 3.11. Each word (luxurious, comfortable, masculine) is a Kansei word. Each participant rates each product for each Kansei word on a scale. Although rating is the most common method of verbal self-report, some others exist. Also different versions of the scales for rating can be used. This is the topic of the next section.

3.4. Collecting data in verbal self-reporting of the Kansei

Figure 3.12 summarizes the methods for measuring the Kansei seen in Section 3.3. All Kansei Engineering studies I am aware of use a verbal self-report form to measure the Kansei. This is also the instrument that will be used for Kansei Engineering studies in this dissertation[35]. From now on, I will focus the discussion on measuring the Kansei through verbal self-report (circled in the Figure). Data collected in this way makes possible a statistical treatment, so several improvements for KE studies will be proposed through the use of statistical methods.

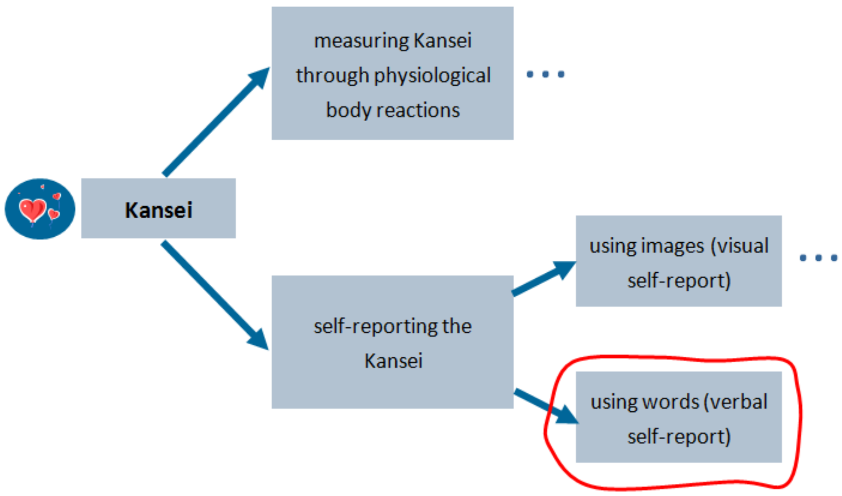


Figure 3.12. A summary of methods for measuring the Kansei

Kansei Engineering, and especially the spanning the space of properties phase, has a link with the discipline of design of experiments. Consider the example illustrated in Table 3.2.

Table 3.2 contains a 2^2 factorial design. There are two factors (color and printing), and each factor has two levels (color can be white or red; printing can be a picture or a text). Each T-shirt acts as a run in a factorial design experiment. Many people will be asked how colorful they consider each one of the T-shirts. The colorfulness is the response of the experiment. Each person giving a value to each T-shirt on the response colorful is a replicate. If 10 people participate in the KE study, each run in the experiment has 10 replicates.

---

[35] However, data from visual self-report instruments like PrEmo could also be used as raw material for many analyses explained in this work. On the contrary, data from physiological bodily reactions would need a different treatment.

The purpose of the KE study is the same as in an industrial factorial design: discovering how properties in the products affect the response (in the case of our example, discovering how the color and the printing affect the perception of colorfulness). In a KE study, however, there is no one response, but several responses: as many as Kansei words used.

Table 3.2. An example of 2^2 factorial design with T-shirts[36]

	Original values			Coded values		As dummy variables			
	Color	Printing		Color	Printing	Color: white	Color: red	Printing: picture	Printing: text
1	White	Picture		-1	-1	1	0	1	0
2	Red	Picture		1	-1	0	1	1	0
3	White	Text		-1	1	1	0	0	1
4	Red	Text		1	1	0	1	0	1

Another important difference between KE studies and factorial designs in industry is that in KE studies the response comes from assessments made by people, while responses in industrial environments usually come from measuring devices. Furthermore, responses coming from these measuring devices are almost always continuous (e.g. density, tensile strength, yield, etc.).

There are basically three different ways of collecting data when the response is assessments made by people, as in KE studies (keep in mind the example of evaluating T-shirts on the Kansei word colorful):

1. Rating the products: each product (T-shirt) is shown, one after the other, and each person has to rate to which degree he or she thinks the T-shirt is colorful. The rating is done, say, on a 5-point scale, where 1 means “the T-shirt is not colorful at all” and 5 means “the T-shirt is very colorful”.
   The procedure is easy to carry out, can be done with or without a computer, but can be a bit tedious for the participants in the study.
2. Ranking the products: all 4 products are shown, and each participant ranks the products. First T-shirt in the ranking is the least colorful, and the last T-shirt is the most colorful. One participant could give this ranking: T-shirt 1, T-

[36] The T-shirts shown are screenshots from a website that allows the creation of customized T-shirts. Unfortunately, I lost the reference of this website.

shirt 3, T-shirt 2 and T-shirt 4, and another participant could give this other ranking: T-shirt 3, T-shirt 1, T-shirt 4 and T-shirt 2.

The procedure can also be done with or without computer. It has the advantage that all stimuli are considered at the same time (not one after the other). The huge problem is that if there are many stimuli, the ranking can be an overwhelming task.

3. Choosing products: products are presented in pairs (or other groupings, such as in threes or fours, but probably not more, as it would be difficult to compare). For example, T-shirts 1 and 4 are presented, and the participant has to choose which one he or she perceives as more colorful. After this selection is done, another pair of T-shirts are presented, and so on.

   It is almost compulsory to use computers when data is collected in this way.

In terms of presentation of stimuli, procedures 1 (rating words) and 2 (ranking products) are similar. When rating words, all products from the design matrix (all four T-shirts in the example) have to be shown. As the order in which products are shown can have an effect on the results, products are presented in a random order to minimize this possibility. When ranking words, all products are presented at the same time (all four T-shirts are shown at the same time, and then ranked by each participant). Of course, analysis techniques differ in the rating and ranking procedures, but no further decision has to be made regarding the presentation of stimuli.

However, the situation in procedure 3 (choosing products) is different. From the four available T-shirts, only two will be presented at the same time. T-shirts in the example can be grouped into six different forms: 1-2, 1-3, 1-4, 2-3, 2-4 and 3-4. We do not want a bigger experimental effort than when rating or ranking: as our factorial experiment is a 2^2 and requires 4 runs, we only want to present 4 pairs, and not all the 6 possible pairs. Which ones do we choose? In which order? Is it possible to use the information from the previous choices to decide on the next pair of products to be presented? Not only will the analysis techniques in this procedure be different from procedures 1 (rating) and 2 (ranking), but some kind of algorithm must also be used to decide on the presentation of stimuli.

These three procedures described above (rating, ranking and choosing) are typical in conjoint analysis. Conjoint analysis is a technique used in market research to determine how different features in products affect people's preference (Hair Jr et al. 1995, pp.556-615). So preference is usually the response in a conjoint analysis study. Each stimulus is presented in a card, describing the characteristics of that stimulus. Figure 3.13 shows an example of cards used in a conjoint analysis study about computers. There are four variables (factors) being studied: computer's brand, processor, RAM memory and hard drive capacity. The question of interest is discovering which of these variables have an effect on the preference people have among these products. Is price the most important variable, no matter the features the computer has? Or are all variables important, except for the brand?

The purpose of conjoint analysis in market research is analogous to that of factorial designs in industrial environments. Conjoint analysis and KE studies share a lot in terms of possibilities for measuring the response, preparation of the design matrix and procedure to present the stimuli. The only difference is that, whereas conjoint analysis usually only has one response (the preference), in KE studies we have as many responses as Kansei words.

Brand: Dell **Processor**: Intel Celeron Dual Core **Memory**: 2 GB **Hard drive**: 160 GB **Price**: 450 €	**Brand**: Dell **Processor**: Intel Celeron Dual Core **Memory**: 4 GB **Hard drive**: 250 GB **Price**: 520 €	**Brand**: HP **Processor**: Intel Core 2 Duo **Memory**: 4 GB **Hard drive**: 150 GB **Price**: 540 €	**Brand**: Acer **Processor**: Intel Core 2 Duo **Memory**: 2 GB **Hard drive**: 400 GB **Price**: 400 €

Figure 3.13. An example of cards for a conjoint analysis on computers

In terms of time and effort needed, rating words is clearly very efficient, and this is especially important in KE, as sometimes there is a huge amount of Kansei words to be rated. The wide literature review on Kansei Engineering presented in Chapter 5 of this dissertation shows that almost all KE studies use the procedure of rating each Kansei word on a scale. So, from now on, I will assume data for a KE study is collected by rating Kansei words (so forms for collecting data will be similar to that shown in Figure 3.11).

However, I believe that exploring these different ways of collecting data for KE studies is a promising research area. Procedure 3 (choosing products), is a conjoint analysis version (called choice-based conjoint) that is especially interesting. It has the advantage of being more natural and closer to the real choice decision process done in the market.

Several topics can be discussed when talking about rating Kansei words:

1. The use of semantic differential (SD) scales or Likert scales
2. The use of ordinal scales or visual analogue (VA) scales.

Figure 3.14 completes the previous diagram in Figure 3.12 and circles these two topics that will be developed in the next subsections.

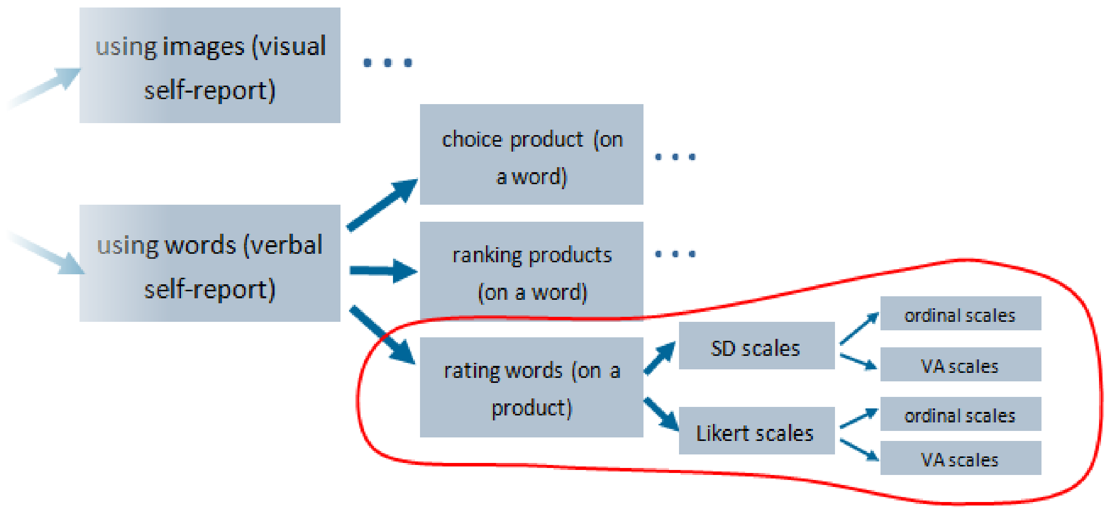


Figure 3.14. Diagram of methods for verbal self-report, focusing on rating words

3.4.1. Semantic differential scales and Likert scales

The semantic differential method was first proposed by psychologist Charles E. Osgood in his book "The measurement of meaning" (Osgood, Suci & Tannenbaum 1957). The aim of the method is to measure the emotional content of a word objectively. This method uses bipolar scales with two words, one on the left and another one on the right, to collect quantitative data. The two words on each side of the scale are antonyms. Figure 3.15 shows an example of a semantic differential scale.

Osgood used multivariate techniques on large collections of semantic differential scales and found that the pairs of words could be summarized in three principal components; he named these three components evaluation, potency and activity. Osgood called the space defined by these three components the semantic space. Evaluation captures the concept of good-bad, potency the concept of strong-weak and activity the concept active-passive. When evaluating a person (the first area where semantic differential method was used) it makes sense to think first if the person represents a danger (is good or bad?), if the person is able to do something (is strong or weak?), and if the person actually has the purpose of doing it (is active or passive?)[37].

[37] The entry on semantic differential method in Wikipedia gives nice concise information and references about the method (http://en.wikipedia.org/wiki/Semantic_differential, accessed October 2009)

This watch is...

	1	2	3	4	5	6	7	
classical	☐	☐	☐	☐	☒	☐	☐	**modern**

Figure 3.15. An example of a semantic differential (SD) scale

Semantic differential (SD) scales have been used extensively for measuring constructs in psychology. But also Likert scales are very popular. Likert scales are named after Rensis Likert, an American psychologist that developed the scale in the 1930s[38]. Basically, it consists of a statement that the respondent is asked to evaluate according to any kind of subjective or objective criterion; generally the level of agreement or disagreement is measured. Figure 3.16 shows an example of a Likert scale.

This watch is modern

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
☐	☐	☐	☒	☐

Figure 3.16. An example of a Likert scale

Both SD scales and Likert scales are subject to different types of bias. Two of them are worth mentioning here:

- Central tendency bias: respondents avoid using extreme response categories. This tendency depends on the person (some people are more “extremist” than others) and perhaps also on the cultural environment (some cultures are more eager to express opinions, while others prefer to remain in a polite “central area”).
  Both SD scales and Likert scales are prone to central tendency bias. A solution to this bias is not having too few response categories (I will delve further into this issue in the next subsection).
- Acquiescence bias (sometimes referred to as "yah-saying"): respondents have a tendency to agree with all the questions.
  An example: imagine a person answers “strongly agree” in the Likert scale shown in Figure 3.16 with the statement “This watch is modern”. If we now ask the statement “This watch is classical” on the same scale, the expected

[38] According to Wikipedia (http://en.wikipedia.org/wiki/Rensis_Likert, accessed October 2009), Rensis Likert developed the 1-5 Likert Scales for his thesis, and eventually became Likert's best-known work. I am quite certain that nothing in this thesis about Kansei Engineering will have such an impact in science!

answer would be “strongly disagree”. However, the real answer could perfectly be “Neither agree nor disagree”.

The problem of acquiescence is not easy to manage. It is more pronounced with Likert scales containing only positively worded items than with Likert scales first asking about the synonym and later about the antonym.

It has been reported that SD scales are better at dealing with the acquiescence bias than Likert scales (Friborg, Martinussen & Rosenvinge 2006). The problem with using SD scales in Kansei Engineering is that a Kansei word and its antonym must be used, and sometimes it is not clear which antonym to use. The election of one can, in fact, add some other kind of response distortion.

Simon Schütte (2005) explains that Japanese researchers like Mitsuo Nagamachi tend to use a SD scale, but always adding “not” to form the antonym (Figure 3.17). Doing so, the problem of choosing antonyms is avoided.

This watch is...

	1	2	3	4	5	6	7	
not modern	☐	☐	☐	☐	☒	☐	☐	**modern**

Figure 3.17. An example of scale as used by Mitsuo Nagamachi

Simon Schütte (2005) proposes a modification of a Likert scale, placing the statement (in this case, the Kansei word) on top in the middle of the scale, and labeling the anchors as “not at all” and “very much” (Figure 3.18).

This watch is modern

	1	2	3	4	5	6	7	
not at all	☐	☐	☐	☐	☒	☐	☐	**very much**

Figure 3.18. An example of scale as proposed by Simon Schütte

Both proposals are a bit tricky, as they look like an SD scale, but in fact are not. I would add a variation of the last proposed scale to the list of possibilities – the one in Figure 3.19. As the left side always means “not at all” and the right side always means “very much”, perhaps these two labels can be erased in each question, and only give the indication of what the anchors mean in the beginning of the survey.

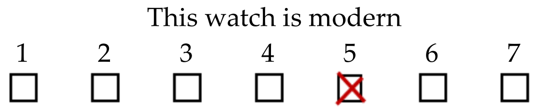


Figure 3.19. A last example of a proposed scale

I am not aware of studies comparing the performance of these last three scales proposed for KE studies (Figure 3.17, Figure 3.18 and Figure 3.19). But I would take the risk of saying that they are quite similar.

3.4.2. Ordinal scales and visual analogue scales

When rating statements (or Kansei words, in our case), both ordinal scales and visual analogue scales can be used:

- Ordinal scales are n-point scales, n can be 3, 4, 5, 6, 7, 9, 11; in fact, any small integer. In Kansei Engineering, 7-point and 5-point scales are the most common. Figure 3.20 (top) shows an example of a 7-point ordinal scale.
  Ordinal scales are easy to administer, either on a computer or on a piece of paper. If an odd number of response categories are used, respondents can mark the central category and thus express neutrality. This is not possible in ordinal scales with an even number of categories.
- In visual analogue scales (VAS), respondents mark a position along a continuous line between two end-points. An example is given in Figure 3.20 (bottom).
  The response is the length of the line from the left (e.g. if the line is 10 cm long, a rule can be used to measure the distance from the left end in millimetres). This manual procedure makes the annotation of responses quite a burdensome procedure if the study is administered on paper (it is easy if using a computer).
  Some computer software for collecting data in KE studies (such as KeSoft[39]) uses VAS.

[39] KESoft (Kansei Engineering Software) is a software developed by the Kansei Engineering group in Linköpings Universitet for collecting and analyzing data in KE studies. I used this program in a course on Kansei Engineering led by Simon Schütte in Barcelona in April 2007.

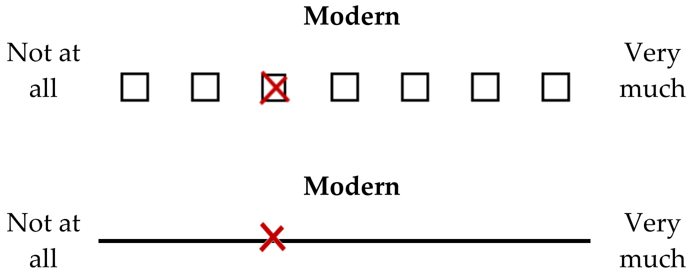


Figure 3.20. An example of a 7-point ordinal scale (top) and a VA scale (bottom)

As scales are always used in surveys, there is a plethora of papers on this topic. Some papers compare ordinal scales and VAS (Grant et al. 1999), others compare 5-point, 7-point, 9-point, 10-point and 11-point ordinal scales (Dawes 2008, Lawless, Popper & Kroll 2010). Sometimes, specific results of papers are contradictory (e.g. one says variance is slightly higher in 7-point scales than in 5-point scales, whereas another one detects no difference). But they all agree on the following main conclusions:

- Reliability and validity is good in scales having at least 5 response categories (3-point scales allow too little differentiation).
- Results are basically the same regardless the number of response categories used, when moving from 5-point scales to 9-point scales.
- Scales with more than 9 points do not seem to be needed, as they do not add more discernibility.
- Visual analogue scales and ordinal scales provide similar results.

Of course, data from ordinal scales is discrete, and this fact has some consequences in the statistical techniques available for analyzing it. As we will see later in this work, data from all participants in KE studies is almost always averaged. Therefore, data for the analysis can be considered continuous thanks to the central limit theorem.

Sometimes, the use of VAS is justified arguing that data coming from VAS scales is directly continuous, so more appropriate for subsequent analysis. I think this is a bit tricky. Consider the example in Figure 3.21. The first rating is 8.4, whereas the second one is 7.7. Can we really consider that the first respondent perceives the shown product as more modern than the second one? Rounding the figures to the integer, and considering that both respondents gave an 8 on a scale from 0 to 10, seems quite reasonable.

So I think that VAS gives the illusion of having continuous data, but I find data from ratings conceptually closer to ordinal data than to continuous data. From now on, I will suppose data in a Kansei Engineering study comes from an ordinal scale. If a VAS is used, data can be treated as ordinal simply discretizing it.

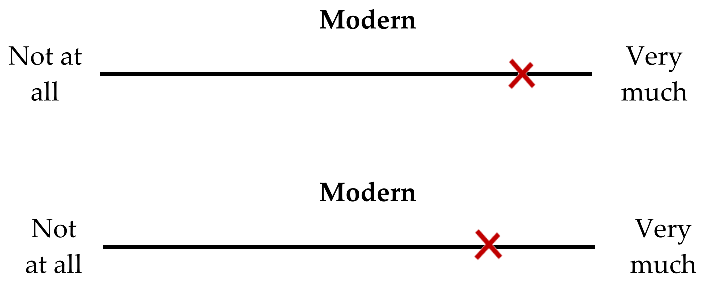


Figure 3.21. Two “different” ratings on a visual analogue scale

I set aside here another issue of research: the effect of labeling a scale with words instead of numbers. The Likert scale in Figure 3.16, for example, had labels from “strongly disagree” to “strongly agree”, instead of numbers from 1 to 5. This labeling of the scale can also be done in an ordinal scale or in a VAS. When labeling a VAS it is even possible not to locate the labels equidistant from each other. I will suppose that, if labels exist, they can be translated into numbers.

4 The Model for Kansei Engineering Studies

This chapter explains, step by step, the proposed model for conducting Kansei Engineering studies. The model provides a framework for further discussions in this dissertation. An experiment with fruit juices is used as an example to illustrate the model and motivate the next chapters in this work.

4.1. The juices experiment: motivation and description

Many challenges in science (if not all) have an origin in real life problems. As I progressed in my research on Kansei Engineering, I realized it was important to conduct a real Kansei Engineering study to detect difficulties and opportunities for deeper research. So in April 2007 I prepared, together with Xavier, my supervisor, and Ana and Eli, two students of mine with a background in statistics, a KE study. Our interest was in the methodological and academic parts of the study, more than in the actual results that we were to obtain. After a brainstorming session, we decided to conduct the study on how the visual presentation of a fruit juice evokes certain sensations[40].

---

[40] Several other options were considered. Our initial idea was conducting the experiment with cocktails. We later decided that fruit juices were easier to manage; they can be purchased

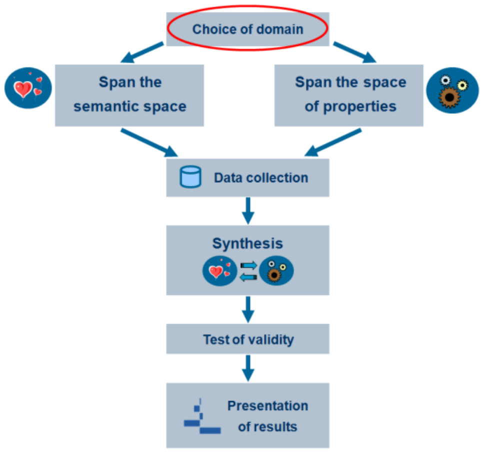
This chapter will explain in detail the model for conducting KE studies first presented in Section 3.2.4 and schematically summarized in Figure 3.6 (reproduced on page 48 of this dissertation). The model is easier to understand when explained together with an example. Next sections will thus go through the different phases in the model, illustrating the techniques used with data from the juice experiment.

Moreover, there are two additional reasons for presenting this case here:

1. Although many examples of KE studies can be found in literature, they usually describe the outcomes, but give limited details on the analysis performed and even fewer details on how data is collected. The juice experiment is a simple example that can be useful for people willing to understand how to conduct a KE study.
2. Several difficulties and challenges in KE will become apparent in the example. These difficulties serve as a motivation for the proposals in following chapters of this work. The juice experiment will appear further on in several chapters (where data will be reanalyzed), and will give continuity to this dissertation.

This application of KE served as the diploma thesis of Ana and Eli, and it was the first complete KE study that I conducted (Peralta Coll, Gómez Muñoz 2007). Besides being employed in this work, data from the juice experiment has already been used in a conference presentation (Marco-Almagro, Tort-Martorell Llabrés & García Subirats 2007) and in another thesis (Álvarez 2009).

4.2. Choice of domain.



The choice of domain is the first step in the proposed model for conducting KE studies. Choosing the domain obviously includes deciding which product is the protagonist of the study. But this is not the only task. In addition to deciding on the product, the choice of domain also requires:

- Defining the target group to which the product is addressed (Schütte et al. 2004)

---

already prepared in supermarkets, are cheaper than alcoholic beverages, and almost everybody has had one at some time in their lives.

Participants in the study have to match the characteristics of the target group (so they are basically a sample from the whole population represented by the target group). Usually, a fairly homogeneous group of people (from the point of view of personal and socioeconomic characteristics) is chosen. In theory, this fact guarantees that ratings done by participants are quite similar. In practice, this is not always the case.

- Defining the kind of presentation for the product.

  There are basically two possibilities: showing real physical products (either prototypes or fully functional products) or showing representations of the product (2D representations, like photographs, or 3D representations, perhaps using a virtual reality environment).

  The number of sensory organs involved in processing the presentation of products has an important effect in the perceived Kansei, and is called the "affective channel width" (Picard 2000). As an example, consider a study done with perfume bottles. The affective channel is wider when presenting a real perfume bottle (so it can be not only seen, but also touched and smelled) than if the perfume bottle is only shown in a photograph. This is what Simon Schütte denominates "proximity of presentation" (Schütte 2005, pp. 67-68).

  The other three aspects affecting the way the Kansei is transferred are: prior experience with the product, interest in the product and allowed degree of interaction. This phenomenon can be described by the term "proximity of interaction" (Eklund, Kiviloog 2003).

- Defining the context for presentation.

  The atmosphere of the place where the experiment is conducted can have an effect on the emotions elicited by the product. The same product can convey different emotions depending on the environment. René Álvarez uses the Japanese term *gemba* for designing the context in which products are presented – an analogy for the use of *gemba* in quality management for designing the manufacturing floor (Álvarez 2009). *Gemba* is so important that he suggests that when measuring the Kansei elicited by a product, we are in fact measuring the Kansei elicited by the whole experience of interacting with the product in that situation.

These three points included in the choice of domain phase besides selecting the actual product (target group, kind of product presentation and context of presentation) receive little attention in many KE studies described in the literature. In this dissertation, the issue of selecting a sample group of people and the consequences of how this is done will be addressed in Chapter 8. Section 4.5 will delve deeper into the topic of how and where the products are presented.

In our experiment, the chosen product for the KE study are fruit juices, and specifically its presentation just before being drunk[41].

Our plan was to present photographs of fruit juices to some people, and to gather the ratings on several Kansei words. One could argue that only showing a photograph gives a narrow affective channel, as no smell or taste is involved. As we were only interested in the visual impression, we thought this was not a big issue. Furthermore, just presenting photographs makes the study more robust to potential effects derived by the context of presentation. At any rate, what was important for us at this point was getting data, more than the results of the study.

As we were dealing with an academic study, and results were to be used only for research in KE and not as practical input for a real product design, we wanted to have quite a big group of participants (say, something over 20), regardless of their age, gender or interest in fruit juices. So we asked friends and family of the three of us leading the experiment – Ana, Eli and myself – to participate. Twenty four people kindly volunteered. The resulting group was rather heterogeneous: there were 13 women and 11 men, with ages ranging from 17 to 59, and different educational levels (from high school to post-graduate studies).

As we have just seen, it is usually not recommended (and not common) to have such a heterogeneous group. Participants in "real life" KE studies done for companies are usually more similar because they are selected from the target customer group. Besides, many KE studies conducted in an academic scenario at universities use students as participants. These students always have very similar ages and interests.

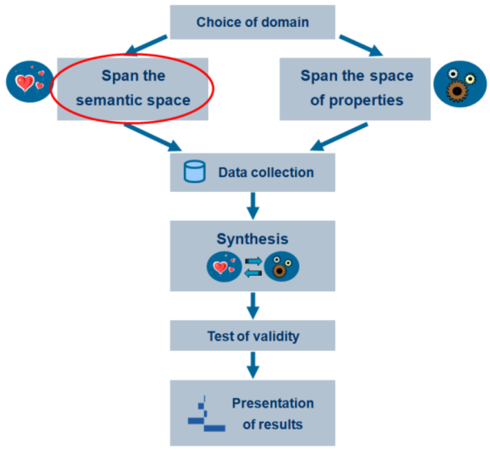
I will explain more about differences among participants in a KE study in Chapter 8. At the moment, it should be noted that our group for the fruit juice experiment was not particularly homogenous and that the participants were all the people we were able to persuade to participate. Nobody got any economic compensation for their collaboration (although we offered them a small snack).

4.3. Spanning the semantic space

We saw in Section 3.3 that Kansei Engineering uses a verbal self-report measurement system for getting data. Participants in the study rate words that express emotions that can be conveyed by a product. These words are called Kansei words in the framework of KE studies.

---

[41] Every time data from the fruit juice experiment is analyzed, this small picture will appear on the left of the page. The picture is by Michael Lorenzo, and is taken from http://www.sxc.hu/photo/1032249 (accessed April 2009)



Explained briefly and simplistically, spanning the semantic space means choosing which Kansei words are going to be used in the study. Usually, Kansei words are adjectives, as the ones shown in Figure 3.11 in page 54 (luxurious, comfortable, masculine). However, verbs or nouns can also be used (Schütte et al. 2004). Sometimes, even sentences can be used as Kansei words if they better describe the desired ideas[42].

Spanning the semantic space comprises three steps:

1. Setting an initial list of Kansei words.
2. Reducing the initial list of Kansei words using qualitative or quantitative methods (or a combination of both).
3. Proposing the final reduced list of Kansei words (the ones that will be used in the data collection phase).

4.3.1. Initial collection of Kansei Words.

As I stated in Section 3.2.1, one characteristic of Kansei Engineering is that there is an attempt to describe the whole range of emotions a product can convey. This endeavor can only be successful if the semantic space really covers all possible emotions elicited by the product. This is why the first step in spanning the semantic space is preparing a long list of Kansei words using all available sources: the longer the better.

These Kansei words can be taken from magazines, manuals, catalogues, websites, users, experts or related KE studies. If possible, ideas (perhaps even extravagant ones) must also be translated into Kansei words in an effort to use KE as a creative product development tool (Schütte et al. 2004). The number of Kansei words on this initial list can be very large, varying from several tens to several hundreds. This first word collection must continue until no new words appear in order to make sure that the complete semantic space is considered. Some studies call this initial list of Kansei words the initial semantic space (Such Pérez 2003).

[42] In a KE study done by the author, sentences like "The product is for athletes" or "The product helps you recover from an effort" were used as Kansei words, as they were the best option for expressing the desired ideas.

A first screening of all Kansei words can be used to remove some of them. For example, words related with materials (e.g. metallic) or specialized terms (e.g., art deco like) can be directly removed from the initial list (Jindo, Hirasago & Nagamachi 1995).

An aside on selecting the words for the initial semantic space

The Kansei words configuring the initial semantic space are, in fact, dependent on the chosen scale for the data collection phase. If a "real" semantic differential scale is chosen, a word and its antonym are needed (e.g. classical-modern). In this situation, perhaps the better procedure is listing the words in pairs from the beginning.

Likert scales (or some modifications of Likert scales to make them "look" as semantic differential scales, compare Section 3.4.1) are more used in KE studies, as they are easier to manage. In this case, one possibility is to directly eliminate from the initial list words that are clear antonyms from another. For example, if both bitter and sweet are in the initial list, only one of them is needed (if a juice is considered bitter, it will get a low rate on the word sweet, so the construct is correctly captured with just one word).

Anyway, the words that are directly excluded depend on the product domain and the criteria of the expert conducting the study. Words must be removed with caution, as the final selection of Kansei words should cover the whole semantic space.

The collection of Kansei words for our juice experiment was done by Eli and Ana searching basically for adjectives and verbs related with fruit juices in books (Millidge 2006, Wilson 2003, Murphy 2005) and websites[43]. The 134 words shown in Table 4.1 were proposed[44]. These words constitute the initial semantic space.

Eleven words were removed from the very beginning, as they were considered more related with physical properties of the juices than with evoked sensations (decorated, with alcohol), too general words (nice, ugly) or simply odd words difficult to relate with juices (deep, ideal). It was decided not to remove antonyms, as it was often difficult to agree on which pairs had to be taken out.

[43] The following websites (all accessed in December 2006) were visited:
http://www.alimentacion-sana.com.ar/informaciones/novedades/frutoterapia.htm,
http://www.enbuenasmanos.com/articulos/muestra.asp?art=1579,
http://www.recetasdecocina.es/listado-recetas-33-Zumos.html

[44] Words were collected in Catalan, as the data collection phase was done in this language. Table 4.1 offers each word translated from Catalan to English. In the data collection phase and in subsequent phases, I will always use the English translation.

Table 4.1. Original list of Kansei words in Catalan and English.

Catalan	English	Catalan	English	Catalan	English
àcid	acidic	excel·lent	excellent	refrescant	refreshing
divertit	amusing	excitant	exciting	regenerador	regenerative
antiestresant	anti-stressing	exòtic	exotic	enfortidor	reinforcing
antifatiga	anti-fatigue	explosiu	explosive	relaxant	relaxing
antioxidant	antioxidant	exquisit	exquisite	remineralitzant	remineralising
aphrodisiac	aphrodisiac	fabulós	fabulous	renovador	renovative
atraient	appealing	fashion	fashionable	reconstituent	restorative
aperitiu	appetizer	festiu	festive	restaurador	restorer
aromatic	aromatic	fibrós	fibrous	revitalitzant	revitalizing
artificial	artificial	vistós	flamboyant	ric en ferro	rich in iron
astringent	astringent	floral	flowery	romàntic	romantic
atractiu	attractive	espumós	foamy	saludable	salutary
dolent	bad	fresc	fresh	saciant	satiable
equilibrat	balanced	futurista	futuristic	saborós	savory
beneficiós	benefical	golós	gluttonous	seductor	seductive
amarg	bitter	bo	good	sensual	sensual
tonificant	bracing	gratificant	gratifying	sedós	silky
genial	brilliant	sa	healthy	senzill	simple
calmant	calming	casolà	home-made	aprimant	slimming
caribeny	caribbean	ideal	ideal	suau	soft
nadalenc	christmas spirit	infantil	infantile	sofisticat	sophisticated
clàssic	classical	intens	intense	agre	sour
fred	cold	vigoritzant	invigorating	picant	spicy
colorit	colorful	irresistible	irresistible	estimulant	stimulating
combinable	combinable	sucós	juicy	reforçant	strengthening
còmode	comfortable	juvenil	youthful	fort	strong
concentrat	concentrated	laxant	laxative	substitutiu	substitutive
consistent	consistent	lleuger	light	sabor subtil	subtle taste
corpulent	corpulent[45]	luxós	luxurious	ensucrat	sugary
cremós	creamy	madur	mature	estiuenc	summery
curatiu	curative	embafador	mawkish	dolç	sweet
refinat	dainty	mediocre	mediocre	silvestre	sylvan
decorat	decorated	hidratant	moisturizing	gustós	tasty
profund	deep	natural	natural	temptador	tempting
delicat	delicate	bonic	nice	tropical	tropical
deliciós	delicious	nutritiu	nutritional	lleig	ugly
espès	dense	apassionat	passionate	vellutat	velvety
desintoxicant	detoxifying	agradable	pleasant	versàtil	versatile
digestiu	digestive	popular	popular	vibrant	vibrant
diürètic	diuretic	potent	powerful	vital	vital
diví	divine	preventiu	preventative	vitamínic	vitamin-rich
sec	dry	proteic	protein-rich	càlid	warm
fàcil de digerir	easy to digest	pur	pure	salvatge	wild
energetic	energetic	depuratiu	purifying	amb alcohol	with alcohol
erotic	erotic	fi	refined		

The initial semantic space was finally constituted by 123 Kansei words.

[45] Means having a large bulky body

4.3.2. Reducing the initial list of Kansei words.

Each Kansei word will play the role of a response in the experiment. Therefore, each participant must rate each Kansei word for every product in the study. As the initial semantic space often has more than a hundred of words, asking participants to rate each product on every word from the initial semantic space is totally unfeasible[46]. So there must be a selection of Kansei words using data reduction methods. Any reduction implies a loss of information, but the idea is removing words that express very similar emotions and are thus essentially redundant. The final list of Kansei words obtained is sometimes called reduced semantic space (Such Pérez 2003).

Three approaches are suitable for this reduction of Kansei words: a qualitative approach, a quantitative approach, or a combination of both.

A qualitative approach for reducing the initial semantic space

Kansei words with similar meanings can be grouped together using an affinity diagram. For each group, one of the Kansei words will be chosen as representative of that group. As the number of words in the initial semantic space can be huge, this grouping can be done in several steps. Imagine we have 300 words in the beginning. An affinity diagram can collapse these words into 40 groups, and a word is selected representing each of these 40 groups. In a second affinity diagram, the 40 words are grouped again and 16 final groups are created, thus producing a final set of 16 words.

Doing the affinity diagram as a one-person task is not appropriate. But in fact not many people are required: probably only those responsible for the KE study. No quantitative data is collected, so there is no need to further analyze the results. The set of representative words of each group from the last affinity diagram conform the reduced semantic space.

A quantitative approach for reducing the initial semantic space

Here, a small collection of data is needed and groups of words are created analyzing this data using multivariate methods. The common procedure is choosing a small set of products, and asking some people to rate the Kansei words using the same scale planned for the data collection phase. Participants in this initial rating can be the

---

[46] I would say a motivated person answering questions in a survey format could stay a maximum of 20 or 30 minutes more or less concentrated. From this point, it is quite sure he or she will lose attention (or even worse, he or she will become fed up and start inventing the answers). Probably this time limit depends on cultural issues (and it is different, say, in Europe and in Asia). Section 5.5.3 has information on time used for the data collection of several KE studies from the literature review.

technicians preparing the KE study or other people (perhaps people that do not strictly belong to the target group, as our only purpose now is to detect similarities among words, and not yet to connect product properties with emotions).

Products for this small rating procedure do not need to be used later in the data collection phase. The whole set of products used in this phase must be quite different from each other, as we are interested in capturing all Kansei words' variations. Imagine our product of interest is watches. If every time a watch is rated high on the word luxurious it is also rated high on the word modern, and on the contrary, every time a watch is rated low on the word luxurious it is also rated low on the word modern, we can infer that Kansei words luxurious and modern are perceived as similar in this domain. However, this is only true if the following two conditions are fulfilled:

- The set of products used for the ratings is not extremely small. If the ratings are done with just one watch, people could perceive that watch as being luxurious and modern at the same time, but it is not possible to infer from this that Kansei words luxurious and modern are equivalent for every watch.
- The set of products is quite different from each other. Imagine that all watches used for the rating have golden faces and rectangular shapes, and that golden faces give the perception of being luxurious and rectangular shapes give the perception of modernity. In this situation, Kansei words luxurious and modern will have similar ratings, and thus will be detected as similar, but this is because all watches had at the same time golden faces and rectangular shapes. Perhaps with more different watches these two words would obtain different profiles of ratings.

Once we have the data, it is summarized calculating averages among all raters, so we finally have a number for each product and Kansei word. Kansei words can then be grouped using two methods:

- A cluster analysis. With this method, several groups of Kansei words are established. Words that are close to other words (in the sense that they have a similar profile of ratings through all products, and thus are perceived as similar) will be part of the same group. I think this method is suitable if the initial number of Kansei words is not extremely large. The advantage with this method is that it is relatively easy to choose one word as representative for each group. The problem is that the number of groups needed is not obvious.
- A principal component analysis. With this method, the principal components, which are linear combinations of the original Kansei words, but orthogonal

among them, are computed. The idea then is to only keep the first principal components, and label them looking at the original Kansei words that contribute more in each principal component. I think this method is useful when the initial number of Kansei words is very large. The advantage here is that common criteria such as keeping the principal components that have an eigenvalue higher than 1 (and thus capture more variability than any of the original Kansei words) can be used in order to decide the number of Kansei words finally considered. The huge problem is that labeling each principal component is sometimes quite difficult (and falling into overinterpreting is easy).

The third way: qualitative + quantitative approach

There is a mixed third way to conduct the reduction of the initial semantic space. It consists in a combination of the qualitative and quantitative approaches explained above. First, an affinity diagram can be used to group Kansei words. This procedure can be repeated more than once. Later, a cluster analysis or a principal component analysis can be conducted to decide on the final Kansei words. If using a non-hierarchical clustering procedure, such as k-means clustering, the groups obtained from the affinity diagram can be used as initial groups for the clustering procedure, thus refining the reduction with collected data.

This last method is the one we used in the juice experience. Remember that the initial semantic space had 123 Kansei words. To reduce the initial semantic space to a smaller and (as much as possible) equivalent set of words, an affinity diagram was first used.

Figure 4.1. Two details of the post-its used for preparing the affinity diagram (first step, left; second step, right)

Each one of the 123 words were written on a yellow post-it. Post-its were stuck on the wall, grouping all words that were considered similar in meaning (Figure 4.1, left). Thirty three groups were obtained. Each group was then labeled choosing the most representative word in the group. This word was written on a blue post-it. Again, groups were made with all blue post-its to further reduce the number of words, and a representative word was chosen for each group and written on a pink post-it (Figure 4.1, right). The initial semantic space was reduced to only 14 words. Table 4.2 summarizes the whole process done with the affinity diagram.

Table 4.2. Kansei words classification after the affinity diagram

			artificial	artificial	artificial
			divine	divine	
		festive	Christmas spirit	festive	festive
			warm	warm	
reinforcing invigorative	energetic powerful	protein-rich revitalising	stimulating strengthening	energetic	energetic
		restorative nutritive	vitamin-rich satiable	nutritive	
relaxing	calming	anti-fatigue	anti-stressing	relaxing	relaxing
erotic sensual	exciting attractive seductive	romantic tempting wild	aphrodisiac explosive passionate	seductive	seductive
	flamboyant	colorful	appealing	appealing	
	fresh	cold	refreshing	refreshing	refreshing
	sweet	sugary	gluttonous	sweet	juvenile
	infantile	juvenile	mature	juvenile	
	vibrant	amusing	gratifying	amusing	
dense	creamy	light	juicy	lightly	lightly
delicate	soft	silky	velvety refined	soft	
		simple	comfortable	simple	
	substitutive	versatile	combinable	combinable	classical
	home-made	popular	classical	classical	
sophisticated	futuristic	luxurious	fashionable dainty	sophisticated	
	caribbean	exotic	tropical	exotic	exotic
			summery	summery	
healthy detoxifying	natural antioxidant salutary	pure moisturizing preventative	curative beneficial vital	healthy	healthy
slimming diuretic	laxative fibrous	digestive astringent	renovative purifying	purifying	

restorer	remineralizing	regenerative	bracing	bracing	
			balanced	balanced	
	excellent bad	brilliant good	fabulous mediocre	good	tasty
	subtle taste	savory	tasty	tasty	
exquisite	irresistible	pleasant	delicious	exquisite	
concentrated	intense	corpulent	consistent	concentrated	concentrated
		dry	strong	strong	
		aromatic	mawkish	aromatic	aromatic
		sylvan	flowery	sylvan	
spicy	acidic	sour	bitter	acidic	

The second part of the reduction procedure required the collection of data. From one of the books used to compile Kansei words (Millidge 2006), 4 photographs of juices were selected (Figure 4.2). The selection was made trying to capture dissimilar styles of juices (different colors, some having a glass and others a goblet, etc.). We selected 6 people (3 men and 3 women) from our research group and asked them to give a rating for each of the 4 juices on the 33 Kansei words obtained after the first affinity diagram (the ones in blue in Table 4.2). Rates were done on 7-point Likert scales (the same ones later used in the data collection phase). These 6 people did not participate in further phases of the study.

Figure 4.2. Fruit juices for the cluster analysis, reproduced from Millidge (2006)

Although these 6 people had to rate 33 Kansei words, the amount of time required was reasonable, as only 4 juices were used. The procedure took about 15 minutes for each participant.

As 6 people were used, each juice had 6 ratings for each Kansei word. These 6 ratings were summarized using the average. Table 4.3 (left) shows the average points each juice got for each Kansei word.

A k-means clustering was then performed; using the command `kmeans` from the base package `stats` in `R`. k-means is a non-hierarchical clustering procedure, so it is possible for two Kansei words to be split into separate clusters after having been joined together. The problem with non-hierarchical clustering techniques is that they are dependent on the initial assignation of observations. If no initial proposal of clusters is made, observations are randomly assigned in the first iteration. But k-means procedures work best when a good starting point for clusters is provided.

Fortunately, we do have a thoughtful starting point: those groups obtained from the affinity diagram. Hence a k-means clustering was done on the 33 Kansei words, assigning each word to one of the initial 14 clusters obtained in the last affinity diagram. Table 4.3 (right) shows the 33 words in a column, each word having a number from 1 to 14 giving the initial assignation to a cluster. After the k-means procedure, only the cluster containing the word “relaxing” is kept without changes. All other clusters are modified a little bit, some become smaller, some become bigger and some keep the same size but change one of the Kansei words. Figure 4.3 graphically summarizes the changes made in the clusters.

Some refining seems quite appropriate. For example, “aromatic”, “sylvan” and “acidic” constituted an initial cluster. After the k-means procedure, this cluster loses the word “aromatic”, and this word is added to the cluster having the words “seductive”, “sophisticated” and “strong”. Actually, “sylvan” and “acidic” are semantically similar words, whereas “aromatic” can be considered somewhat different, so changing this last word to another cluster seems reasonable (although perhaps the resulting cluster with words “seductive”, “sophisticated”, “strong” and “aromatic” is more difficult to justify). Anyway, the final clusters can be considered at least as logical as the initial ones, and have the advantage of incorporating the information extracted from actual ratings using juices.

One word from each of the final clusters shown in the right part of Figure 4.3 should be chosen to name the cluster. This selected word is in bold in Figure 4.3. Choosing the word that better synthesizes the ideas conveyed by each cluster can be a rather awkward task. Consider again the cluster with the words “seductive”, “aromatic”, “sophisticated” and “strong”. “Seductive” was selected as the word representing this cluster, but it is hard to justify why this choice is better than, say, “sophisticated”.

Table 4.3. Average ratings for the 4 initial juices (left) and groups before and after the k-means cluster analysis (right)

				
acidic	3,50	2,17	4,33	4,00
amusing	4,67	2,67	4,17	4,33
appealing	5,33	4,33	4,17	4,67
aromatic	4,17	3,50	4,17	3,50
artificial	4,33	3,50	3,50	3,33
balanced	3,50	4,83	4,50	4,67
bracing	3,33	4,50	4,00	4,67
classical	2,50	3,50	3,17	4,50
combinable	4,33	3,83	2,83	3,83
concentrated	4,17	4,50	4,00	4,00
divine	3,67	2,33	2,83	2,67
energetic	3,67	3,33	4,67	4,33
exotic	5,00	5,50	4,50	2,67
exquisite	3,67	3,83	3,50	4,17
festive	5,83	3,83	5,00	4,00
good	4,33	4,00	4,33	5,67
healthy	3,83	4,50	5,50	5,50
juvenile	4,17	3,83	3,83	4,50
light	3,83	3,33	4,17	5,33
nutritive	3,83	4,83	5,50	5,00
purifying	3,67	3,33	4,17	4,50
refreshing	5,00	3,83	4,33	5,00
relaxing	5,00	4,33	3,50	4,67
seductive	4,50	3,33	4,50	3,83
simple	2,83	3,00	4,00	5,50
soft	2,83	4,17	4,17	5,00
sophisticated	4,50	3,83	4,00	3,33
strong	4,50	3,17	4,00	3,83
summery	5,33	3,83	5,00	5,67
sweet	3,83	4,67	3,50	3,83
sylvan	3,50	2,50	3,83	2,83
tasty	4,33	5,00	4,67	5,33
warm	2,83	2,17	2,50	2,33

Before		After	
artificial	1	1	divine
divine	1	1	*warm*
festive	2	2	*artificial*
warm	2	3	nutritional
energetic	3	3	*healthy*
nutritive	3	4	relaxing
relaxing	4	6	*appealing*
appealing	5	6	refreshing
seductive	5	6	*summery*
refreshing	6	7	*energetic*
amusing	7	7	amusing
juvenile	7	7	juvenile
sweet	7	7	*purifying*
light	8	8	light
simple	8	8	simple
soft	8	8	soft
classical	9	8	*classical*
combinable	9	9	*sweet*
sophisticated	9	9	combinable
exotic	10	9	*exquisite*
summery	10	9	*concentrated*
balanced	11	10	*festive*
bracing	11	10	exotic
healthy	11	11	balanced
purifying	11	11	bracing
exquisite	12	12	good
good	12	12	tasty
tasty	12	13	*seductive*
concentrated	13	13	*sophisticated*
strong	13	13	strong
acidic	14	13	*aromatic*
aromatic	14	14	acidic
sylvan	14	14	sylvan

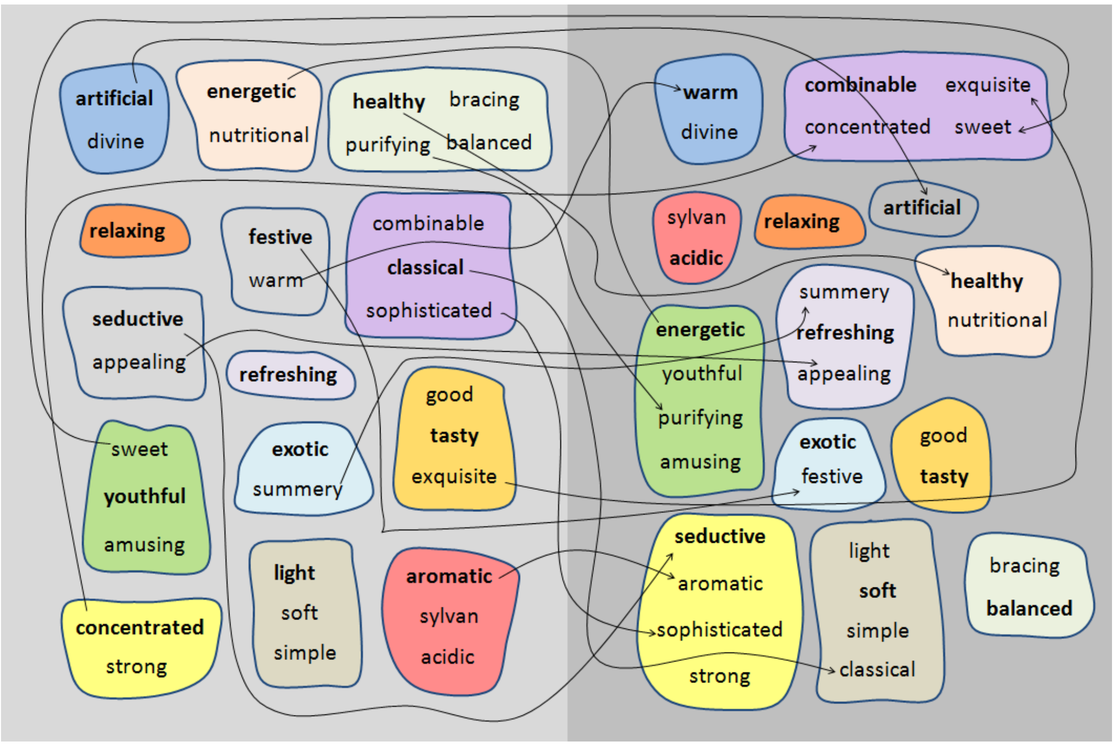


Figure 4.3. Rearranging of groups after the k-means cluster analysis

This selection of a representative word for each cluster becomes even more complex if the quantitative procedure for reducing the semantic space is a principal component analysis (PCA). In this case, the final words come from labeling the first principal components focusing on the original Kansei words that contribute more to each principal component: sometimes quite a "creative" job.

A nice representation of the clusters is obtained if a PCA is performed on data in Table 4.3, a variables factor map for the first two components is drawn and words are color-coded according to its cluster (Figure 4.4). If the first two principal components incorporate a good amount of the total data variability (say, at least, 50% considering both components), words in the same cluster should be located close in the graph, and hence this representation somehow checks the validity of the obtained clusters.

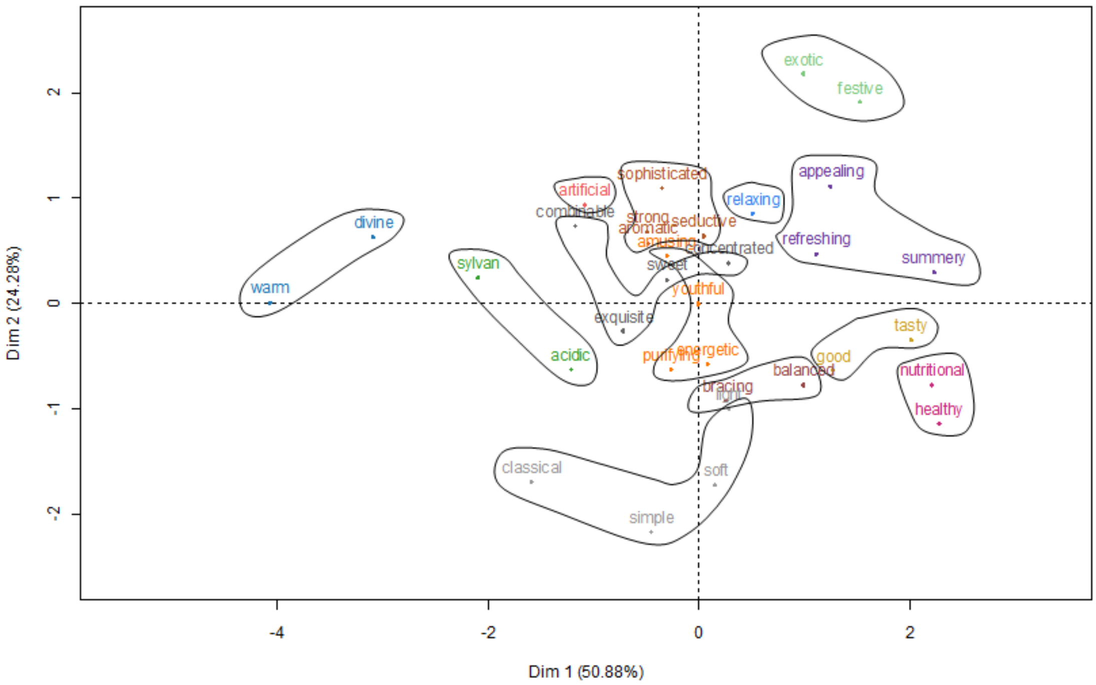


Figure 4.4. A variables factor map from the PCA made with the initial ratings on juices, with words colored according to their cluster membership

4.3.3. Final list of Kansei words.

The semantic space is completely deployed when the list of Kansei words that will be used in the data collection phase is clear and finished. The data reduction procedures explained in the previous section help a lot in obtaining this final list.

However, it sometimes happens that people responsible for the product have a special interest in knowing if a specific emotion is elicited by the product, and they want to incorporate this word to the list. Or people conducting the experience decide that the list of words is too long, as the experimental effort required would be too great. In these cases, words can be added or eliminated. But it is essential to keep in mind that the final set of Kansei words should cover the whole semantic space, and that the initial collection of words and the subsequent reduction guarantees (at least in theory) this wide coverage.

In our experiment with juices, 13 words constituted the reduced semantic space: warm, combinable, acidic, relaxing, artificial, energetic, refreshing, healthy, exotic, tasty, seductive, soft and balanced. We thought 13 words were too many for the data collection procedure, so we decided to use only some of them[47]. As our main interest in the study was having a good quality data set (more than the actual conclusions that were to be obtained), we eliminated words that we considered difficult to interpret related to juices (soft, balanced) or that were more attached to the physical properties of the juices than the others (warm, combinable, acidic).

Eight Kansei words remained: relaxing, artificial, energetic, refreshing, healthy, exotic, tasty and seductive. Two more decisions were made:

1. The word energetic was also eliminated, as it was also considered to be attached to the physical properties of the juices (the energy the juice can provide is something objective that can be measured in calories). I now think this was a mistake. Regardless of the fact that we can know the number of calories a juice provides, the perception of the juice as being energetic or not can be relevant. In any case, energetic was not considered in the study.
2. All Kansei words but one conveyed perceptions that could be qualified as positive: relaxing, refreshing, healthy, exotic, tasty and seductive. Artificial, on the contrary, gave a negative idea. We thought that this fact could confuse participants and have some undesired influence in the ratings. So it was decided to change the word from artificial to natural, thus all 7 Kansei words had positive ascriptions. Again, I am not sure now if this was a good idea. The word natural was originally in the initial semantic space (Table 4.1), but was dropped in the first affinity diagram. Words healthy and natural were located in the same manually done cluster, and healthy was selected as representative of that cluster (Table 4.2). Artificial has always been a one word cluster and has been kept until the last step. Perhaps artificial and natural mean more in the semantic space than just being antonyms.

Either way, the final Kansei words used in the study were: refreshing, healthy, exotic, seductive, natural, relaxing and tasty.

[47] In fact, 13 Kansei words are not that many in common KE studies. But we wanted to repeat the whole rating procedure twice (I will explain the reason later), and we did not want to have more than 7 or 8 words.

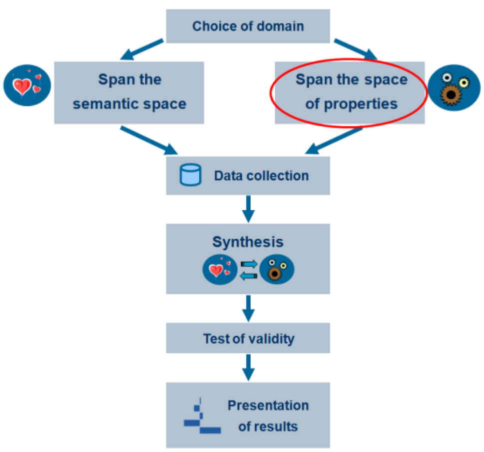
4.4. Spanning the space of properties

As the Kansei words span the semantic space and try to represent the whole Kansei the product in study can convey, the space of properties lists the physical properties of the product that can have an effect in the elicited Kansei. Spanning the space of properties is sometimes considered similar to spanning the semantic space (Schütte et al. 2004). And in some way it is similar, as there are many potential physical properties that could be used, but at the end we will only work with a subset of these physical properties.

Spanning the space of properties comprises the following steps:

1. Making a list of all possible physical product properties and selecting the ones that apparently have the largest impact in the users' Kansei.
2. Preparing the design matrix (that is, a matrix saying how many products will be used for the experiment, and establishing the value of each property for each one of the products).
3. Selecting the products (or producing product prototypes) according to the design matrix.

4.4.1. Selection of product properties



I think spanning the space of properties in Kansei Engineering is similar to choosing factors in a design of experiments. In an experiment conducted in an industrial environment, previous knowledge from the process is used to select the variables (factors) that are more prone to have an effect in the response. The values that those factors take in the experiment are also carefully chosen to maximize the probability of detecting the factors' effects, if these effects really exist. In this context, each value a factor takes in the experiment is called a level.

In Kansei Engineering, the nomenclature "factors and levels" is not common. Factors are usually called items, and levels are usually called categories. Probably this is due to the fact that Quantification Theory Type I (QT1), a common technique used in the synthesis phase of KE studies, employs this nomenclature of items and categories[48]. The word category seems appropriate to refer to the different values a categorical

[48] Quantification Theory Type I will be explained in detail in Chapter 6, together with some recommendations and suggestions to improve the technique.

variable can have. This is just the case in QT1, where all variables used are categorical. On the contrary, in the field of design of experiments factors do not need to be categorical (in fact, in industrial experiments, they are more often continuous than categorical) and the word level fits better. From now on, I will basically refer to "factors and levels" and not to "items and categories" in this work. The main reason for this decision is that I will try to connect the space of properties in Kansei Engineering with the design of experimental matrices in the field of design of experiments, where the terms "factors and levels" are used.[49]

Imagine our domain of interest is watches. Physical properties that are under producers' control are, for example, shape of the watch or color of the face. These two properties could be factors in a KE study. The factor shape of the watch could have two levels: round shape and rectangular shape. The factor color of the face could have more than two levels: golden color, white color, light blue color, etc. The selection of factors implies the selection of levels for each factor. Sometimes, the process goes the other way: we have in mind the levels (e.g. watches are round and watches are rectangular), and then we choose the factor where these levels can be accommodated (e.g. shape of watches).

The space of properties has received little attention in Kansei Engineering literature (on the contrary, a lot more can be found about the semantic space). Perhaps one of the only articles entirely devoted to this aspect of KE is a paper by Simon Schütte (2006). In this paper, a flowchart is suggested to help in the process of selecting the factors. Some appealing ideas are presented, such as the necessity to create new concepts and identify new properties to make KE more innovative. Also a method for selecting the factors that will be used in a KE study is proposed in the paper. The proposal is that each person on the team votes on, say, the two most important factors in his or her opinion. All votes are added up, and a Pareto chart with the total votes is drawn. Those factors with higher bars in the Pareto chart are the ones used for the study[50].

Anyway, engineers or technicians usually choose factors merely using previous experience and intuition. How many factors must be selected? The answer should be as many factors as necessary to describe the whole space of properties of the product. In practice, as we will later see, the number of factors we can choose is somehow determined by the number of products that will be rated by participants in the KE study.

[49] An accessorial reason is that "factors and levels" sounds more familiar to me.
[50] This method of voting for selecting the factors was used by Simon Schütte in a course about Kansei Engineering offered in Barcelona in April 2007. The results from this voting were very reasonable.

In our KE study with juices, a brainstorming session was conducted to produce as many factors as possible for juices. We thought about properties that could impact in the visual perception of the juices and that were easy to modify when taking the photographs. Finally, five factors were selected (Table 4.4).

Factor	Levels
Straw	Yes
	No
Decoration	Yes
	No
Ice	Yes
	No
Container	Glass
	Goblet
Color	Yellow
	Orange

Table 4.4. Factors and levels in the juices experiment.

Some factors naturally had just two levels: yes and no. Recipient and color could have had more levels. But I wanted to use just two levels for each factor, because then all the theory of two-level factorial designs could be used for the analysis, and this was something desired for this "academic" experience.

4.4.2. Matrix of experiments and selection of products.

Once the factors that will be used in the KE study are selected, the set of products that will be rated by participants must be prepared. Remember that each product is rated by each participant on every Kansei word. A KE study can be done (in theory) not only with physical products, but also with intangible products (what is usually called a service). As I find it a bit strange calling product a service, often in this work I use the word stimulus instead of product meaning the element that each participant rates on every Kansei word.

There are two very different situations regarding the selection of products for the study:

1. When the products are created.
   If data is collected using real products and not photographs or computer animations, prototypes can be used. These prototypes perhaps do not work, but have the same appearance as the final products. If photographs or computer animations are used, there is no need to have physical prototypes,

as Computer Aided Design (CAD) and rendering can be used to create realistic stimulus.

The advantage of this approach is that the design matrix of the experiment dictates which prototypes must be prepared. This matrix can have the desired characteristics (such as orthogonality) that allow a good estimation of the factors' effects. The problem is that preparing those prototypes can be expensive.

In this approach, first the design matrix is built and later the prototypes are prepared following the design matrix.

2. When existing products are used.

   In this approach, a set of existing products is collected trying to cover as many combinations of factors' levels as possible. Sometimes, not many products are available, and simply "you do what you can". In this situation, the available products define the design matrix.

   The problem here is that, most probably, this matrix is not orthogonal and not all factors' effects can be estimated. Imagine for example that we want to check if rectangular watches are perceived as more modern than round watches, but we only have round watches in our set of products. Obviously, it will be impossible to detect the effect of shape in watches. The advantage is simplicity: you just need to collect the products you have at hand.

   In this approach, first the products are collected and later the design matrix is built according to the products.

   Another element of caution in this second approach must be taken into account: not showing products that have its brand on it (or that its brand is easily recognizable although it is not explicitly shown, as with the Coca Cola bottle). As we saw in Section 2.4.2, some brands have a deep emotional impact that could distort the results of our KE study.

Even though the first approach (building prototypes) is much better for having good estimations of effects in the synthesis phase, the second approach (using what you already have) is more common in Kansei Engineering literature.

I think there is a need to look deeper into the consequences of using "ill-conditioned" design matrices in KE studies, and that some simple ideas can improve the performance of many methods used in the synthesis phase. I will study these topics in Chapter 7 of this dissertation.

In two-level factorial designs, it is common to use –1 and +1 for denoting both levels of each factor in the design matrix. In KE studies it is usual to have factors with more than two levels, so dummy variables are usually employed. For each factor, each dummy variable takes the value 1 or 0 to indicate the presence or absence of some level in each stimulus.

In the juice experiment, the prototypes were prepared according to the design matrix. The design matrix is a 2^{5-1} factorial design, which has resolution V. This means that main effects are confounded with interactions of order four and higher, whereas second order interactions (in which we are also interested) are confounded with interactions of order three and higher. Hence it is a design with a very nice alias structure. Table 4.5 shows the design matrix for the juice experiment.

As we can see from the design matrix, 16 different photographs of juices must be prepared following the specifications. For example, the first juice must have no straw, no decoration, no ice, use a glass as a container and be orange in color. We took all the photographs in one day, in a seminar room at the university. Ana and Eli brought a glass and a goblet, straws and ice (we kept the ice in a freezer until the last moment) to the seminar room. I bought several juice packages; Granini multifruit nectar for orange color and Granini tropical multivitamin nectar for yellow color (both of these nectars have very pure colors, that was the reason for choosing them. The smell or the taste do not matter at all, as we were only going to show photographs). Ana and Eli also cut orange and lemon peels for decorating some of the juices (the orange ones with the orange peels and the yellow ones with the lemon peels).

Table 4.5. Design matrix for the fruit juice experiment

	Straw	Decoration	Ice	Container	Color
1	No	No	No	Glass	Orange
2	Yes	No	No	Glass	Yellow
3	No	Yes	No	Glass	Yellow
4	Yes	Yes	No	Glass	Orange
5	No	No	Yes	Glass	Yellow
6	Yes	No	Yes	Glass	Orange
7	No	Yes	Yes	Glass	Orange
8	Yes	Yes	Yes	Glass	Yellow
9	No	No	No	Goblet	Yellow
10	Yes	No	No	Goblet	Orange
11	No	Yes	No	Goblet	Orange
12	Yes	Yes	No	Goblet	Yellow
13	No	No	Yes	Goblet	Orange
14	Yes	No	Yes	Goblet	Yellow
15	No	Yes	Yes	Goblet	Yellow
16	Yes	Yes	Yes	Goblet	Orange

After taking the photos, I retouched them a little bit using the photo editor Paint Shop Pro (basically, the original background was erased to have the juice with a white background). The 16 photographs are shown in Figure 4.5. These 16 photographs were used as stimulus in the data collection phase.

Figure 4.5. The 16 stimulus created according to the design matrix.

The design matrix of the experiment is shown again in Table 4.6, but now coded using dummy variables.

Table 4.6. Design matrix for the fruit juice experiment, expressed using dummy variables

	Straw		Decoration		Ice		Container		Color	
	No	Yes	No	Yes	No	Yes	Glass	Goblet	Orange	Yellow
1	1	0	1	0	1	0	1	0	1	0
2	0	1	1	0	1	0	1	0	0	1
3	1	0	0	1	1	0	1	0	0	1
4	0	1	0	1	1	0	1	0	1	0
5	1	0	1	0	0	1	1	0	0	1
6	0	1	1	0	0	1	1	0	1	0
7	1	0	0	1	0	1	1	0	1	0
8	0	1	0	1	0	1	1	0	0	1
9	1	0	1	0	1	0	0	1	0	1
10	0	1	1	0	1	0	0	1	1	0
11	1	0	0	1	1	0	0	1	1	0
12	0	1	0	1	1	0	0	1	0	1
13	1	0	1	0	0	1	0	1	1	0
14	0	1	1	0	0	1	0	1	0	1
15	1	0	0	1	0	1	0	1	0	1
16	0	1	0	1	0	1	0	1	1	0

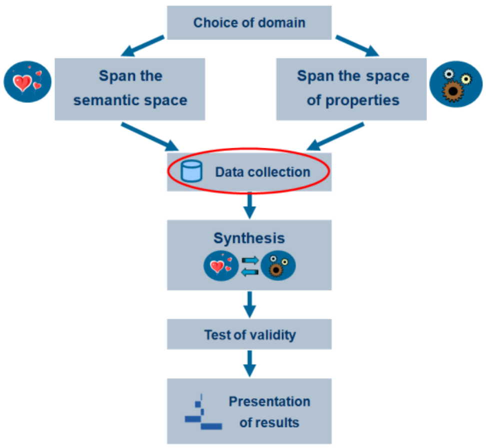
4.5. Data Collection

In my opinion, the data collection phase is one of the most important ones (if not the most important) in a KE study. The principle of "garbage in, garbage out" applies here: if our raw data is not reliable, the conclusions from our study will be poor or even erroneous, no matter the sophistication of the statistical analysis performed later. Moreover, the way the data is collected determines, at least to some degree, the kind of analysis that can be conducted afterward.

4.5.1. Location and time needed

Data for a KE study can be collected face to face, presentially or over the Internet. Using a survey on the Internet is easy and not very expensive, so the number of participants can be quite high. However, it poses several problems:

- It is not possible to use real products as stimuli. Although not only photographs, but also videos can be used to present the products, the affective channel width will probably be narrower than in a presential KE study (senses as smell or touch cannot be involved in the evaluation)[51].
- Some kind of Internet platform is needed for collecting the data (KESoft, developed by the Kansei Engineering Group in Linköpings Universitet, is an example). The data collection cannot be done – obviously – in the more "traditional" pen and paper way.
- Although it is easier to have a bigger group of respondents, if they are invited to participate by e-mail, the risk of having an auto-selected sample – and thus some bias in the results – appears.



But perhaps the biggest problem is that control over how the survey is filled out is completely lost. As I have stated in Section 4.2 when talking about choice of domain, defining the context for presentation is a very important issue in a KE study. This atmosphere can only be assured if data is collected face to face. Moreover, having all

[51] Perhaps, in the near future, virtual reality can be used to present products over the Internet, without the need for special devices. Today this is not yet possible or at least not easy to do. Chapter 9 contains a brief discussion on the use of virtual reality for presenting stimuli in KE studies.

participants in a room allows the researcher to control how respondents behave, to answer any doubts they have, to check if they remain focused on the task, etc.

Of course, collecting data on a computer (not necessarily over the Internet) also has its advantages:

- All data is immediately available in electronic form (without the need to copy data from pieces of paper to the computer afterward, which is undoubtedly a source of mistakes).
- Real-time algorithms to present stimuli can be used (something done, for example, in adaptive conjoint analysis). Randomization of stimuli for each participant is also straightforward when using a computer.
- Other responses besides the ratings, such as time needed for giving the rate or selecting a product, are easy to measure using a computer. Some analysis techniques could be used to take advantage of these measures [52].

One idea for having the advantages but not the problems would be to use computers to collect data, but having all participants in a room at the same time.

Time needed to complete the survey must also be taken into account. Common sense tells us that the shorter the better. There are plenty of websites devoted to online surveys, and they usually recommend a maximum of 10 or 15 minutes. It seems proven that substantial length of surveys affects response rates and the quality of those responses which are obtained (Burchell, Marsh 1992). Although time needed to complete the data collection phase in a KE study is not usually reported, the number of Kansei words and stimuli used in some studies surely involve half an hour surveys or even longer.

In my opinion, suitable length is also linked to the fact of being together with other participants and the person leading the KE study. Probably temptations to abandon the study or to start answering in a crazy way are higher if sitting alone in front of the computer instead of being with other participants in a prepared room. Furthermore, the time a person can be concentrated on a task obviously depends on the person and his or her motivation. Also cultural differences can be relevant, and in some societies people might be more "docile" and thus easier to keep rating many products.

[52] For example, products that need little time to get a rating are those with most significant properties affecting that rating. Some kind of weighting depending on this time could be proposed for the synthesis phase. This idea is, for example, exploited in Barone, Lombardo & Tarantino (2007).

If I had to give a number, I would say that a KE study, done with motivated participants who are all simultaneously focused on the task of rating the stimuli, should have a data collection no longer than 30 minutes.

In the juice experiment, data collection was done in a seminar room in the Faculty of Mathematics and Statistics (FME), in Barcelona[53] (a familiar place to both Xavier and me, as we give lectures there, and to Ana and Eli, as they were students in that Faculty). We wanted to repeat the whole study twice, in order to have two replicates of each rating. All studies will be done with data from the first replicate until Chapter 7. However, Chapter 8 will use data from both replicates, as repeatability in KE studies will be addressed.

Ten minutes were used to welcome participants and explain the procedure for the data collection. The first replicate was collected in 10 minutes. Then there was a 25 minute break, during which a small snack was offered (no fruit juices of any kind were available, the only drinks were mineral water, coke and beer). The second replicate also took 10 minutes. Five additional minutes were used to thank participants for their collaboration and say goodbye. The whole study took, approximately, one hour.

4.5.2. Procedure

Before starting the actual data collection, it is very important to be sure that every participant understands the same thing when reading each Kansei word. To guarantee this, a definition of each Kansei word can be given before starting the data collection. This has the problem of rationalizing the procedure even more, when asking people to think about each word before rating the products to ensure they cognitively understand the meaning. But I think this is better than discovering that some words were unclear once the data collection has already finished.

The data collection has three dimensions: subjects, Kansei words and stimuli. There are two possibilities for collecting data:

1. Each participant is presented with a product. He or she then rates this product on all the Kansei words. (Figure 4.6, left)
2. Each participant is presented with a Kansei word. He or she then rates all products on this Kansei word. (Figure 4.6, right)

---

[53] More precisely, it was done in rooms ℝ and ℚ of the Faculty of Mathematics and Statistics, in C. Pau Gargallo 5, 08028 Barcelona.

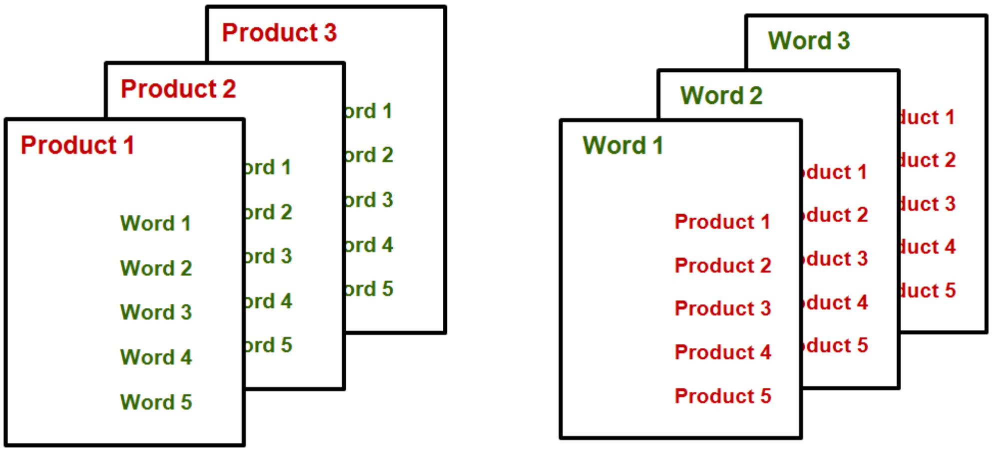


Figure 4.6. Two ways of rating: Kansei words on each product (left) or products on each Kansei word (right)

The first option is the most commonly used, and I think it is the most natural and appropriate. The respondent can concentrate on a product, and after experiencing the Kansei he or she feels with that product, give a rating on each Kansei word.

To avoid any bias, products and Kansei words are randomized as much as possible. If the data collection is done with a computer, this randomization is easier to achieve.

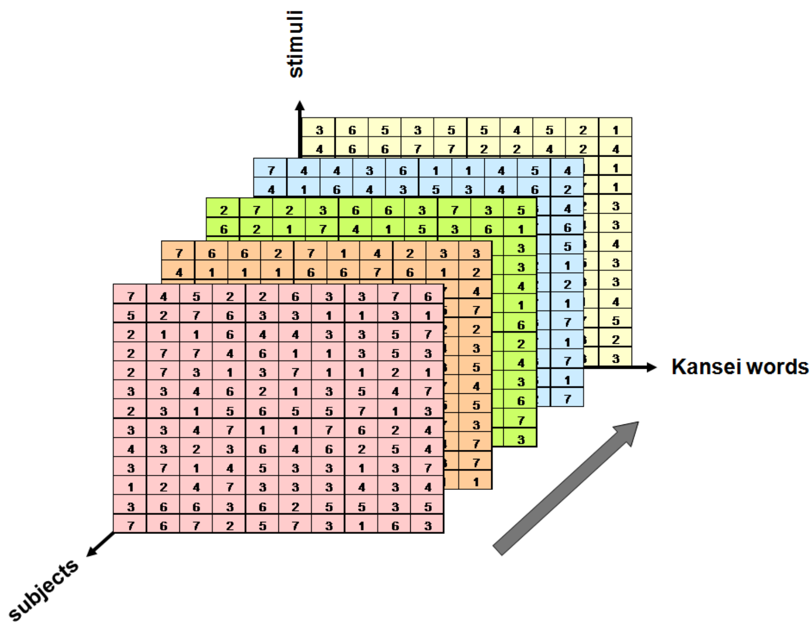


Figure 4.7. A three dimensional matrix with all the data from a Kansei Engineering Study

At the end of the process, a three dimensional matrix of data is obtained (Figure 4.7). Usually, this matrix is collapsed following the grey arrow in Figure 4.7: the subjects' dimension is lost, and the average for all participants on each stimulus and Kansei word is used.

In the juice experiment, an introduction was given to participants to explain the purpose of the study and the procedure. A legend with all Kansei words and a short definition of each one was distributed before starting the rating process and participants were urged to read them and encouraged to ask any questions they might have (Figure 4.8).

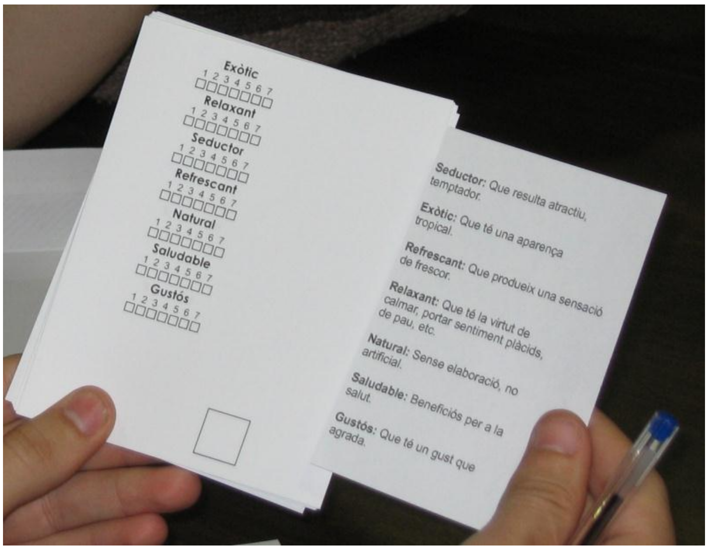


Figure 4.8. The legend with a definition of the Kansei words, and one of the forms for rating Kansei words.

All participants (24) were randomly divided into 4 groups, with 6 people each. Each group had a laptop, where the photograph of each of the 16 juices was shown (Figure 4.9). Each participant had forms like the one in Figure 4.8, one for each juice (but each form had Kansei words in different order, so hundreds of different forms were prepared). When all members in a group had finished rating all words on a form, the next juice was shown (there was no time limit for rating). Each group had the juices presented in a randomized order.

After all members in the 4 groups had rated all 16 juices, a small snack was offered, and the whole procedure was repeated in order to have a second replicate (as stated above, this second replicate will be used in Chapter 8 of this dissertation).

Figure 4.9. Groups of 5 or 6 people rating the juices shown on the screen.

4.5.3. Motivation of participants

We have already seen that one of the tasks in the choice of domain is selecting the target group. The participants for the study must be a representative sample of this target group. Nothing more is usually assured about the target group in Kansei Engineering literature.

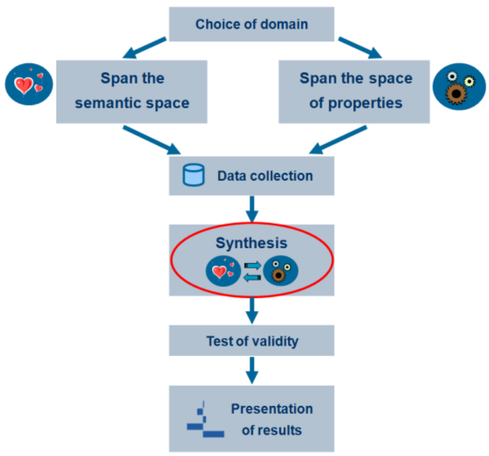
However, I think one commonly forgotten issue is vital for guaranteeing the reliability of data collected in a KE study: motivation of participants. If participants understand the purpose of the study, take it seriously and try to do a good job, data quality is much better.

In our experience, participants were quite heterogeneous (much more than in a "real" KE study). Table 4.7 shows name, gender and age of participants in the study. But they were all family and friends, so they were personally interested in the success of the study and thus tried to do their best all the time, without complaints.

Table 4.7. Name, age and gender of the participants in the study.

Name	Age	Gender	Name	Age	Gender
Antonio	25	Male	Lydia	58	Female
Belén	17	Female	Manolo	34	Male
Núria	22	Female	Marta	33	Female
Carla	23	Female	Merce	30	Female
David	25	Male	Miriam	22	Female
Enrique	23	Male	Mònica	46	Female
Erli	29	Female	Helena	30	Female
Eva	21	Female	Pere	50	Male
Guillem	21	Male	Raquel	30	Female
Héctor	44	Male	Sergi	18	Male
Imma	26	Female	Vicenç	21	Male
Jose	59	Male	Xavi	32	Male

4.6. Synthesis



Once the data collection is completed, data must be analyzed. The synthesis phase is the core of Kansei Engineering, as the link between emotions and physical properties of products is made here. When using quantitative data, there are basically two approaches for linking the semantic space with the space of properties:

- Using statistical tools.
- Using automatic learning techniques.

We also face two possibilities for treating our raw data:

- Averaging data on all subjects (and thus having continuous data, thanks to the central limit theorem).
- Treating data at an individual level, thus having ordinal data.

Having continuous data or ordinal data obviously affects the suitable tools that can be used in the synthesis phase.

I will for the moment stop the discussion on the synthesis phase here (and not because this topic is unimportant; on the contrary, it is so vital that Chapter 5 will analyze the most used techniques for completing the synthesis phase, and Chapter 6 and 7 will look deeper at these commonly used techniques).

The fruit juice experiment specifically used a 2^{5-1} factorial design. With this design, main effects and second order interactions can be estimated only confounded with three order or higher order interactions. Half normal probability plots were used to detect the significant effects[54]. Table 4.8 summarizes the significant effects for each Kansei word.

Kansei word	Significant effect
Refreshing	Ice
Healthy	-
Exotic	Decoration Color
Seductive	Color Decoration
Natural	-
Relaxing	-
Tasty	Color

Table 4.8. Significant effects for each Kansei word.

There are no significant interactions, and it seems that all factors are inert for words healthy, natural and relaxing. Figure 4.10 shows the main effects plots for those words with significant effects.

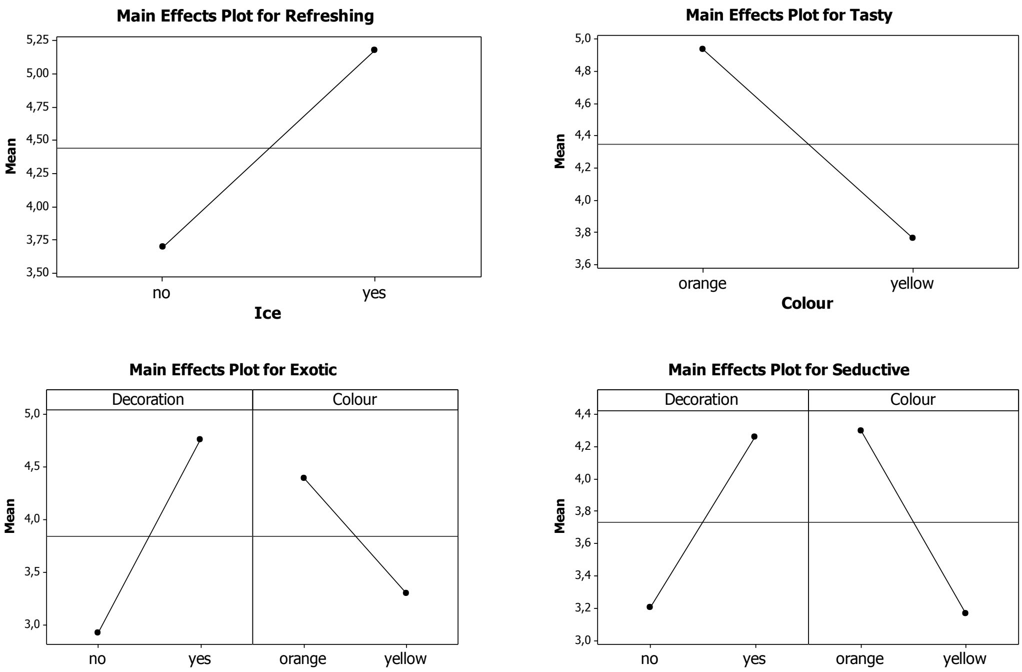
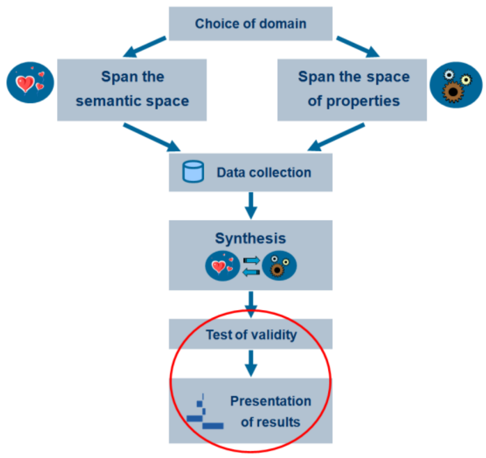


Figure 4.10. Main effects plot for significant effects

---

[54] Note that the approach for analyzing the data as a factorial design implied averaging data over all subjects. These data can also be treated at an individual level (I will show this in Chapter 6, where I introduce ordinal logistic regression).

4.7. Test of validity and presentation of results



Once the synthesis phase is completed, significant factors are revealed and their effect on each of the Kansei words is discovered. Not much is written in the available literature about the test of validity phase. In my opinion, some confirmatory experiments must be run in the test of validity phase to confirm the results from the synthesis phase. For example, as we have discovered that ice affects the perception of being refreshing, a juice with ice must be shown to some people (different from the participants in the main study, but belonging to the same target group), and specifically ask them if they consider that juice refreshing. In this confirmatory experiments, we do not try to discover which factors affect each Kansei word (this has already been done), but to explicitly validate our discoveries.

If all confirmatory experiments are successful, conclusions must be compiled to prepare the presentation of results step. If some confirmatory experiments do not provide the expected result, previous phases of the KE study must be revisited.

The last phase in the model for conducting KE studies was renamed from model building to presentation of results to emphasize the importance of giving results in a digestible way. "Clients" of KE studies are often industrial designers, marketing experts, and in general technicians not necessarily used to statistical terms. This is an important reason to prefer visual rather than numerical outputs. In my opinion, two kinds of outputs should be used for presenting results from a KE study:

- Descriptive graphs, easy to understand for everybody with a bit of effort, and that directly translate numerical data into useful information. For example, radar plots and profiles for Kansei words can be used to check similarities among words; scatterplots of the mean vs. the standard deviation for each Kansei word can give information on the homogeneity of ratings among participants.
- Graphs consequence of analyzing data from the KE study (mainly using tools for the synthesis phase). These graphs substitute numerical outputs and must incorporate aids to interpretation (making incorrect interpretation almost impossible). For example, individuals and variables factor maps from a principal component analysis can be used to check similarities among words

and stimuli. Other "invented graphs" introduced in the next chapters of this dissertation also serve this purpose.

Although it would have been convenient, no confirmatory experiments were done in the juices experiment (in fact, as two replicates were obtained for each rating, this uncommon second replicate was used for checking the likelihood of results from the first replicate).

Analysis of significant effects from the fractional factorial design (Figure 4.10) led to the following conclusions:

- Juices with ice are perceived as more refreshing.
- Orange juices are perceived as more tasty.
- Orange juices with decoration are perceived as more exotic and seductive.

If desired, an equation for obtaining the mean rating for Kansei words refreshing, tasty, exotic and seductive could be derived from the analysis (although a visual representation is always more appealing).

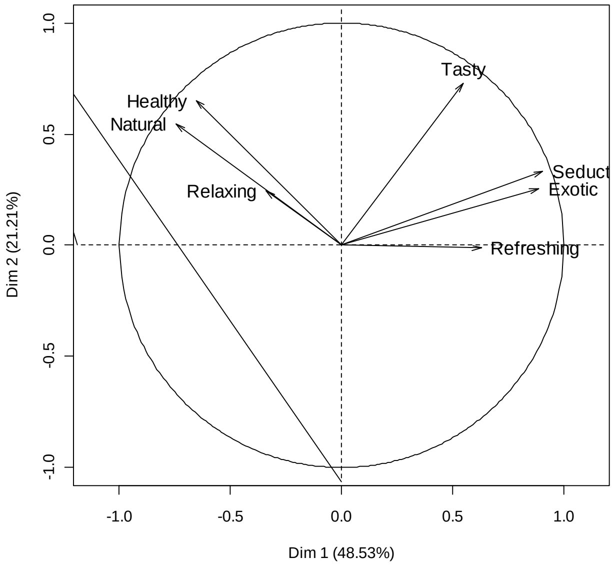


Figure 4.11. The variables factor map from a PCA with data from the juice experiment.

As I have stated above, many conclusions can be extracted by just carefully looking at descriptive graphs that do not require complex calculations. For instance, Figure 4.12 shows radar plots for all Kansei words (drawn with the function `stars` from the `R` package `graphics`). Look at the word Refreshing: juices 1, 2, 3, 4, 9, 10, 11 and 12 have short segments; whereas juices 5, 6, 7, 8, 13, 14, 15 and 16 have long segments. Juices with short segments do not have ice, but juices with long segments do have ice (collate the design matrix in Table 4.5). This explains why ice is the only significant effect for the word refreshing, and that juices with ice are perceived as more refreshing than juices without ice.

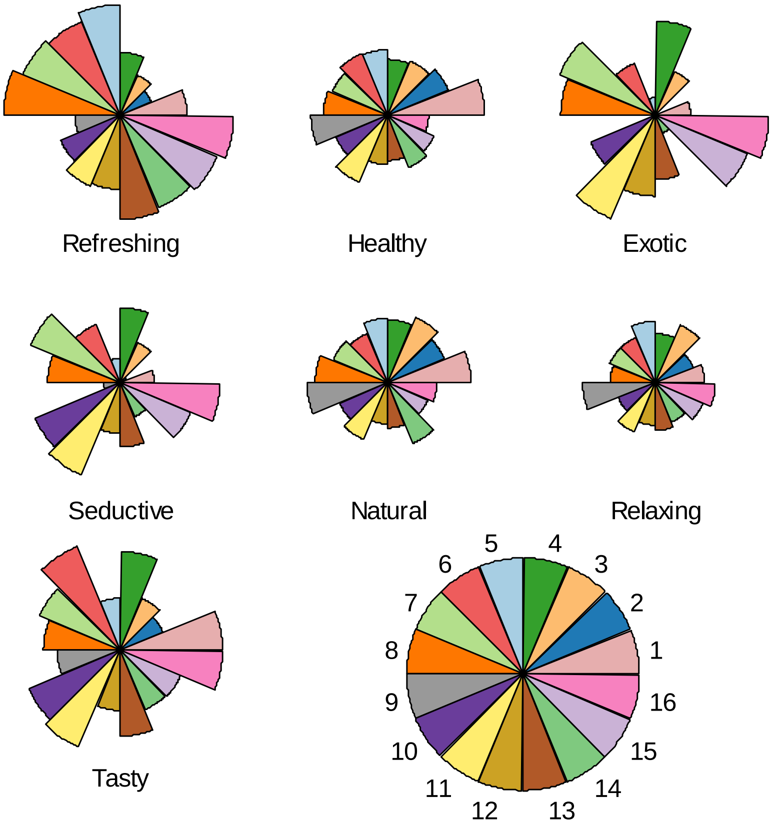


Figure 4.12. Radar plots for Kansei words.

Figure 4.11 shows the variables factor map from a principal component analysis (done with the function `PCA` from the `R` package `FactoMineR`). Words seductive and exotic are very close to each other. This is logical, as they both have color and decoration as significant effects, and with values in the same direction. The radar plots of seductive and exotic in Figure 4.12 also have a similar shape, thus confirming the fact that both words are perceived as similar. Although a decision was made to keep exotic and seductive as two different Kansei words, the results show that they could have been considered the same.

No factors have been detected as significant for Kansei words healthy, natural and relaxing. All 16 juices get a similar averaged rating for these three words (so all 16 segments in the radar plots have more or less a similar size, or at least less differences than segments in Kansei word refreshing, for instance). However, the variables factor map in Figure 4.11 places words healthy, natural and relaxing rather close to each other. Are these words also perceived as similar, as the variables factor map seems to convey?

Profiles for words healthy, natural and relaxing are shown in Figure 4.13. The graph on the left shows the mean rating for each juice. All three lines are quite parallel, thus confirming that the three words are perceived as similar. A 95% confidence interval for each word and for each juice can be calculated and added to the graph (Figure 4.13, right). Confidence intervals must overlap for all stimuli if words are rated similar (in this case, all three CI overlap for juices 2 to 16).

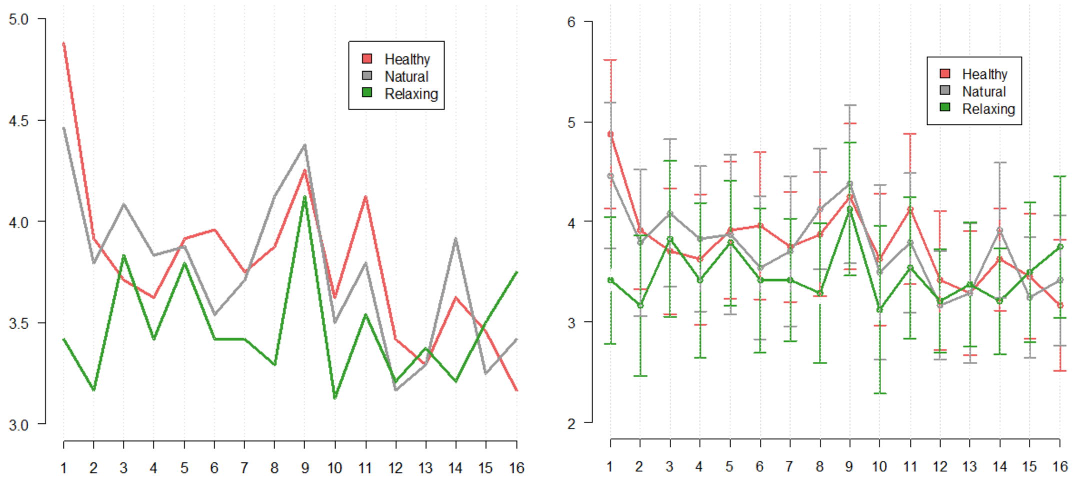


Figure 4.13. Profile for Kansei words healthy, natural and relaxing, without (left) and with (right) 95% confidence intervals for the mean.

Of course, many other graphs could be drawn to further describe the set of data from the juice experiment. Besides descriptive graphs, also graphs for visualizing the built model are needed. I will either "invent" these graphs or improve existing ones in Chapters 6 and 7.

4.8. Some thoughts looking at data from the juice experiment

I will finish this chapter with some thoughts that motivate the discussions in Chapter 8. Figure 4.14 shows ratings from the 24 participants for the Kansei word refreshing. Low ratings are red, middle ratings are yellow and high ratings are green.

	Imma	Lydia	Mònica	David	Erli	Eva	Belén	Antonio	Héctor	Helena	Vicenç	Jose	Xavi	Raquel	Merce	Guillem	Marta	Pere	Núria	Esther	Sandra	Enrique	Sergi	Manolo
1	4	4	5	4	3	7	4	2	4	4	4	2	3	4	5	2	2	5	1	2	4	6	7	7
2	1	5	3	1	1	1	2	4	4	6	3	4	3	3	3	6	4	4	1	2	1	5	3	1
3	1	2	3	1	3	1	6	5	1	5	2	4	4	5	2	5	5	3	2	1	1	4	5	7
4	1	3	3	2	3	5	6	3	5	4	3	3	4	3	4	6	3	4	4	1	4	4	7	7
5	1	6	4	5	7	4	7	5	5	6	5	6	6	7	1	6	5	7	7	6	7	6	1	7
6	2	5	4	5	7	4	6	3	6	3	6	3	4	6	4	5	3	7	5	6	7	5	7	7
7	3	1	4	7	7	5	7	2	4	3	6	6	4	6	4	6	4	5	5	6	7	7	7	7
8	4	4	5	6	7	4	3	5	5	6	5	5	5	7	1	7	5	7	7	6	7	6	7	7
9	2	5	3	1	4	3	2	5	1	6	1	4	5	5	2	3	6	4	4	1	2	5	4	1
10	1	3	3	4	2	4	6	3	4	3	3	4	5	5	4	5	2	3	4	2	4	5	7	7
11	2	3	5	6	3	7	6	2	4	4	4	5	5	6	5	2	4	3	3	2	6	4	7	4
12	1	4	3	1	7	5	1	5	4	6	4	6	4	7	2	6	4	3	4	2	1	6	7	7
13	4	1	3	4	7	5	6	3	5	3	6	6	4	6	6	6	4	5	6	7	7	6	5	7
14	4	4	4	5	7	4	2	4	7	6	4	6	4	3	3	3	5	5	5	7	7	7	6	7
15	3	3	5	6	7	7	2	3	5	5	5	4	4	7	3	4	5	6	7	5	6	7	7	7
16	4	2	4	4	7	7	6	2	7	5	6	3	4	7	4	7	4	6	7	7	7	5	7	7

Figure 4.14. Ratings for the Kansei word Refreshing

Looking at this data set for the Kansei word refreshing, we can see that:

- Some people tend to give extreme scoring (Manolo, last column, basically rates using 1 and 7), while others give ratings in the middle part of the scale (Mònica, third column, never gives 1, 2, 6 or 7).
- Some people generally give low ratings (Imma, first column, gives points from 1 to 4), while others give high ratings (Enrique, third column from the end, gives points from 4 to 7).

Although each row is the same juice, ratings from different participants almost always range from the minimum to the maximum allowed rating! If data from all participants is summarized using the average, this variability among respondents is completely removed. This is not the case when data is treated on an individual level, as we will see in Chapter 6.

A big effort is made in choosing fairly homogenous participants for KE studies[55]. However, probably variability among respondents exists even if all participants fit well in the target group (especially because socioeconomic variables are usually used to configure the target group, but this does not guarantee homogeneity from an emotional point of view). Why not take it one step further and try to segment participants – if possible and necessary – according to their emotional response? This will be one of the topics of Chapter 8.

Although I only used data from the first round in this chapter, the whole data collection process in the juice experiment was repeated once more. We have, thus, two replicates for each person, product and Kansei word.

---

55 Remember this was not done in our juices experiment, as we invited family and friends (in fact, our family and friends are probably not very heterogeneous…)

	Ana 1st round	Ana 2nd round	Difference
1	6	6	0
2	4	4	0
3	2	3	-1
4	6	6	0
5	5	5	0
6	7	6	1
7	7	6	1
8	6	4	2
9	4	4	0
10	6	5	1
11	6	6	0
12	4	4	0
13	6	6	0
14	5	5	0
15	5	6	-1
16	6	6	0

	Guillem 1st round	Guillem 2nd round	Difference
1	3	5	-2
2	7	5	2
3	3	5	-2
4	5	5	0
5	4	5	-1
6	5	6	-1
7	3	5	-2
8	6	6	0
9	5	6	-1
10	7	4	3
11	5	5	0
12	5	3	2
13	5	4	1
14	6	3	3
15	6	5	1
16	3	6	-3

Figure 4.15. First and second round ratings on the Kansei word Tasty for two participants (Ana and Guillem)

Figure 4.15 shows the first and second ratings for all 16 juices from two respondents on the Kansei word Tasty. Some participants are quite consistent in the first and second round scoring. Ana, for example, gives basically the same ratings for all juices: there are never more than 2 points of difference between the first and the second round, and more than 60% of the times she gives exactly the same rating. But some other participants fail to express the same opinion in the first and second round. Guillem, for instance, expresses quite different opinions between the two rounds in juice 10, 14 and 16, and gives the same rating in less than 20% of the juices.

Can we have some kind of replicate in KE studies that allows us to distinguish between consistent people (such as Ana) and inconsistent people (such as Guillem)? Can we use this information in some way? The second part of Chapter 8 deals with repeatability in KE studies.

5 A Literature Review on Kansei Engineering

This chapter describes a literature review done on the topic of Kansei Engineering. Statistics on the use of tools and other issues are offered based on the analyzed papers, complementing the explanations in the last chapter for the phases in the model.

5.1. Selection of works for the literature review

On March 2009 a search was done in the databases ISI Web of Knowledge from Thomson Scientific, ScienceDirect and Engineering Village from Elsevier and InformaWorld from Informa PLC[56]. The quest was performed on the words "Kansei Engineering" (inside quotation marks) in abstract, title and keywords of journal articles. Table 5.1 shows the number of results retrieved.

As some journals feed more than one of the databases, some results were duplicated. After removing the duplicates, 114 different papers were obtained. From these, 10 papers were in Japanese and 20 in Chinese. These 30 non-English papers were not considered for the selection. A total of 84 papers remained.

---

[56] The search was updated in January 2010 to include papers published during the whole of 2009.

Table 5.1. Number of results from every database

Database	Website	Number of results
ISI Web of Knowledge	www.isiknowledge.com	49
ScienceDirect	www.sciencedirect.com	39
Engineering Village	www.engineeringvillage.com	125
InformaWorld	www.informaworld.com	6

The number of journal articles devoted to Kansei Engineering has increased in the last few years, as we can see when analyzing the publication year of these 84 selected articles (Table 5.2).

Table 5.2. Number of retrieved journal articles on Kansei Engineering per year

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
1	1	0	0	0	8	1	7	1	1

2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
1	0	2	2	7	4	12	9	9	18

Figure 5.1 shows a C control chart that underlines the increase of Kansei Engineering related articles in the last few years.

From this list of 84 papers, the following criteria has been used to select the most relevant ones:

- To consider the latest research trends in Kansei Engineering, all of the most recent journal papers (from 2007, 2008 and 2009) but one, impossible to find, (Shi, Sun & Xu 2009) have been included (35 papers).
- For each one of the 48 papers published in 2006 or before, the number of times the paper was cited in other papers was sought. This is a rather cumbersome task. The procedure was to try to find the number of times a paper was cited using the cited reference search utility in Web of Knowledge (WoK). When there was no information from WoK, Google Scholar (GS) was used instead. Eighty percent of the references got the number of times cited from WoK, and the rest from GS[57]. It has not been possible to obtain the number of times 3 papers were cited: these 3 papers have not been considered anymore.

[57] As stated in several studies comparing tools for citation counts, GS tends to give slightly higher numbers than WoK. The reason is that GS searches are not limited to refereed journals, and include more results from conference proceedings and non-English language journals. (Meho, Yang 2007)

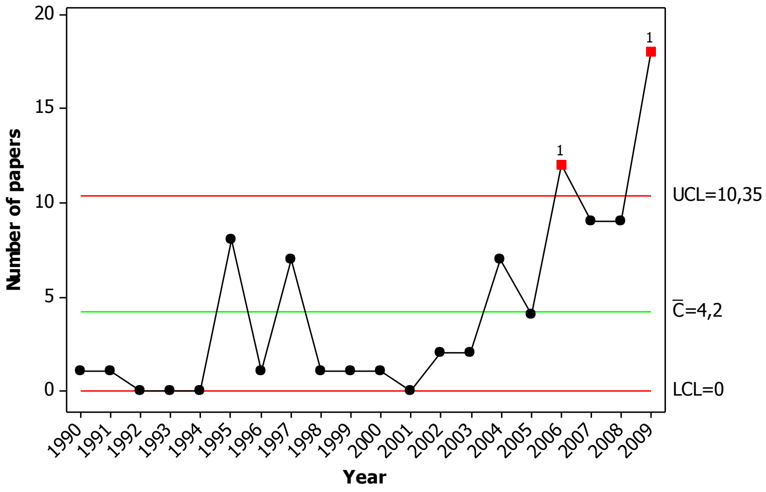


Figure 5.1. Number of retrieved journal articles about Kansei Engineering per year

The remaining 45 papers were sorted by the ratio of number of times cited per year since their publication. Papers with a ratio higher than 1 have been included (23 papers). Figure 5.2 graphically shows this cut off point.

The sole author of the two papers with the highest ratio of times cited per year is Prof. Nagamachi. The seminal paper "Kansei Engineering - a New Ergonomic Consumer-Oriented Technology for Product Development" (Nagamachi 1995a) is the journal article most frequently cited per year. It was published in 1995 and has been cited 130 times.

Therefore, a total of 58 papers have been analyzed in this literature review. All statistics and comments in the rest of this chapter are based on the study of these 58 papers.

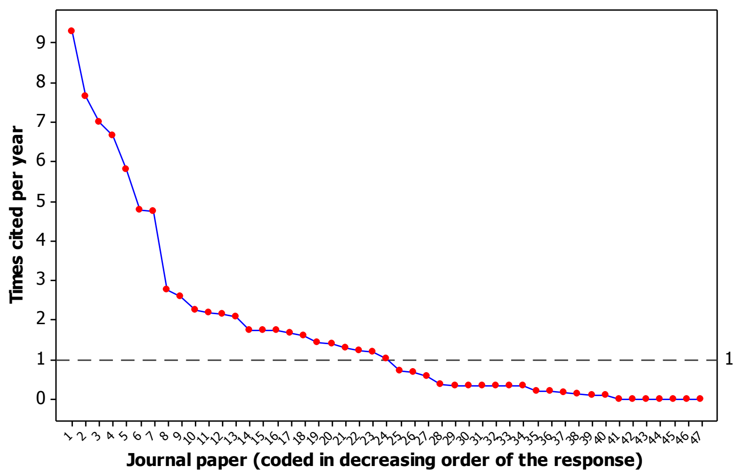


Figure 5.2. Number of times cited per year since publication for journal papers on Kansei Engineering

As Kansei Engineering studies can be applied to many different products, it is possible to find examples of KE in conferences devoted to quite dissimilar topics. However, I think two international conferences have the highest connection with Kansei Engineering:

- The First and Second European Conference in Kansei Engineering, held in Helsingborg, Sweden, in June 2007 and August 2008 under the umbrella of the Quality Management and Organisational Development (QMOD) Conference. The QMOD conference has traditionally offered many presentations on KE, but those in 2007 and 2008 had special sessions devoted to KE.
- The International Conference on Kansei Engineering and Emotion Research (KEER), held in Sapporo, Japan, in October 2007, in Osaka, Japan, in March 2009, and in Paris, France, in March 2010. This conference is devoted to what I called emotional design in Section 3.2.1.

Although no conference proceedings have been included in this literature review, I attended the First and Second European Conference in Kansei Engineering (Helsingborg, 2007 and 2008) and the International Conference on Kansei Engineering and Emotion Research (Paris, 2010). My perception from these conferences (that can be corroborated by reading the proceedings) is that the journal papers selected for this literature review correctly capture all dimensions of KE studies covered in the conferences.

5.2. Origin of the papers and common applications of Kansei Engineering

This section locates the geographical origin of the first author of the reviewed paper, gives statistics about the most common journals with papers on Kansei Engineering, and lists frequent domains where KE is applied.

5.2.1. Origin of the papers

A total of 61% of the reviewed papers come from Asia, 36% come from Europe and 3% from America. Placing the research center of the first author of each one of the reviewed papers on a map provides a good visual overview of areas where Kansei Engineering is popular. The maps in Figure 5.3 and Figure 5.4 show color-coded signs locating the author's city: a red sign means 4 or more papers, dark blue sign 2 or 3 papers, and light blue sign 1 paper. The name of cities with more than 1 paper can be found on the right. In Europe, I have decided to also mention some cities where, in my opinion, researchers on KE are doing interesting work.

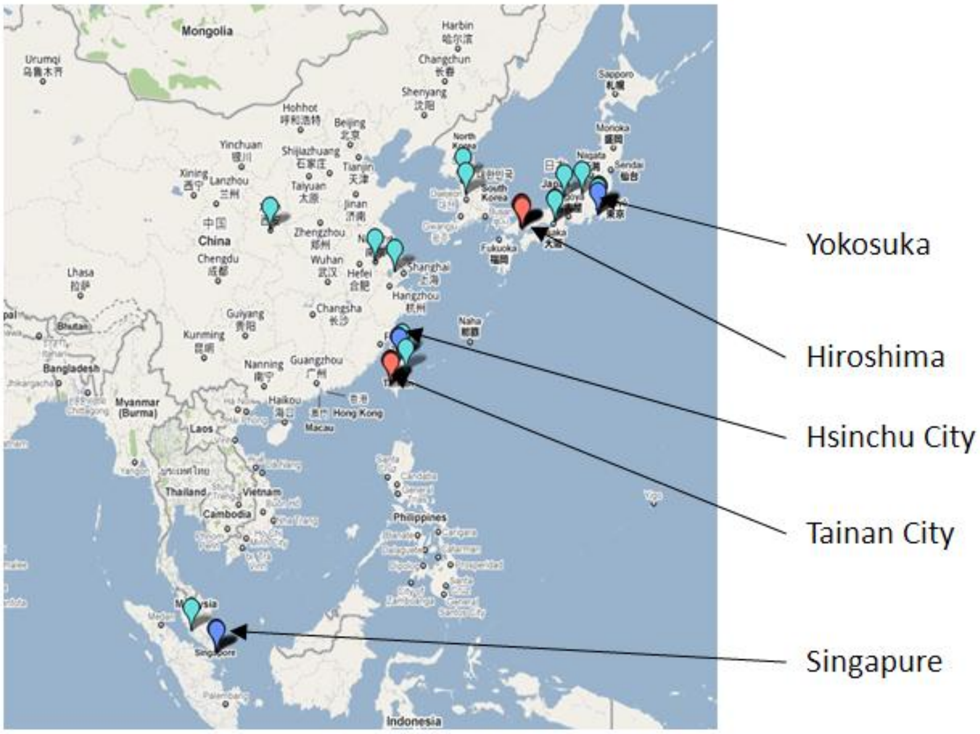


Figure 5.3. Asia map with location of first author of reviewed papers.

Figure 5.5 shows a tag cloud of the cities. Hiroshima (Japan) clearly stands out, with 8 papers, all generated in the research group led by Mitsuo Nagamachi (Shigekazu Ishihara, Keiko Ishihara, Yukihiro Matsubara and Tatsuo Nishino are other relevant researchers from this group). Tainan City (Taiwan) and Leeds (UK) come second, with 5 papers. Papers from Tainan City come from researchers in the Department of Industrial Design, in the National Cheng Kung University, such as Hua-Cheng Chang, Hsin-Hsi Lai and Yu-Ming Chang. Papers from Leeds come from the Affective Engineering Research Group, in the School of Mechanical Engineering from the University of Leeds. Tom Childs, Stephen Lillford, Brian Henson and Cathy Barnes belong to this group. Four papers come from Linköpings (Sweden), where Simon Schütte, Jörgen Eklund and Ebru Ayas, all three initially at Linköpings Universitet, have been working together with the team led by Mitsuo Nagamachi on Kansei Engineering topics.

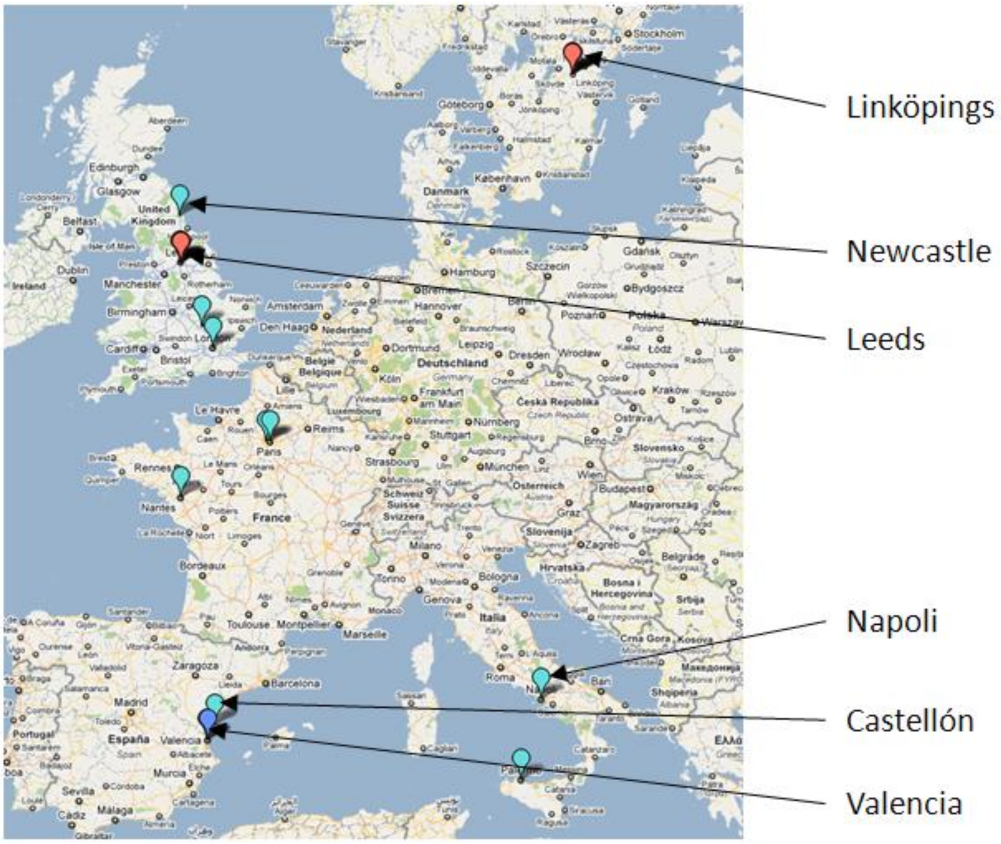


Figure 5.4. Europe map with location of first author of reviewed papers (red sign means 4 or more papers, blue sign 2 or 3 papers, and light blue sign 1 paper).

Three papers come from the School of Mechanical and Aerospace Engineering of Nanyang Technological University in Singapore, and two papers from the Institute of Applied Arts from the National Chiao Tung University, in Hsinchu City (Taiwan). Also two papers have been written by Carmen Llinares and Álvaro Page, from the Universidad Politécnica de Valencia (Spain), and two more papers by Tomio Jindo, from the Nissan Research Center in Yososuka (Japan).

Figure 5.5. A tag cloud with cities locating the first author of reviewed papers[58].

[58] This tag cloud has been created in February 2010 in the website http://tagcloud.oclc.org

Other European researchers relevant – in my opinion – for the quality of their work are Margarita Vergara and Salvador Mondragón, from the Universitat Jaume I in Castellón (Spain); Shirley Coleman from Newcastle University's Industrial Statistics Research Unit (UK), Antonio Lanzotti and Michele Staiano from the Università degli Studi di Napoli Federico II (Italy) and Stefano Barone from the Università degli Studi di Palermo (Italy)[59].

5.2.2. Journals with papers on Kansei Engineering

Table 5.3 shows a list of the journals where the papers appeared. An asterisk marks journals indexed in Journal Citations Report from ISI Web of Knowledge. Kansei Engineering is a discipline with origins in ergonomics – when founded by Mitsuo Nagamachi – and this becomes apparent in the list: more than 30% of the reviewed papers appeared in the "International Journal of Industrial Ergonomics". The 6 papers appeared in "The TQM Journal" were presentations from the First and Second European Conference in Kansei Engineering, selected for publication together with some other papers from those two QMOD conferences.

Other journals with papers on Kansei Engineering also belong to the area of knowledge of ergonomics, together with some others devoted to industrial design, product development and human-computer interaction. As Kansei Engineering studies can be applied to a huge range of products, some papers have appeared in journals committed to those products: clothing, medical devices, urban planning, etc.

Table 5.3. Number of reviewed papers in each journal (those with an asterisk are journals indexed in Journal Citations Report from ISI Web of Knowledge)

Journal	Papers
International Journal of Industrial Ergonomics *	19
The TQM Journal	6
Applied Ergonomics *	3
Expert Systems with Applications *	2
International Journal of Product Development	2
Journal of Engineering Design *	2
Artificial Intelligence for Engineering Design, Analysis and Manufacturing	1
Building and Environment *	1
CIRP Journal of Manufacturing Science and Technology	1
CoDesign: International Journal of CoCreation in Design and the Arts	1
Computers & Operations Research *	1

[59] Pietro Tarantino and Giovanna Matrone also worked with these Italian researchers in the field of Kansei Engineering.

Computers in Industry *	1
Ergonomics in Design	1
Frontiers of Mechanical Engineering in China	1
Human Factors and Ergonomics in Manufacturing *	1
Information-an International Interdisciplinary Journal	1
Internation Journal of Production Economics	1
International Journal of Biomedical Engineering and Technology	1
International Journal of Clothing Science and Technology *	1
International Journal of Computer Applications in Technology	1
International Journal of Production Economics *	1
International Journal of System of Systems Engineering	1
International Journal of Vehicle Design *	1
Journal of Applied Statistics *	1
Landscape and Urban Planning *	1
Lecture Notes in Business Information Processing	1
Quality and Reliability Engineering International *	1
Quality Management Journal	1
Theoretical Issues in Ergonomics Science	1
Wear	1

5.2.3. Common applications of Kansei Engineering

When looking at the products where KE has been applied in the reviewed papers, the inventory is quite extensive. Figure 5.6 shows a summary of the applications found in the reviewed papers. There are many case studies related with cars and phones. Many times, papers dealing with methodological aspects of KE, or introducing new methods of analysis, use mobile phones as an example of application (Chen, Chuang 2008, Jiao, Zhang & Helander 2006, Lin, Lai & Yeh 2007, Zhai, Khoo & Zhong 2009b). This is probably due to the fact that mobile phones are products that are easy to find, it is possible to select a wide range of mobile phones with different characteristics, and they are relatively cheap and small enough to be effortlessly carried from one place to another. On the other hand, papers dealing with car parts belong more to the case study category, where no huge methodological suggestions appear, but simply a case study with real implications in industry is described. Among others, there are studies about car profiles (Chang, Lai & Chang 2006, Chang, Chen 2007, Lai, Chang & Chang 2005), steering wheels (Jindo, Hirasago 1997) or car dashboards (Tanoue, Ishizaka & Nagamachi 1997, Ben Ahmed 2009).

Automotive	Car profile, analog speedometers, car dashboard, etc.
Phones	Mobile phones, telephones, etc.
Packaging	Glass bottles (for beer, for sauce), packaging for a laundry product, for a cosmetic; wrapper for a chocolate bar, etc.
Clothes	Shoes, socks, etc.
House	Wallpaper, flooring material, in the kitchen: knives, glasses, faucets.
Industrial	Rocker switches, machining centres, vehicle operated seats, color of mini earth moving
Office	Office chairs, color copy machines
Websites	e-commerce website for clothes, flowers…
Others	Massage chair, Oil paintings, train interior design, real state promotions, neighborhoods in a city, waiting areas in healthcare centers

Figure 5.6. Examples of applications of Kansei Engineering in the reviewed papers. [60]

Several papers are devoted to packaging: glass bottles for a new sauce (Barnes, Lillford 2007) or for a beer (Barnes, Lillford 2009), containers for cosmetic products (Nagamachi 2002b) or for laundry powder (Delin et al. 2007), wrappers for chocolate bars (Dahlgaard et al. 2008), or even studies about sensations elicited when touching materials later used in packaging (Chen et al. 2009). Other papers are devoted to clothes – shoes being the most common topic addressed (Ishihara et al. 1997, Pearce, Coleman 2008, Bouchard 2009), house elements – flooring materials and wallpapers (Bahn et al. 2009), kitchen faucets (Aktar Demirtas, Anagun & Koksal 2009) and knives (Chen, Chang 2009) or front doors for houses (Matsubara, Nagamachi 1997).

[60] The height of each colored box is proportional to the number of applications found

It is common to find studies on emotional design applied to some consumer products, such as electronic consumer goods or packaging, but not on industrial machines. It seems like one of those areas where functionality and beauty are – mistakenly – opposed. Instead, several reviewed papers show industrial applications of KE: rocker switches (Schütte, Eklund 2005), machining centers (Mondragon, Company & Vergara 2005) or vehicle operated seats in earth moving machines (Nakada 1997).

Given Kansei Engineering's profound links with ergonomics from its origins, it is no surprise to find applications such as with office chairs (Jindo, Hirasago & Nagamachi 1995).

The vast majority of reviewed papers show applications to specific products. However, some papers have a more holistic approach, focusing on what could be called an application of KE to experiences. For example, some studies refer to the buying experience in an e-commerce website (Akioka, Fukumori & Muraoka 2009, Lokman, Noor & Nagamachi 2009, Ishihara et al. 2008). Some others refer to the general impression given by different scenarios: a train interior (Lanzotti, Tarantino 2008), a waiting area in a healthcare center (Ayas, Eklund & Ishihara 2008) or even a whole neighborhood in a city (Llinares, Page 2008).

Having in mind KE applications not only coming from the reviewed papers, but also from presentations in the conferences appointed in Section 5.1, it can be safely said that Kansei Engineering is applied to a large range of products.

5.3. Motivation for Kansei Engineering

Many of the reviewed papers start justifying the need for Kansei Engineering. According to Delin et al. (2007), *"over 80% of new products fail"*. The reason is not that they do not work, but that they do not appeal to users. Some authors clearly explain this situation. Mondragon, Company & Vergara (2005), for example, state the following when dealing with machine tools:

> *Meeting the technical specifications (TS) is important but not always enough for a product to succeed. There are aspects that are difficult to quantify but which do influence the design and/or selection of a machine. For instance, a machine tool may meet all the TS of current safety norms but, in spite of this, it may be perceived by the operator as unsafe. Operators will consequently be less concerned with production and more worried about their personal safety, which they perceive as being exposed to danger by the machine.*

In fact, some authors consider that users will forgive performance lacks or even dysfunctions if the product is attractive enough:

> *Modern consumers not only place importance on a product's physical quality, but also employ their sentimental responses when deciding whether or not to purchase a particular product (…) The latter phenomenon is particularly evident in the case of mature consumer products such as cars, cell phones, electrical and electronic appliances, furniture, etc. It has often been shown (e.g. by Apple's iMac computer) that if products possess superior feeling features, such as form and color, they can still sell well and be well liked even if they lack obvious advanced technologies and functions. Accordingly, designing products with enhanced feeling qualities is a vital means of gaining market advantages.* (Lai, Chang & Chang 2005)

The idea that in a highly competitive market, products must not only attend to consumer demands regarding performance and function, but also attractiveness, is shared in many papers (Nagamachi 2002a, Nadaka 1997, Petiot, Yannou 2004, Yang, Nagamachi & Lee 1999, Zhai, Khoo & Zhong 2009a). Attractiveness is also extended to services: *"To be competitive, businesses must design services that do not just satisfy customers, they must also delight them"* (Gonzalez, Mueller & Mack 2007).

Despite this necessity to consider attractiveness in product design, *"the subject of emotional appeal (…) or desirability is often neglected as designers tend to pay more attention to issues of usefulness and usability"* (Lokman, Noor & Nagamachi 2009). Other times designers do integrate emotional attributes in their designs, but in an intuitive way. This is good, but not enough:

> *Traditionally, industrial designers use inspiration and cultural observation to translate feelings into product properties, but these methods pose problems which range from misinterpreting the design brief, to poor support in decision making and a lack of rationale. This in turn can contribute to the poor market performance of many new products.* (Barnes et al. 2008)

When trying to integrate emotional properties to products, "*the primary technique used is focus groups, which is often cited as the reason for low success rates for new products*" (Barnes, Lillford 2009). Too often, "*form design or styling activities are (…) reduced to a discussion based on opinion and subjectivity, with no theoretical basis*" (Petiot, Yannou 2004).

This use of qualitative tools only is not sufficient in the endeavor of designing attractive products. When talking about automotive design, Chang and Chen (2007) state:

> *Traditionally, automotive form designs were generated in accordance with the individual designer's intuitive perception and related product experience (…) However, this approach lacks objectivity since the image perception of a designer may well differ from that of the product's target consumers (…) Therefore, a requirement exists to develop tools to assist the designers to objectively and effectively generate automotive designs from a consumer image perception perspective.*

Kansei Engineering appears as a quantitative approach to emotional design: *"Quantitative data on the relationship between design elements and user evaluations is useful to product designers and managers in formulating design strategies"* (Hsu, Chuang & Chang 2000).

The need for this quantitative approach is very well summarized in the following excerpt from Lai, Chang & Chang (2005):

> *The consumer's feeling evoked by a particular product is generally regarded as an abstract or uncontrollable product feature. When developing a product, designers are commonly supplied with a target feeling generated on the basis of market analysis. With this target in mind, the designer then employs his or her subjective experiences to develop the physical product. However, under this approach, there are no target feeling criteria against which to test the success or otherwise of the finished design. Hence, the risk exists that the product is actually a failure before it even enters the market. Therefore, it is clearly necessary to develop scientific methods and procedures to facilitate the estimation, review and improvement of the feeling qualities of a design.*

Some papers – such as Barnes, Lillford (2007), Barnes, Lillford (2009), Barone, Lombardo & Tarantino (2007), Lanzotti, Tarantino (2008) and Nagamachi (2008) – make an effort to introduce Kansei Engineering in early stages of the product design process, as a method to "*reduce the arbitrary decision making that is often applied in the creative product design process*" (Roy, Goatman & Khangura 2009).

Kansei Engineering appears as an attractive area of research where many contributions can be made, as this categorical statement in Lanzotti, Tarantino (2008) suggests: "*The KE approach is still lacking a solid scientific basis*". Statistical tools in Kansei Engineering can help in solving this. The following sections in this chapter will analyze what the reviewed papers discuss about the steps followed in a Kansei Engineering study.

5.4. Spanning the semantic space and the space of properties.

As stated later, not all the reviewed papers follow the KE studies' model shown in Figure 3.5. The reason is that not all studies include the synthesis phase, where the linking between Kansei words and design elements is made. But, definitely, all studies include the description of the semantic space – in the form of Kansei words – and the space of properties – at least, to define a set of stimuli to be evaluated. In this section, both topics will be analyzed:

1. First, I will describe how the Kansei words are commonly – and not so commonly – selected, and how many are used. Some original ideas on how to rank Kansei words according to its importance will also be explained.
2. Second, I will explain how the stimuli (the products) to be rated by participants in the KE study are selected. The set of stimuli is often selected according to some kind of experimental design. I will describe how factors are chosen and, consequently, which experimental designs are used.

5.4.1. How Kansei words are selected, and how many are used.

The first step in defining the semantic space is collecting a long list of possible Kansei words. These Kansei words are almost always extracted from all available sources relevant to the domain of the KE study: magazines, dictionaries, documents, newspapers, manuals, catalogues, websites... This first list of Kansei words can be incredibly long – more than 600 in Jindo, Hirasago & Nagamachi (1995) – somewhere in the middle – with 142 adjectives in Llinares, Page (2008) – or quite short – 39 initial Kansei words in Lanzotti, Tarantino (2008).

Of course, this initial list has to be reduced, as only a few Kansei words can be used finally in the survey. There are not many rules on how to make this reduction, and papers often skip this point giving scarce details. However, I think this is an essential step and must be done carefully. As written in Delin et al. (2007):

> *The (…) task is to reduce the list back down to the 10 or 20 best candidates that can be presented to participants in semantic differential experiments. Given that the process as described so far can generate several thousand words, the accuracy of the reduction process is crucial: it is important to ensure that the more insightful and useful words remain in the list, whereas those that are less descriptive of important qualities (…) are removed. It is also vital to remove words that are obscure or may be unknown to participants, that have more than one relevant interpretation, or that are simply difficult to apply to products of the nature of that being tested. Any such words would introduce error and noise into the data produced from this process*

Besides Delin et al. (2007), some other authors give guidelines on how to perform this initial purge of Kansei words (Barnes et al. 2008): remove adjectives requiring additional context to be understood, remove comparative adjectives, remove non-gradable adjectives… In any case, the difficulty of the process is evident: "*The collection of adjectival words has been straightforward. The question is more how to reduce and reject words*" (Barnes et al. 2004).

After this first purge, there are still too many words, and they must be reduced in some way. Basically two approaches are used (I named these two approaches the qualitative and quantitative approach in Section 4.3.2):

1. Simply doing an affinity diagram (qualitative approach), with the participation of product's experts, and selecting a representing word from each group. This is done, for instance, in Roy, Goatman & Khangura (2009), Chang, Chen (2007), Aktar Demirtas, Anagun & Koksal (2009) and Llinares, Page (2008).
2. Doing a first data collection, and then performing a factor analysis [61] (quantitative approach). This is done, for instance, in Llinares, Page (2007), Nakada (1997), Kuang, Jiang (2009) and Lai et al. (2006). A name is then given to the selected factors depending on the original Kansei words contributing more to each factor. Those are the final Kansei words used for the main data collection phase.

The quantitative approach with a factor analysis has the advantage of being data-based, but the inconvenience is that it requires a previous data collection similar – although probably smaller – to the main survey. This fact of having to collect data twice is not very clearly stated in some papers, although some others give an understandable explanation:

> *The SD evaluations are done twice in the form of evaluation experiment. The subjects check a point among the scaled numbers that they think is appropriate to the pictures presented in front of them on each Kansei word and each sample on the prepared sheets (…) The first step is to evaluate the suitability of many Kansei words collected from various kinds of sources. Factor analysis is introduced for compressing information into a smaller number of synthesized variables and for finding axes of semantic space after this SD evaluation. Kansei words are mapped in the semantic space based on their principal component loading and are grouped together. By factor analysis, Kansei words are summarized from hundreds of words into several scores of*

[61] Sometimes, a principal component analysis (PCA) is done, instead of a factor analysis (FA). PCA and FA are two different techniques, but have many things in common and they are often confounded. I will try to bring some clarity on this topic in Section 5.6.1.

ones. These words are used again in the second SD evaluation which is to analyze the relationship between each word and subject's image about each sample. (Yang, Nagamachi & Lee 1999)

In Section 4.3.2 I proposed the use of a third way to reduce the initial semantic space: start with an affinity diagram and, later, use a cluster analysis to refine the groups extracted from the affinity diagram. This is similar to the proposal of Choi and Jun (2007), where first a hierarchical clustering is done, and later a K-means clustering reallocates the Kansei words to obtain better partitions, but substituting the hierarchical clustering by the affinity diagram. I think this approach is quite reasonable, and in fact rather original, as it is not commonly used. Lanzotti and Tarantino (2008) do perform an affinity diagram and a factor analysis, but separately with the same initial Kansei words, to conclude that both methods gave similar results.

After the initial purge of Kansei words and the reduction made (via the qualitative approach, the quantitative approach or a combination of both), the final semantic space is defined. The final number of Kansei words used for the next steps of the KE study varies among studies. Figure 5.7 shows a dotplot of the number of Kansei words used in the studies from the reviewed papers.

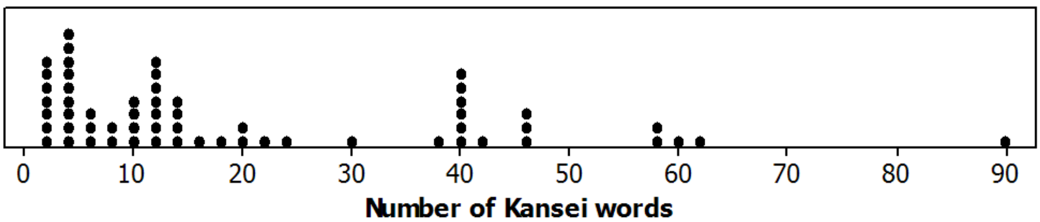


Figure 5.7. Dotplot with the number of Kansei words in KE studies from the reviewed papers.

35% of the studies use 8 or less Kansei words. The median is 12 Kansei words. When lots of Kansei words are used in the study, these words are only used for the data collection. For interpretation of results, not all Kansei words are usually taken as a response separately. On the contrary, a factor analysis or a principal component analysis is performed to identify the main components, a name is given to each component, and only these few components are later linked with the design properties in the synthesis phase. For example, Jindo, Hirasago & Nagamachi (1995) – the study having 90 Kansei words – reduce all 90 words to 4 axes (a design factor, a comfort factor, a refinement factor and a power factor), and refers the results to these four components.

Two interesting ideas for the semantic space

I think there are – at least – two problems with the process of collecting Kansei words and reducing them in the common way I have explained above:

1. The selected words come from people deeply involved in the KE study, but not from the final users. The participants in the study give ratings on a predefined list of Kansei words written by the team in charge of the study, but have no possibility of adding Kansei words that are perhaps more relevant to them than those proposed. Qualitative techniques such as focus groups or personal interviews do not have this problem.
2. All Kansei words in the list are at the same level, with no ranking of importance at all. For example, we ask if subjects find a chair comfortable, modern, and so on. But is it equally important for a chair to give the sensation of being comfortable than of being modern? Perhaps comfortable is more important than modern in a chair…

Some studies introduce a very clever procedure for collecting Kansei words that solves, in part, the first problem stated above (and even the second one). The idea is asking participants to verbally express their perceptions on the product in a descriptive open talk exercise. Those descriptions are later analyzed, and the most frequent adjectives employed can be used as Kansei words. This is done in Petiot, Yannou (2004), Barnes, Lillford (2009) and Ayas, Eklund & Ishihara (2008). In Ayas, Eklund & Ishihara (2008), for example, waiting areas in health centers are evaluated. Kansei words are extracted from personal interviews with patients while they were waiting to see a physician or to have laboratory tests. I think this is a very natural way to extract Kansei words, and if the interview is well done, rich qualitative information can also be obtained as a collateral benefit.

The verbal description of emotions also has the advantage of adding some prioritization to the list of Kansei words: the words more often stated by participants are probably the more important or more relevant ones for them.

Another clever way to solve the second problem of having some prioritization for the Kansei words is proposed in Chen, Chuang (2008). Here, a Kano questionnaire is used to classify each Kansei word as attractive, one-dimensional or must-be quality. The suggestion is focusing on those Kansei words related with attractive quality.

Some papers, such as Barnes, Lillford (2009), Hsu, Chuang & Chang (2000), Llinares, Page (2007) and Aktar Demirtas, Anagun & Koksal (2009), collect an overall liking score for each product. This idea was also suggested in Kensys[62] (Coleman 2011). A regression can be later done using this liking score as dependent variable, and the scores from each Kansei word (averages of each subject's rating on each Kansei word and stimuli) as independent variables. The most important Kansei words are those significant in this regression.

5.4.2. How many factors, how many stimuli and how the design matrix is created

Spanning the space of properties involves two main tasks:

- Factors (usually called items in the context of Kansei Engineering) probably affecting the Kansei words are chosen. Several levels (also called categories) for each factor must be selected. Factors in KE studies are almost always categorical.
- A set of products to be used later in the data collection phase is prepared. These products can be a selection of already existing products, or prototypes set for the study (perhaps not working prototypes, but only visual representations of the products). This set of products is usually arranged according to some criteria (for example, following a factorial design matrix).

In some occasions, when only a location study is done, without linking factors with Kansei words, it is not strictly necessary to define a proper list of factors. However, the suspected main factors are always taken into account – either explicitly or implicitly – as varying these factors guarantees having a selection of products covering a large range of possibilities. For example, flower arrangements are analyzed in Ishihara et al. (2008). Although no formal linking of the semantic space and the space of properties is made, flower arrangements are selected by clearly varying type of flower and colors. In Barnes and Lillford (2009), the stimuli consist of a set of existing beer bottles. Although there is no indication of which factors are being changed to select the set of bottles, one can easily realize that there are clear and dark bottles, and bottles with a straight profile and with a more rounded profile.

Selection of factors

When a complete KE study is done, factors are absolutely needed: the synthesis phase will reveal which factors are significant for each of the Kansei words. How are these factors selected? Obviously, knowledge of the product is important here: "*In a*

[62] Kensys was a project on Kansei Engineering funded by the European Comission under the Fifth Framework Programme (Project Reference: IPS-2001-42075).

practical manner, product design attributes should be selected based on the design practitioners' expert knowledge and design guidelines" (Bahn et al. 2009).

The selection of factors is basically done by technicians, users or a combination of both:

1. When only technicians are used in the selection, a common way to choose the factors is asking experts in a product to write down design properties they think are influential. Later, those experts form a focus group to decide on the final factors (Lai, Lin & Yeh 2005). In Nakada (1997), the initial set of factors are ranked by importance by technicians to only use those in the highest positions.
2. Combining technicians and users as source for the factors' selection is also common. Consider, for example, the procedure in Dahlgaard et al. (2008):

   > *In the chocolate bar example an original set of 14 different product items (properties) were identified and presented to target group customers who rated them by assigning a total of ten points between those items (properties) they personally believed to be the most important ones. The data was used to construct a Pareto-diagram, and the four most important items (properties) were chosen for evaluation: size, shape, colour, and brand. The remaining items (properties) were dismissed.*

   A similar approach is adopted in Lanzotti and Tarantino (2008): 65 initial factors in train interiors are selected; travelers are asked to select the ten most important factors from the list of 65. The most frequently selected factors are the ones that are finally used. This mixed combination of factors selected by technicians but later prioritized by users is also adopted in Bouchard (2009).

3. It is very strange to only have users as source for the factor selection. However, this is an appealing approach: what really attracts attention to users are the relevant factors. A very original way to define the list of factors is explained in Barnes et al. (2008). The studied product here is bottles. Based on the repertory grid technique, a triadic sorting exercise[63] is done: participants are asked to select three bottles from the set and identify a likeness between two bottles and define the contrast with the third. For example, a consumer

[63] The repertory grid is an interviewing technique which uses factor analysis to determine an idiographic measure of personality. It was devised by George Kelly in around 1955 and is based on his Personal Constructs theory of personality. ("Repertory grid" *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=Repertory_grid&oldid=322893481, accessed April 2010). Repertory grids will briefly appear again in Chapter 8 of this dissertation.

might say, "bottles 4 and 9 are similar because they have a circular footprint and bottle 14 is different as the footprint is square". One factor could be then footprint, with circular and square as levels for these factors.

It is not clear from the literature review if it is better to have only technicians to select the factors, or, on the contrary, if users' involvement in this part of the study is desirable. The mixed approach of technicians doing a first initial list, but also involving users in the final selection, seems reasonable.

Number of factors and levels. Design matrix

Figure 5.8 shows a dotplot of the number of factors in KE studies from the reviewed papers. Only studies explicitly defining the number of factors used are considered. Almost 40% of the studies have 4 or less factors. The median is 5.

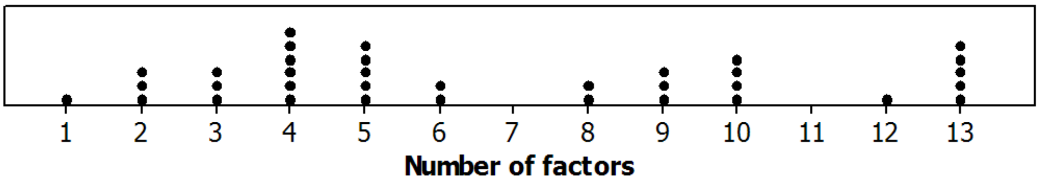


Figure 5.8. Dotplot with the number of factors in KE studies from the reviewed papers.

There are few references in the literature to how the levels for each factor are defined. It seems that whereas factors are selected according to their envisaged significance, the number of levels for each factor is more determined by "statistical limitations". The set of stimuli to be used is sometimes decided according to the design matrix of a full or fractional factorial design. As two-level factorial designs are very popular, factors are often restricted to just two levels when using design matrices coming from factorial designs – this is what is done in Barone, Lombardo & Tarantino (2007), for example. Other times, however, factorial designs are used with some factors having more than two levels (Bahn et al. 2009).

Taguchi orthogonal arrays (OA) are also rather spread. The array that accommodates the desired number of factors and levels is the chosen one: in Ogawa, Nagano & Yukawa (2009), 5 three-level factors and 1 two-level factor are accommodated in a L18 OA; in Chen and Chuang (2008), 9 four-level factors and 1 two-level factor are accommodated in a L32 OA, in Lai, Chang & Chang (2005) a L27 OA is used for 13 three-level factors.

As stated in Section 4.4.2 when talking about the design matrix, there are two different situations regarding the selection of products for the study: when prototypes of the product are created – perhaps simply using a drawing program, as in Chen and Chuang (2008) – or when existing products are used. From the reviewed papers, approximately 80% of them use a selection of already existing products as a set of stimuli, whereas 20% build prototypes (normally images using a drawing program).

The use of a design matrix that determines the combinations of design elements for each product is easier when prototypes are created (they are simply created following the specifications by the matrix). If already existing products are used, it is necessary to find existing products that follow the needed combinations of design elements: this is not always easy to achieve, and sometimes it is impossible, inducing situations of non-orthogonality. Even in these non-orthogonality situations, efforts are made to collect a set of products that are as balanced as possible regarding combinations of factors: "*In order to control validity, the representative product samples were chosen in such a way that a certain product property were presented by at least two different product samples*" (Schütte, Eklund 2005)[64].

As explained in Section 2.4.2, brands can have an impact on the perception of products. To avoid this effect, brands are usually removed from samples when using real products: "*We prepared those test samples by putting each photograph of automobile steering wheels on the market into Macintosh computers using a scanner, and then masking the backgrounds and eliminating corporate marks on the steering wheels by graphic processing of 'Photo Shop', an image processing software.*" (Jindo 1997). This is also done in Delin et al. (2007) and in Mondragon, Company & Vergara (2005).

Number of stimuli

The number of stimuli finally used in KE studies varies from just a few to more than 100. Figure 5.9 shows a dotplot with the number of stimuli from the reviewed papers.

[64] Section 7.2 of this dissertation will make suggestions on how to achieve design matrices suitable for KE studies (and will discuss why this is important).

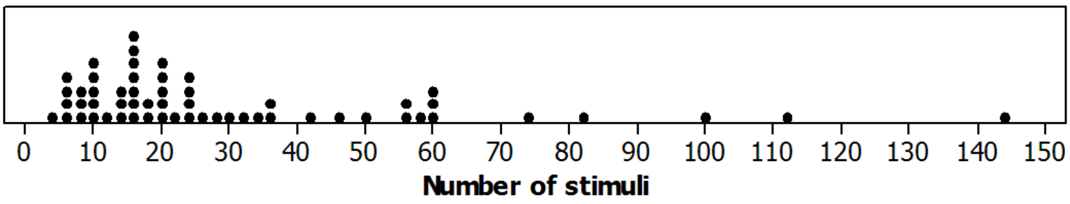


Figure 5.9. Dotplot with the number of stimuli in KE studies from the reviewed papers.

More than 75% of the studies have 36 stimuli or less. Studies with lots of stimuli are special for several reasons. The study with 144 stimuli (Chang, Chen 2007) uses car profiles as stimuli, and these profiles are not rated but ordered in a two steps process[65]. The next studies with more stimuli have more than 100 stimuli in total, but only a few are rated by each participant in the study (consequently, not all stimuli are rated by all participants, and only the average of the ratings is used in the analysis). For example, in Llinares and Page (2007), each participant only rates 3 stimuli, although a total of 112 stimuli are used.

In general, the more factors involved, the more stimuli used. Figure 5.10 (obviously drawn with only the KE studies that state both the number of factors and the number of stimuli) shows this relationship.

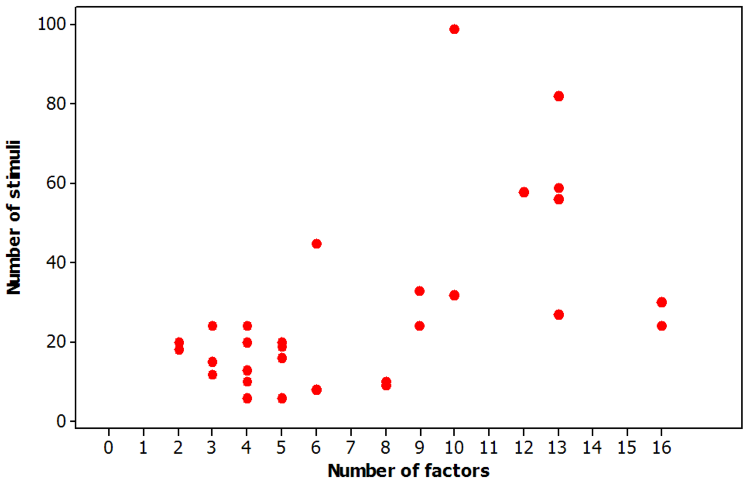


Figure 5.10. Relationship between number of factors and number of stimuli.

[65] More information about this KE study is offered in Section 5.5.4, when talking about how data is collected.

5.5. Kansei Engineering Data Collection

Collecting data is a key point in any KE study. The principle of garbage in, garbage out applies: if data is not reliable, the conclusions might be wrong, no matter the sophistication of the analysis. As stated in Ben Ahmed (2009), *"the precision and reliability of these models essentially depend on the reliability and precision of the data used. Indeed, if the data used are not representative of reality, the reliability of the model may be deceptive"*. Probably the complexity of the data collection process leads to the desire to conduct a pilot study before the main one, as done in Barnes et al. (2004) and Hsu, Chuang & Chang (2000).

In this section, several topics regarding KE data collection will be discussed. Some scholars of Kansei Engineering state that too few participants are usually used in the studies. But to properly affirm this, it is necessary to "use data" and investigate the number of participants in studies reported in the literature. I will present this information in Section 5.5.1. Other topics will be the use of images or real products and how the session is organized (including time needed to collect all data). Finally, the kind of scales typically used will be considered.

5.5.1. How many subjects participate, and how they are selected.

Statisticians often face the question: which sample size do I need for my study? Consequently, one could think that the number of participants necessary for a Kansei Engineering study is a common topic found in papers. However, it is not. Only Aktar Demirtas, Anagun & Koksal (2009) tries to find this sample size with the formula used for estimating proportions with a certain margin of error (its appropriateness in this case is arguable).

Some authors are aware of this methodological lack in Kansei Engineering: *"Careful analysis of the results of the research is needed to clearly define how many participants are needed for statistical significance of the results"* (Barnes, Lillford 2007). But the sentence appears in the future research agenda section of the paper. [66]

Anyway, the amount of participants varies significantly among different KE studies. Figure 5.11 shows the number of participants in KE studies from the reviewed papers: the minimum is 9, the maximum 159 and the median 40. 30% of the studies are done with 24 subjects or less, and only 10% are done with more than 80 subjects.

---

[66] In my opinion, the number of participants needed is not something absolute, but related with the homogeneity of the response inside different emotional groups. I will make some contributions on this topic in Chapters 8 and 9 of this dissertation.

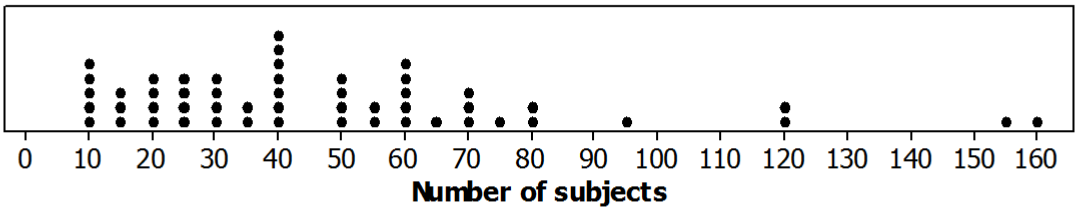


Figure 5.11. Dotplot with the number of subjects in KE studies from the reviewed papers.

In general, participants are selected so *"that they represent the target customer group well"* (Yang, Nagamachi & Lee 1999). To do this, and *"as the product is designed for entire market segments, representative users from diverse backgrounds, such as different age, occupation, and gender, should be considered, and the user distribution along each division must be reasonable."* (Kuang, Jiang 2009). What *reasonable* means is not clearly defined. Normally, when considering all participants as a whole, it is supposed that they are basically homogeneous. Sometimes, individual data is graphically analyzed to check for this homogeneity (Schütte, Eklund 2005). In Petiot and Yannou (2004), one participant (from a total of 11) is even removed because it is considered anomalous[67].

As several KE studies are done for research in an academic atmosphere[68], university students are a huge source of participants for Kansei Engineering studies. For example, all participants in the studies by Barnes et al. (2004), Barone, Lombardo & Tarantino (2007), Chang, Lai & Chang (2006), Yang, Nagamachi & Lee (1999) and Schütte and Eklund (2005) are students.

Comparisons among groups

Many times demographic and socioeconomic data (age, gender, level of studies, etc.) is recorded for each participant. Although this information could be used after the synthesis phase, this is never done. In fact, individual data is normally averaged among participants when starting the synthesis phase, so missing the opportunity to perform analysis at an individual level. Roy, Goatman & Khangura (2009) propose characterizing subjects using a personality categorization indicator such as the 16PF[69].

---

[67] I will explain in Chapter 8 a methodology, first applied to Kansei Engineering in Álvarez (2009), to detect anomalous values in a semi-automatic way. The topic of what an anomalous value in the context of Kansei Engineering means will also be addressed.

[68] It could be that the number of Kansei Engineering studies done at universities is over-represented in the reviewed papers. The reason is simple: academic studies are more frequently published than those done in industry.

[69] The 16PF, or Sixteen Personality Factor Questionnaire, is a multiple choice personality questionnaire developed over several decades research by the psychologist Raymond B. Cattell and his colleagues. More information can be found in "16PF Questionnaire" *Wikipedia,*

The paper does the analysis in a completely qualitative way. In my opinion, using a personality test is a good idea, as probably personality traits are more important when segmenting people according to their emotional reaction to products than the common demographic and socioeconomic data[70].

Although an *a posteriori* segmentation of participants based on their responses is never done, several studies try to compare the emotional reaction of *a priori* defined groups. These groups are studied separately because of the suspicion that they are different. A popular example is comparing between designers and users (Bahn et al. 2009, Chang, Chen 2007, Hsu, Chuang & Chang 2000, Jiao, Zhang & Helander 2006). The case study shown in Hsu, Chuang & Chang (2000) is especially interesting, as it studies variability in the ratings in both groups: users present less variability than designers, because they tend to rate in the middle of the scale; on the contrary, designers have much more extreme opinions, thus having a higher range of ratings.

In Huang et al. (2009), the study is done on four different age groups of people: children at about 10, students at about 20, adults from 10-50 and elderly people from 60-70. In Ogawa, Nagano & Yukawa (2009) the perception of a massage chair is compared between people with stiff shoulders and relax-conscious persons. In Mondragon, Company & Vergara (2005), machining centers are evaluated by production managers, university lecturers in manufacturing engineering and by machine tool operators. Pearce and Coleman (2008) present a study on shoes made in two countries (UK and Spain). Besides comparing the results between both countries, this study also evaluates differences in response between genders and between participants viewing real products or reproductions.

An interesting and original proposal related with groups can be found in Barnes and Lillford (2009). The idea is asking each participant for an overall liking score for each product. A cluster analysis is then performed on these overall liking scores to determine if the analysis can be defined around different user groups.

5.5.2. The use of images vs. real products

Either photographs of the stimuli or real products can be presented in the data collection phase of a KE study.

The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=16PF_Questionnaire&oldid=355890033, accessed April 2010.

[70] Chapter 8 of this dissertation will be devoted to how to segment participants based on their emotional reactions to the product being studied, and how to characterize these segments.

When photos are used, only the Kansei elicited by the sense of sight is considered. This is obviously limiting, in the sense that the affective channel width is somehow narrow. But the widespread use of photos for KE studies is normal, as sight is an important sense – if not the most - in perceiving the Kansei conveyed by a product and getting photos nowadays is cheap and easy. According to Nakada (1997), *"human beings acquire 80% of their information through their sense of sight, which has a great impact on impression"*. This same author states that *"evaluation by slides has been proven to produce similar results to evaluations carried out with the actual object"*. Approximately 70% of the reviewed papers use images for presenting the stimuli.

When KE studies are performed through the Internet, products are necessarily displayed using images (or videos, in some very rare occasions). In some studies done with web-based questionnaires, written descriptions of the products complement the images (Roy, Goatman & Khangura 2009). Of course, this has the problem of making the experience less emotional and more rational. Sometimes, participants are required to make an effort to stretch their imagination in order to rate the stimuli, as in Llinares and Page (2008):

> *The set of stimuli used to develop the field study consisted of a total of 74 images of the different neighbourhoods in the city of Valencia. In differential semantics, interviewee perceptions are obtained by showing them a sample of the product to be evaluated (a table, a chair, a telephone, . . .). In this case, given the difficulty of taking the people to the neighbourhoods to be evaluated, stimuli or images of the areas the interviewees had to recognise were prepared. The aim is to collect citizens' perceptions and opinions on the different neighbourhoods as a place to live. The subject must observe the stimuli, recognise the neighbourhood and mentally reproduce what he thinks of it, its atmosphere, noise, colour, smell,… and therefore the stimulus must be clear enough to evoke the neighbourhood for the interviewees*

Another solution for rating the neighborhoods in Llinares and Page (2008) would have been physically visiting them, so using "the real product". This is the approach in Ayas, Eklund & Ishihara (2008), where the "product" is waiting areas in health centers (however, each person only rates one health center, the one he or she is visiting).

Real products are used in approximately 30% of the reviewed papers. In this case participants are usually allowed not only to look at the prototypes, but also to interact with them, touching and playing (Ogawa, Nagano & Yukawa 2009, Barnes, Lillford 2007, Hsu, Chuang & Chang 2000, Petiot, Yannou 2004). Other senses, such as the sense of touch, are thus involved in this kink of KE studies. Sometimes, only the sense of touch is involved: Barnes et al. (2004) present a KE study for assessing the emotions conveyed by the sliding contact of a finger over rough glass surfaces (intended for cosmetics packaging).

The effort to use imagination sometimes requested from participants in the study when using images can also be reproduced when using real products, as in Jindo, Hirasago & Nagamachi (1995):

> *The subjects walked around the chairs to view them thoroughly and then evaluated them on the basis of the impressions they received. In making their evaluations, the subjects were asked to disregard as much as possible their impression of the color of the chair.*

A mixed option combining images and real representation of the products is using a more immersive atmosphere, like some kind of virtual reality (Barone, Lombardo & Tarantino 2007)[71].

5.5.3. How the session is conducted

The way the data collection session is conducted depends on the fact of having participants physically in a room or over the Internet. Very few papers give recommendations on what to do when subjects answer questions on a web-based questionnaire. Roy, Goatman & Khangura (2009) state that *"a questionnaire like this is that it must take a very short time and be enjoyable and undemanding to answer"*. It seems a reasonable piece of advice. When participants are present in a room, the recommendation is to conduct the experiment in a place that is as aseptic as possible, like *"a pleasant but neutrally-furnished room"* (Chen et al. 2009).

Sometimes, like in Roy, Goatman & Khangura (2009), a written description of the Kansei words is offered to assure that everybody understands the same before starting. Almost all studies propose a process for data collection that tries to guarantee that order of presentation will not affect the results (randomization is a common suggestion). For example, this is the proposal done in Nakada (1997):

> *Using Kansei techniques an evaluation was conducted on 56 types of vehicle operator seats using slides. First a dummy run of two or three slides was shown to accustom the evaluators to the images of the vehicles themselves. In order to avoid the seriate affect of tiredness and a decrease in strictness of evaluations, subjects were divided into 6 groups of 7 or 8 people, the groups were given breaks, and the seat evaluation sequence was altered for each group.*

---

[71] I think the use of virtual reality for presenting the stimuli will increase in the near future: I will explain a bit more on this topic in Chapter 9.

Very few papers state the time needed to record the data. In my opinion, this information is very relevant, as the more time allowed the more stimuli that can be rated, but at the price of having a more demanding experiment. It seems that there are differences in the time considered reasonable in European papers and Asian papers. Barnes and Lillford (2007), talking about a study done in the UK, state:

> *Our experience has shown that, for robust results, we have about 30 min to brief an individual and then ask him or her to respond to our stimuli. Assuming each response takes about 1 min and a 10-min briefing is required, this gives us a maximum of 20 different stimuli which have to cover combinations of the range of experimental variables under investigation.*

The same authors, in Barnes et al. (2008), assert that *"the length of the survey and its repetitive tasks can be extremely demanding on participants and lead to questions about the repeatability and robustness of the results"*, so a 30 minutes maximum is recommended. On the contrary, Lokman, Noor & Nagamachi (2009), a KE study done in Malaysia, report spending 2 hours per Kansei session.

Several "tricks" are also used to reduce the time needed to perform the data collection. In Llinares and Page (2008), not everybody rates every stimuli: *"It was considered that an interviewee could reply to a maximum of 3 questionnaires before losing interest"*. Using a verbal description of emotions for each stimuli, instead of rating predefined Kansei words, as in Ayas, Eklund & Ishihara (2008), can also reduce time needed: *"The free association method used in this study may help reducing long data collection times in Kansei engineering studies and reduce the time for participants"*.

5.5.4. Kinds of scales used to rate the stimuli

Rating Kansei words is, by far, the most commonly used method for collecting data in KE studies: more than 95% of the reviewed papers. From these papers, 46% of them use 5-point scales, 44% use 7-point scales and 10% use 9-point scales. So both 5-point and 7-point scales are equally frequent. Visual analogue scales (VAS) are very uncommon: it is only used in Bahn et al. (2009) and Schütte and Eklund (2005).

In fact, Schütte and Eklund (2005) is a remarkable paper because it is the only one comparing the use of VAS scales with 7-point scales. The conclusion is that *"there is no significant difference between data from 7-point and VAS scales. Therefore, Kansei data gained from a 7-point scale can be treated in the similar way as data from VAS scales allowing corresponding statistical evaluations"*.

Another key decision regarding scales is the fact of using Likert scales or semantic differential scales (see Section 3.4.1). This selection affects the list of Kansei words, as

Likert scales are unipolar (they demand just one Kansei word), but semantic differential scales are bipolar (they demand Kansei words in pairs). From the reviewed papers, 37% use semantic differential scales and 63% use Likert scales.

Normally, 7-point scales are labeled with numbers 1-2-3-4-5-6-7. A variation is proposed in Mondragon, Company & Vergara (2005), where the sequence used is 3-2-1-0-1-2-3: *"To prevent prejudgments, only positive scales were used, with a range from 0 to 3 for each of the opposite descriptors belonging to the same pair"*. The scale is later converted to numbers from -3 to 3 for analyzing the results.

Kuang and Jiang (2009) study mobile phones, and not only use a 7-point semantic differential scale, but also a ranking of all 15 mobile phones done by each participant. It is the only reviewed paper that uses rankings.

Rating stimuli is adequate when the number of stimuli is not very high (something common: more than 75% of the studies have 36 stimuli or less, see Figure 5.9). However, it can be too long and challenging when there are lots of stimuli. In Chang and Chen (2007), a rate is given to 144 stimuli, but each of the 32 participants used the following procedure for each of the 6 Kansei words:

> *Taking the first image word, each subject divided the 144 automotive profile samples into three groups, i.e. is low (L), medium (M) and high (H), in accordance with his or her intuitive perception of the degree to which the image word accurately described the feelings induced by the profile. Each of the three groups was then further divided into three sub-groups. In this way, the 144 automotive profile samples were divided into a total of nine groups. Each automotive profile was then assigned a score from one to nine (LL to HH) according to the group to which it belonged.*

Finally, each stimulus has a rating on a 9-point scale for each Kansei word, so all common techniques of analysis in the synthesis phase can be used.

5.5.5. Capturing data through physiological body reactions

In Section 3.3.1 I introduced how to measure the Kansei using physiological body reactions. This approach of collecting data for KE studies is only followed in one of the reviewed papers (Shimizu et al. 2004). The domain of this Kansei study is socks and belts. An electroencephalography (EEG) is done for different degrees of pressure on the abdomen by the waist belts. The intensity of alpha waves is also evaluated. For the socks, the muscular activity of the lower leg during walking and sole pressure when wearing different kinds of socks is captured.

This dissertation is focused on the usual way of capturing data for Kansei Engineering studies: self-reporting the Kansei through Kansei words. However, I think the use of physiological measures is a promising area of research, especially for those products that physically interact with our body. Ogawa, Nagano & Yukawa (2009), in a paper describing a case study with massage chairs done with Kansei words – the common way – states: "*We believe that it becomes increasingly necessary to quantitatively define the effect and efficacy of massage chairs and also make evaluations using the physiological indicators in the future* "

5.6. Kansei Engineering Data Analysis

Once the data collection has finished, it is time to analyze it and reach conclusions. Two "families of tools" are mainly used:

1. Tools for a descriptive analysis of the data: lots of information can be extracted simply by using clever graphs, such as profiles or radar plots of the average ratings of products or Kansei words. I have presented some of these graphs in Section 4.7. I also include under this umbrella principal component analysis, as I consider it a descriptive tool.
2. Tools for the synthesis phase: a key characteristic of classical KE studies is linking the semantic space with the space of properties. This is done in the synthesis phase. Tools used in this phase essentially fall under two categories: statistical tools (some kind of regression analysis) or automatic learning tools (mainly neural networks).

When just a location study (placing products in a two-dimensional map) is performed, descriptive tools are often the only ones used. When there is an attempt to find the relationship between Kansei words and design elements, either statistical or automatic learning tools necessarily make an appearance. Many times, studies with a proper synthesis phase also perform descriptive analysis of the data.

5.6.1. Tools for a descriptive analysis of the data

Profiles or radar plots are convenient tools to graphically summarize multivariate data (Manly 2005). I have exemplified their use in Section 4.7 of this dissertation. Although one could expect a massive use of these graphs in Kansei Engineering, they are not so common[72]. When using profiles, there are two options:

[72] In my opinion, graphs are a very suitable tool for presenting results in Kansei Engineering studies (and in any data-based study). I will suggest the use of several graphs to present the results from a KE study in the next chapters.

1. Having a plot for each product, with Kansei words in the horizontal axis and the average rating in the vertical axis, such as the picture in Figure 5.12. This is done in Bouchard (2009), Schütte and Eklund (2005), Nakada (1997) and Mondragon, Company & Vergara (2005).
2. Having a plot for each Kansei word, with products (or levels of a significant factor) in the horizontal axis, and the average rating in the vertical axis. This is done in van Lottum, Pearce & Coleman (2006) and Bahn et al. (2009).

Of course, both options can also be used in the same study. Radar plots are equivalent to profiles, with each Kansei word in one radius of the radar plot (Llinares, Page 2007).

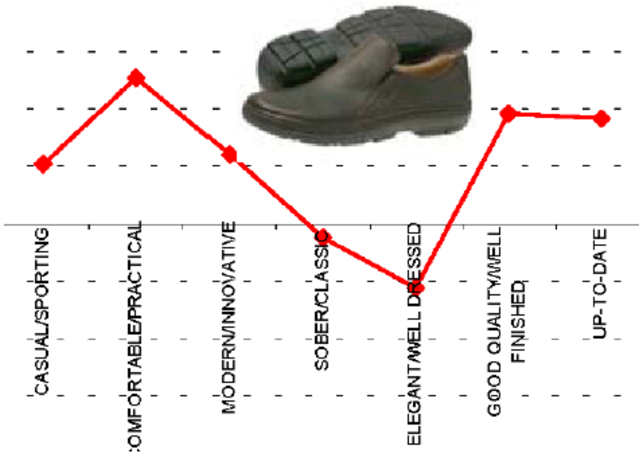


Figure 5.12. A profile of Kansei words for a shoe, reproduced from Bouchard (2009)

Principal component analysis or factor analysis is extensively used in Kansei Engineering studies (Chen et al. 2009, Barnes, Lillford 2007, Barnes et al. 2004, Barnes et al. 2008, Choi, Jun 2007, Delin et al. 2007, Ishihara et al. 2008, Jindo, Hirasago & Nagamachi 1995, Llinares, Page 2007, Llinares, Page 2008, Petiot, Yannou 2004, Barnes, Lillford 2009, Hsu, Chuang & Chang 2000, Huang et al. 2009). This is not surprising, as its visual output can summarize results in a very appealing way.

These multivariate techniques are basically used in KE studies for two purposes:

1. For reducing the dimensionality of the semantic space, that is, for grouping Kansei words, once the main data collection is finished. For example, in Llinares and Page (2008), instead of working with all 60 original Kansei words, the first 11 factors capturing almost all the variability of the original dataset are used. In this case, these factors are given a name (or at least a description composed by those Kansei words that contribute the most to each factor). This approach facilitates reaching conclusions.

2. For locating products (or Kansei words, or both) in the semantic space. For example, Figure 5.13, taken from a case study described in Ishihara et al. (2008), locates all products (in this case, flower arrangements) in the two-dimensional semantic space. This second use of principal component analysis and factor analysis is by far the most widespread.

I think there is some confusion – not only found in KE studies – between factor analysis and principal component analysis. So sometimes a factor analysis is – strictly speaking – performed, although it is described as principal component analysis in the text (Barnes et al. 2004, Delin et al. 2007). Next section will try to clarify the confusion.

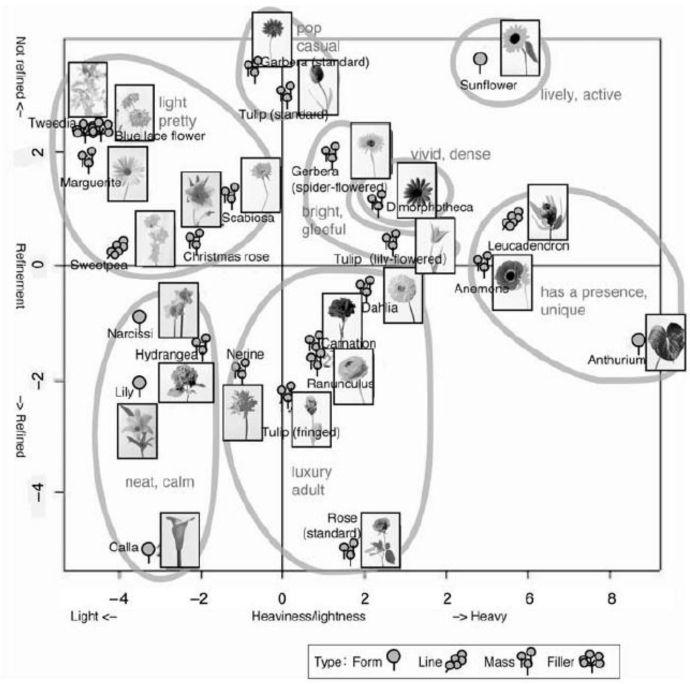


Figure 5.13. Flower arrangements located in the semantic space, reproduced from Ishihara et al. (2008)

Bringing clarity to principal component analysis and factor analysis

One cause for this confusion could be the fact that the software package SPSS – one of the most commonly used in KE studies – only has factor analysis as a menu, but when conducting the analysis with all options by default (principal components extraction, no rotation) it is in fact doing a principal component analysis[73].

A principal component analysis (PCA) is a purely descriptive method. It helps for understanding the underlying data structure and reducing the number of original

[73] On the contrary, Minitab (another software also used in KE studies) clearly distinguishes both methods in two separate options under the Multivariate menu: Principal Components and Factor Analysis.

variables into a small number of uncorrelated variables (called principal components). Principal components are extracted with an eigenanalysis of the correlation matrix. There are several criteria to decide on the number of principal components (PC) finally used: only selecting those PC with eigenvalues greater than 1 (so no PC is considered that accounts for less variability than a single original variable), selecting PCs until a certain percentage of variance is accumulated, or using a scree plot to only consider the most important PCs (like in a Pareto chart).

On the other hand, factor analysis (FA), in contrast to principal component analysis, is not purely descriptive, but a model is underneath. The aim is also to summarize the data structure in fewer dimensions, but the focus is now in identifying the underlying factors (latent variables) that might exist. The idea is that the whole variability of the data can be divided into a common part – explained by the factors – and a unique part – lost when only using the extracted factors. Sometimes some psychological or sociological theory is beneath the procedure, and the number of factors to be extracted comes from this theory[74]. Other times, the number of factors is decided with the same procedures employed in PCA.

One method of factor extraction in FA is using principal components. In this case, factors are the same as in a principal component analysis. If only two factors are selected, these two factors are exactly the same as the first two principal components. But other methods of factor extraction are available; a common one is maximum likelihood. In this case, if only two factors are selected, they will be different from the first two principal components (and different from selecting three factors or just one).

Especially in FA, there is an attempt to give a name to each of the extracted factors. This is usually easier achieved when factors are rotated. In the rotation of factors, the reference axes of the factors are turned about the origin until some other position is reached. The rotation can be orthogonal (in which the axes are maintained at 90 degrees) or oblique (when the former constriction is removed). A popular rotation method is the varimax rotation, which maximizes the sum of variances of the loadings. The rotation of axes is something traditionally done in FA and not in PCA. However, principal components can also be rotated, and it is thus sometimes done.

When PCA or FA is used in Kansei Engineering studies, the purpose is basically locating products and Kansei words in a pair of axes (as I have done in Figure 4.11).

[74] For example, if a psychological theory maintains that there is just a latent variable – general intelligence – that explains the performance of children in an intelligence test, only one factor should be extracted using points from the whole set of questions. On the contrary, if the theory supports the idea that two latent variables are needed – say, mathematical and linguistic skills – two factors should be extracted.

Thus only two factors are needed. The amount of variability captured by these two factors is not always stated (we hope it is enough to correctly summarize all the variability of the data). Sometimes, three factors are extracted, and a 3D representation is then used. But never more, because the visual representation of products and Kansei words would then be not direct, and surely more complicated.

In KE studies, there is usually no theory that supports the number of factors that must be extracted. PCA and FA are used in a descriptive way, mainly to locate products and Kansei words (sometimes in a biplot) and visually decide about similarities and dissimilarities. Thus, in my opinion, a PCA is more appropriate – and enough – for the intended purposes. If there is the desire to give a name to each axis – something that is not always easy and, in fact, not needed – the axes can be rotated to facilitate this task.

Anyway, empirical studies have shown that results from principal component analysis and from factor analysis are similar in many cases (Velicer, Jackson 1990).

5.6.2. Tools for the synthesis phase

About 25% of the reviewed papers present studies – even though defining themselves as Kansei Engineering studies – that are only "location studies" (products and Kansei words are placed in the semantic space). Surely, authors of these studies are aware of this fact. Mondragon, Company & Vergara (2005), in a case study on machine tools (MC), state:

> *The next step should be to determine which design elements are responsible for those affective dimensions of the MC. Working groups with designers, managers and users should perform MCs' morphological analysis and extract the design elements taking part on them. By considering these elements as independent variables and the semantic attributes as dependent ones (…) the relationships between design elements and semantics could be determined from the same data from the expanded study.*

The synthesis phase is a key part of KE studies. Unlike other steps in the KE model, where basically the same tools are used in every study, the synthesis phase is the step with the most variety in the reviewed papers. Several tools are available for linking the Kansei words and the products' design properties, and it is not clear which ones are better in which situation. Basically, two kinds of tools are used: statistical tools (regression models) and automatic learning tools.

Figure 5.14 shows a Pareto chart of tools used in the synthesis phase. Several remarks can be made when looking at this graph:

- QT1 (quantification theory type I) is the most popular tool in the synthesis phase. It is especially common in papers from Asian countries.
- Summing up studies using MLR (multiple linear regression) and OLR (ordinal logistic regression), 20% of the reviewed papers use these techniques. These papers are basically from European countries.
- As QT1 is nothing more than a multiple linear regression (with some special characteristics, as it will be explained in Section 6.2), all these three techniques – QT1, MLR and OLR – are statistical regression models. Using regression models is by far the most common way to complete the synthesis phase: more than 60% of the reviewed papers follow this approach.
- Especially NN (neural networks), but also RS (rough sets), two tools from the automatic learning approach, are used in approximately 20% of the reviewed papers.
- Other tools used with much fewer frequencies are fuzzy logic (Lin, Lai & Yeh 2007, Ma, Nakamori 2007), Bayesian networks (Ben Ahmed 2009) or partial least squares (PLS) (Lokman, Noor & Nagamachi 2009).

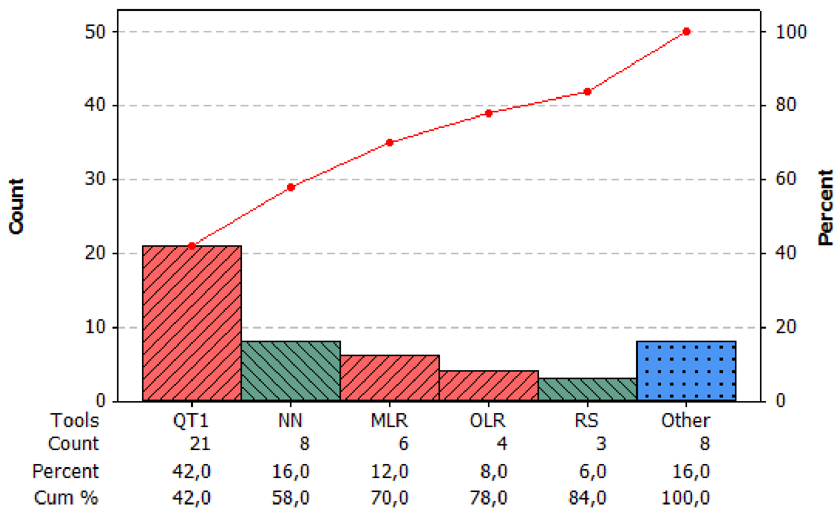


Figure 5.14. Pareto chart of tools used for the synthesis phase in the reviewed papers.

QT1 is used in many papers in a standard way (Bahn et al. 2009, Chen, Chuang 2008, Dahlgaard et al. 2008, Fukushima et al. 1995, Jindo 1997, Matsubara, Nagamachi 1997, Nagamachi 1995b, Nakada 1997, Schütte, Eklund 2005 Tanoue, Ishizaka & Nagamachi 1997, Yang, Nagamachi & Lee 1999). The significance of factors in QT1 is checked looking at the partial correlation coefficient (PCC). Usually, factors with a PCC greater than 0.7 are considered significant (Schütte et al. 2004). However, there is a tendency to take all factors into account, regardless of its PCC (Jindo 1997).

A very interesting approach with QT1 is the one proposed in Ogawa, Nagano & Yukawa (2009). Here, category scores (which are, basically, the coefficients in the regression) are shown with its confidence intervals. In fact, Yang, Nagamachi & Lee (1999) correctly state that an F-test can be done for the whole regression, and t-test for category scores (although none of the reviewed papers make use of them). p-values are never calculated in QT1. Despite this tendency to consider that all factors affect the response in QT1, the fact that not all of them are significant is clear in some papers, such as Lai, Chang & Chang (2005):

> *Although establishing the optimal settings facilitates the design of a car profile which closely matches the target feeling, it is known that some factors are of high influence, while others are of lesser significance. The purpose of the improvement stage (…) is not to renovate all the design factors, but simply to redesign those factors which have a significant influence upon the feeling quality. In other words, the intention is to obtain the greatest improvement in feeling quality through the minimum of redesign activity.*

Some papers perform a common multiple linear regression for the synthesis phase. In Chen and Chang (2009), the variance inflation factor (VIF) is used to check collinearity of the factors. In this paper, only significant factors are considered: they are selected using a stepwise procedure. Some others papers follow the typical procedure from conjoint analysis: calculating main effects (called utilities or part-worths in the context of conjoint analysis) (Hsu, Chuang & Chang 2000, Jiao, Zhang & Helander 2006).

When QT1 or multiple linear regressions are used, data is summarized using the average of all participants in the study for each stimuli and Kansei word. Of course, this has a huge limitation, as stated by Schütte and Eklund (2005):

> *The data ready for model building is typically a three dimensional data matrix where concepts (product samples) and scales (Kansei words) span a two-dimensional plane whereas subjects represent the third dimension (…) However, most statistical methods require two-dimensional data. Consequently, the number of dimensions is usually reduced by calculating the average value over the subjects (…) However, doing this means a crucial loss of information, since the variation in the participants' opinions vanishes.*

An ordinal logistic regression allows the use of individual data (ratings on a 5-point or 7-point scale) coming from each participant. Barone, Lombardo & Tarantino (2007), Lanzotti and Tarantino (2008), Bouchard (2009) and Aktar Demirtas, Anagun & Koksal (2009) use an ordinal logistic regression in the synthesis phase. Aktar Demirtas, Anagun & Koksal (2009) also perform a multiple linear regression with the same data. By the way, this last paper is the only one that looks at the residuals of the regression to verify the assumptions of the regression model: in general, model checking is not a very common practice in KE studies…

Barone, Lombardo & Tarantino (2007) propose a very original addition to the analysis of results in the synthesis phase. When collecting data, respondent choice time is also measured: the less time the subjects need to give an opinion – a rating – on a particular product, the more clear it is that the product elicits that emotion. Later, when analyzing the data, an attribute importance weight is calculated and incorporated into a weighted ordinal logistic regression.

Rough sets, an automatic learning technique, has been recently introduced as a tool for the synthesis phase by Tatsuo Nishino (Nishino, Nagamachi & Tanaka 2006, Nishino et al. 2007, Nishino et al. 2008). Nagamachi (2008) states: *"Recently, it became clear that Rough Sets Theory may be better to find decision rules fit to the designers' kansei thinking"*. The output from rough sets is very different from the output given by regression analysis: the design elements are also linked with Kansei words, but in the form of rules. According to Zhai, Khoo & Zhong (2009a), *"the design elements are the condition attributes and the scores attained by Kansei words are the decision attributes. Therefore, the Kansei engineering problem has been transformed into a multi-criteria decision-making issue, which can be addressed using advanced knowledge discovery tools such as rough set theory"*. Ayas, Eklund & Ishihara (2008) show a case study using rough sets. Zhai, Khoo & Zhong (2009a, 2009b) also present an example, but focus the study more on offering some methodological suggestions[75].

Non-linearity (interactions and quadratic terms) in the synthesis phase

A popular critique to QT1 in particular, and to regression models used in KE studies in general, is that they cannot accommodate non-linearities. The following excerpt from Ishihara et al. (1995) expresses this concern:

> *Weiss noted that the linear regression method assumes that all predictors relate linearly to all other predictors. He mentioned this as "If the relationship among predictors deviates from a straight line relationship, that nonlinearity (interaction) will also lower the accuracy of the predictions". Thus, if there is interaction between items or categories, then the deletion or combination of these are needed (...) There are many interactions between items in actual design components. For example, in clothing, one usually does not put on a necktie with a collarless blouse.*

---

[75] In my opinion, rough sets are an interesting approach to the synthesis phase of KE studies, and can at least complement the conclusions extracted from statistical tools. Chapter 6 includes a long discussion on the use of rough sets in Kansei Engineering, with several methodological suggestions.

Curiously enough, non-linearities basically mean interactions here, more than quadratic terms. Of course, interactions can be estimated using regression models if the design matrix allows it: *"if a linear regression fit is unsatisfactory, quadratic or higher-order polynomial regression can be employed to consider interactions among individual design parameters"* (Kuang, Jiang 2009). This paper proposes a central composite design using 3 factors for being able to estimate both interactions and quadratic terms. However, this is not very common in the reviewed papers. On the contrary, neural networks are introduced as a way to include non-linearities in the model (Chang, Chen 2007, Huang et al. 2009, Ishihara et al. 1995, 1997, Lai, Lin & Yeh 2005). Sometimes, both neural networks and QT1 are used (Lai et al. 2006, Nagamachi 2002b).

When using neural networks (NN), a back propagation neural network is commonly employed (Chang, Chen 2007). NN are used more for predicting the Kansei elicited by products than for understanding how the factors affect each Kansei: *"NNs are well suited to formulate the product design process for matching the product form (the input) to the consumer's perception of product image (the output), which is often a black box and cannot be precisely described"* (Lai, Lin & Yeh 2005).

Papers using neural networks often implement some kind of user interface to predict the Kansei. For example, Chang and Chen (2007) use a NN trained with 144 vehicle pictures to predict the consumer perception of an automotive profile expressed in the form of a numerical definition: the profile is defined by 92 points, that can be "moved" on the screen thanks to a program in Visual Basic. Each of the 92 points is an input neuron in the NN, and each Kansei word an output neuron. When the designer interactively "moves" a point, the elicited Kansei is shown on the screen. This use of NN in Kansei Engineering is, in my opinion, promising and appealing. However, NN have the problem of overfitting and, as stated in Zhai, Khoo & Zhong (2009a), *"these methods are often transparent to designers and consumers (…) Moreover, compared with the widely used statistical analysis, they lack formal frameworks when employed in Kansei Engineering applications."* [76]

The idea of taking interactions among factors into consideration is present also in KE applied to services, such as looking at the emotions conveyed by waiting areas in Swedish health centers (Ayas, Eklund 2008). I have the hunch that interactions in KE studies with services can be even more important than in KE studies with products, and thus must not be undervalued.

---

[76] More on neural networks and Kansei Engineering can be found in Section 6.1.2 of this dissertation.

Variability in KE datasets

I consider variability among participants in a KE study a very relevant issue. Intuitively, it is easy to imagine that the emotions elicited by the same single product can be very different depending on the person. Several papers include sentences like "*a great variability in the emotional reaction appeared*" (Bouchard 2009). Mitsuo Nagamachi himself is conscious of this fact when he lists "*how to treat the individual differences of Kansei*" as a problem still to be solved in his seminal paper "Kansei Engineering - a New Ergonomic Consumer-Oriented Technology for Product Development" (Nagamachi 1995a). Sometimes, the variability seems so high that the authors show doubts about how to reach conclusions: "*It would (…) be questionable whether results obtained with the inherent variance of an open web-based survey may be considered to provide data suitable for valid scientific processing*" (Roy, Goatman & Khangura 2009).

Nothing is usually done to deal with this variability. A common recommendation is having a group of participants as homogeneous as possible, hoping that this will keep variability to a minimum:

> *Research (…) has shown that when the consumers' characteristics are more uniform, their evaluation responses are likely to be broadly similar. Therefore, maintaining a consistency of consumers' characteristics is an important aspect of marketing.* (Lai, Chang & Chang 2005)

Jiao, Zhang & Helander (2006) express the same idea: "*Within the same segment, customer affective needs and rule patterns are similar*". This homogeneity of participants is almost always supposed, but not proven. Sometimes, but very rarely, the response distributions are examined to judge the level of consensus. When double peaks are observed, this is an indication of disagreement in the elicited emotions. Barnes and Lillford (2009) proposal is then analyzing both groups separately. But clear indications on how to do this are never offered. Schütte et al. (2004) also suggest the previous analysis of the responses before analyzing the data.

The common procedure in Kansei Engineering studies is averaging data among participants, thus losing the variability among subjects. In my opinion, the first step for taking variability among subjects into account is not averaging their ratings. The use of an ordinal logistic regression, as done in Barone, Lombardo & Tarantino (2007), Lanzotti and Tarantino (2008), Bouchard (2009) and Aktar Demirtas, Anagun & Koksal (2009), allows this, as the response is directly the ratings done by participants.

One step further is explicitly considering not only means, but also variances:

> *Previous studies on Kansei engineering and aesthetics used questionnaire-collected data to examine customer subjective evaluations based on a mean scale rating. However, the evaluation of aesthetics is subjective and highly individualistic. Aesthetics evaluation based solely on mean scale ratings, without considering variation in customer evaluations, is not appropriate. (…) The robust design approach focuses on bringing the mean closer to the desired target and simultaneously reducing quality variation. This design may be successfully used in subjective quality management.* (Chen, Chuang 2008)

This robust design approach – rarely found in the reviewed papers – is done following the original proposals by Genichi Taguchi, using Taguchi's orthogonal arrays as design matrices and analyzing the collected data through the signal-to-noise ratio (Chen, Chuang 2008). Some papers, such as Lai, Chang & Chang (2005), explicitly define the noise factors. In this paper, noise factors are four psychological or behavioral individualized characteristics (involvement, personal trait, peer relation, and social support; each one with 3 levels). Each one of these psychological characteristics is measured with standard personality questionnaires documented in psychological literature. The inner array is a L27 OA (13 three-level control factors), whereas the outer array is an L9 (4 three-level noise factors). The justification in the paper for performing a robust design approach is the following:

> *Analysts frequently employ demographic characteristics to segment the total market into particular consumer groups comprising individuals with common characteristics. Powerful psychological or behavioral individualized characteristics are generally neglected since they tend to be very difficult to investigate reliably. However, the influences of such characteristics are important since they represent uncontrollable factors and may introduce significant variances into the feeling evaluations of a product. If it is infeasible to exclude the influence of such uncontrollable factors completely, then it is clearly prudent to take steps to at least reduce their influence.*

This last excerpt also prompts an appealing idea: it would be interesting to segment people not only by demographic characteristics, but also according to psychological characteristics. I think the idea of finding segments of people can be further explored[77]. More methodological advice on how to deal with variability in KE studies is needed. Some authors are aware of this need, but they propose this research for a later moment: "*In the future, we would like to challenge research issues targeting the individual variability taking one step further from the research on average preference*". (Ogawa, Nagano & Yukawa 2009)

---

[77] In fact, the first part of Chapter 8 of this dissertation is devoted to finding groups of people with the same emotional reaction on a product, from the data collected in a KE study.

5.7. Test of validity and presentation of results

The last steps in the model for conducting a Kansei Engineering study is checking to which degree our results are correct doing some confirmatory experiments, and finally presenting our conclusions in a useful way.

5.7.1. Confirmatory experiments

Some Kansei engineering studies do – and report they have done – confirmatory experiments, but not many of them. Among those explaining the confirmatory experiments one can find Roy, Goatman & Khangura (2009), Chang and Chen (2007), Chen and Chuang (2008), Lai et al. (2006), Lai, Chang & Chang (2005), Lanzotti and Tarantino (2008) and Aktar Demirtas, Anagun & Koksal (2009).

Sometimes, although the confirmatory experiments are not done, an explanation is given for not having done that:

> *One question arising was whether it would be feasible to evaluate the resulting concepts and check whether the impact was as intended. Adding such a step into the Kansei Engineering process would offer the possibility of verifying the final result. If it matches the intended feeling it can be released for production, if not, there can be an iterative process detecting the problem, fixing it and coming up with a new suggestion. Such a step was not done in this study since rocker switches are difficult to rate only visually. A mock-up or model is needed which was not produced here.* (Schütte, Eklund 2005)

Confirmatory experiments are usually done adding one or more new stimuli to be rated by the same group of participants of the main data collection. These new stimuli are obviously created following the pieces of advice extracted from the synthesis phase. The new stimuli are presented in the same way as the main set of products was presented.

The confirmatory experiments almost always confirm the conclusions from the synthesis phase. An interesting exception is Lai, Chang & Chang (2005), where the results are not as good as expected. The authors suggest the fact of not having considered interactions in the analysis of data as a possible explanation: "*The interactions among profile factors could represent influential elements of the feeling difference, but are not considered in the present experiment*".

Chen and Chang (2009) propose an exciting experience once the synthesis phase is concluded. In this paper, knives are the studied product. A designer is asked to create two new “elegant” knives, based on his own personal intuition. Other two designers are asked to create two new "modern" and two new "individualized" knives respectively, but based on setting the significant factors at the appropriate levels, according to the results from the synthesis phase. All six knives are then rated on a 9-point semantic scale. The results are very good for the "modern" and "individualized" knives – designed with the help of the output from the Kansei Engineering study –, but very poor for the "elegant" knives – created by just using the designer's intuition.

5.7.2. Presentation of results

Results from a Kansei Engineering study are addressed, basically, to products' designers. In the same way that the joined work of experts in statistical methods and technicians from an industrial process is a winning combination for improving quality, the collaboration between experts in Kansei Engineering and designers can be very productive:

> *We always collaborate with the designer group in the final design stage because the designers are invaluably key persons in creating excellent products which the customers are satisfied with and can enjoy. Kansei engineering/Kansei ergonomics just provides the sensible and sensitive data analyzed by a technology based on the human kansei. The collaboration of the kansei engineer with the designer is needed in the final stage for the success of kansei product development.* (Nagamachi 2008)

The majority of the reviewed papers focus on spanning the semantic space and the space of properties, first; and on linking both spaces with the synthesis phase, later. Very few lines are devoted to discuss how to present the conclusions in an easily "digestible" way. However, in my opinion, this is an important issue[78].

Some papers implement the results from the KE study in a kind of database accessible by a computer interface (Jindo 1997, Jindo, Hirasago & Nagamachi 1995, Yang, Nagamachi & Lee 1999). In Jindo, Hirasago & Nagamachi (1995), to give an example, a program generates a 3D view of an office chair according to the selected Kansei word, by setting the design elements to the levels specified by the Kansei Engineering model: for instance, it generates a chair that is perceived as stylish. This approach is close to the so-called Kansei Engineering System (KES); using the computer database, two directions can be followed:

[78] I will make proposals on presenting results from Kansei Engineering studies in Chapter 8 and 9 of this dissertation.

- A forward inference (from Kansei to design elements): introducing the desired Kansei, the system tells the needed values for the design elements (or at least shows existing products that elicit that Kansei).
- A backward inference (from candidate design to Kansei): introducing the values of the design elements (perhaps in an indirect way, from a designer's rough sketch, for example), the Kansei the product will elicit is revealed.

Although acceptable in some occasions, in general I am not eager to use Kansei Engineering as a pure automatic tool. I feel closer to the opinion expressed by Barnes and Lillford (2009): "*It should be emphasised that the tools provide support for designers in their creativity and decision-making rather than providing an automated design engine intended to substitute them*".

An interesting idea is suggested in Kuang and Jiang (2009), a paper devoted to how Kansei Engineering can be used in mass customization (MC):

> *Normally, MC is accomplished through developing product platforms. A product platform is the set of common parameters, features, and/or components that remain constant from product to product within a given product family. (…) The platform parameters are defined as those parameters that have nearly no influence on the customers' perception. On the other hand, individual parameters are defined as those parameters that have a large influence on the customers'perception. In product platform design, however, it is unreasonable for a designer to subjectively judge the influence of a parameter.*

So Kansei Engineering can be used to detect those design characteristics that do not affect customers' perception (platform parameters) and to know which design elements must be changed for individualizing each product to convey the desired perception to each person.

In the context of quality improvement, documenting the steps followed to improve a process and its results can be inspiring for future projects (and some ideas can be even reutilized). In the same way, the design decisions coming from Kansei Engineering studies constitute a valuable patrimony for a company: "*After a product has been launched the record of how a selection decision was taken becomes useful as a legacy to be consulted before subsequent product revisions*" (Chen et al. 2009).

5.8. What has been left aside in this literature review

Although exhaustive, this literature review is not complete, in the sense that does not incorporate all available written material about Kansei Engineering studies. In particular, only papers published in English journals have been considered, disregarding conference proceedings.

Undoubtedly, some papers that fall under the knowledge area considered in this literature review have not been retrieved in the literature search. The reason is that they did not state the words “Kansei Engineering” in the title, abstract or keywords (in fact, some of them do not include the words “Kansei Engineering” in any part of the article). Some of these papers can be found searching for “semantic differential method” – for example, Chuang, Chang & Hsu (2001), Komine et al. (2007), Jeong, Lim (2009). These papers, however, perform a principal component analysis or a factor analysis with data collected from questionnaires, but do not try to link Kansei words with design elements, a key ingredient of KE studies.

On the contrary, a paper has been retrieved that has the words “Kansei Engineering” in the abstract, but, while interesting, does not make any contribution to the specific area of knowledge of KE studies (Leon 2009). In fact, aside from the abstract, this paper does not have the words “Kansei Engineering” or simply “Kansei” anywhere else.

The decision of not including conference proceedings in the literature review has necessarily reduced the range of methods used in KE. Some conference proceedings are available through ISI Web of Knowledge and Engineering Village, but not all of them. Having include this works in the literature review could have supposed some kind of bias (although an unavoidable bias is also introduced not considering any conference proceeding). From a practical time perspective, conference proceedings have not been included to limit the amount of paper reading, already big.[79] However, the main reason for not including conference proceedings is that journal papers are usually more polished, and I have considered that they offer all the knowledge from the topic in a distilled way. In fact, many papers started as a presentation at a conference and ended up as an article in a journal.[80]

At least two projects funded by the European Commission were devoted, with more or less depth, to Kansei Engineering: Kensys (Kansei Engineering System), funded

[79] This chapter devoted to the literature review has been one of the most costly in time terms.

[80] Both external referees that reviewed this thesis stated that the published papers are only a subset of the research which is taking place. In particular, one referee states that this method of selection has underrepresented location studies (those that stop in the data collection, without entering in the synthesis phase). Other interesting studies (such as studies comparing different forms of presentation of stimuli) have been also left aside. Thanks to a fast review I did on conference proceedings, I would also say that quite a lot of them describe methods from the automatic learning approach (rough sets, neural networks, fuzzy logic, genetic algorithms, etc; topics not deeply treated in this dissertation).

under the Fifth Framework Programme (Project Reference: IPS-2001-42075), from 2003 to 2006, and ENGAGE (Engineering Emotional Design), funded under the Sixth Framework Programme (Project Reference: 510998) from 2004 to 2007. This projects produced a lot of interesting output related with Kansei Engineering, but has not been included neither (in part because access to the material is not easy).

6 Tools for the Synthesis Phase, in Depth

This chapter explains in detail some commonly used methods for linking the semantic space (the Kansei words) with the space of properties (the technical properties). Each tool is introduced with several examples, and contributions to improve understandability of some techniques are proposed. Some new techniques not found yet in KE studies are also suggested.

6.1. An overview of tools for the synthesis phase.

The literature review done in Chapter 5 showed several tools used for the synthesis phase in KE studies. Figure 5.14 summarizes the list of most common techniques. From this list of techniques, two main methods of linking the semantic space and the space of properties can be derived. These two methods have a totally different approach to relate the independent variables (product properties) and the dependent responses (the Kansei words):

- The statistical approach, with QT1 as the main representative (42% of the reviewed papers used QT1). Twelve percent of papers did not use QT1, but simply a multiple linear regression (MLR). As QT1 is a kind of MLR, an overwhelming 54% of the papers use a linear regression in the synthesis phase. This implies estimating model parameters through a statistical model. Typical assumptions on the data required for regression models must be fulfilled. Some papers – less than 10% of the reviewed ones, and all of them from European authors – use an ordinal logistic regression.

- The automatic learning approach, with neural networks and rough sets as main representatives. An advantage of this approach over the statistical approach is that there is no need to make any assumptions on the data (the disadvantages are different depending on the use of neural networks or rough sets, and will be later stated).

In this chapter, some tools for the synthesis phase will be explained in detail, and new ideas to improve its behavior and understanding will be proposed. Basically, I will cover Quantification Theory Type I and ordinal logistic regression in Section 6.2, and rough sets in Section 6.3. But before entering into these topics, I will briefly describe neural networks and categorical regression (Section 6.1.2). Neural networks are quite common in KE studies, so I want to give an overview of them for the sake of completeness. However, they will not be used in this dissertation – and the reasons for discarding them will also be clarified. Categorical regression is scarcely used in Kansei Engineering, but it could perfectly be used more frequently, and I find it curious enough to explain a little bit of it.

Figure 6.1 graphically summarizes the tools that will be explained in this chapter. The main part of this chapter is devoted to the following tools:

- Under the statistical approach, Quantification Theory Type I (QT1) and ordinal logistic regression (OLR). The ideas supporting QT1 are very simple (a multiple linear regression with dummy variables). However, although literature often gives QT1 results in case studies, it is not clear how to get these results: some detailed algorithmic procedures in QT1 always remain somehow obscure. They will be divulged in Section 6.2.

  But my preferred statistical tool for the synthesis phase is not QT1. Instead, ordinal logistic regression is a natural choice taking into account the kind of data used in KE studies. Output from an OLR is much complex than that of QT1, but many visual enhancements will be proposed to facilitate interpretation – and make it as easy as in QT1. A last step in this search of rigor in Kansei Engineering data analysis will be achieved with the incorporation of participants in the study as random effects in the ordinal logistic regression model.

- Under the automatic learning approach, rough sets. Simply stated, rough sets is an algorithm for extracting patterns in a dataset (the same that could be done by hand, but performed automatically). I think it can complement well the output from the logistic regression. I will cover two flavours of rough sets: one works with data collapsed among participants (the same as QT1), whereas the other works with data at an individual level (the same as OLR).

Statistical approach	Automatic learning approach
Quantification theory type I (QT1) Ordinal logistic regression (OLR) Mixed-effects ordinal logistic regression (mOLR) Categorical regression (CatReg)	Original rough sets (ORS) Variable Precision Bayesian Rough Set (VPBRS) Neural networks (NN)

Figure 6.1. A summary of methods for the synthesis phase

All tools will be exemplified in Chapter 6 with an invented dataset: the T-shirts study (described in the next section). Chapter 7 will show all tools in action when faced with data from a real Kansei Engineering study (the juice experiment)

6.1.1. An invented example of KE study: T-shirts.

A simple and invented dataset will be used to exemplify the proposed methods for the synthesis phase of a KE study. This example was originally presented in the First European Conference in Kansei Engineering (Marco-Almagro, Tort-Martorell Llabrés & García Subirats 2007). The choice of domain for this example is T-shirts (Figure 6.2).

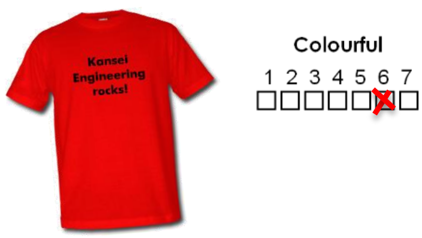


Figure 6.2. T-shirts are rated on the Kansei word colorful

As the purpose of this example is just illustrating several methods of analysis, only one response is used (the Kansei word Colorful). Four different persons evaluate 12 different T-shirts. Three factors are used, the first two with 2 levels and the third one with 3 levels:

x_1: Color: x_{11} = white; x_{12} = red.

x_2: Sleeves: x_{21} = long-sleeved; x_{22} = short-sleeved.

x_3: Printing: x_{31} = picture; x_{32} = plain; x_{33} = text.

Table 6.1 shows all combinations of T-shirts used in the study. Each T-shirt has an identification number for easy referring to it later. The design matrix is a mixed-level full factorial design ($2^2 \cdot 3^1$)

Table 6.1. All combinations of T-shirts in the study.

	Color	Sleeves	Printing			Color	Sleeves	Printing	
1	White	Long	Picture		7	White	Short	Plain	
2	Red	Long	Picture		8	Red	Short	Plain	
3	White	Short	Picture		9	White	Long	Text	
4	Red	Short	Picture		10	Red	Long	Text	
5	White	Long	Plain		11	White	Short	Text	
6	Red	Long	Plain		12	Red	Short	Text	

Each person (Carla, Joan, Marc and Maria) rates each T-shirt on a Likert scale, from 1 to 7. The (invented) ratings have been generated so that factors Color and Printing affect the response, whereas Sleeves is an inert factor. The results are given in Table 6.2.

Table 6.2. Results from the T-shirts experiment.

	Color	Sleeves	Printing	Carla	Joan	Marc	Maria	MEAN	
1	Red	Long	Picture	5	6	7	5	5.75	1
2	White	Long	Picture	3	4	5	3	3.75	2
3	Red	Short	Picture	7	4	5	6	5.50	3
4	White	Short	Picture	4	4	4	3	3.75	4
5	Red	Long	Plain	1	4	5	1	2.75	5
6	White	Long	Plain	1	3	2	1	1.75	6
7	Red	Short	Plain	4	5	5	2	4.00	7
8	White	Short	Plain	2	4	5	1	3.00	8
9	Red	Long	Text	5	6	6	4	5.25	9
10	White	Long	Text	1	2	5	1	2.25	10
11	Red	Short	Text	5	6	7	4	5.50	11
12	White	Short	Text	2	4	3	2	2.75	12

Looking carefully at the data in Table 6.2 – and with a little effort – it is possible to discover which factors affect the response Colorful. T-shirts 5, 6, 7 and 8 (the ones that have plain printing) have basically lower punctuation than T-shirts 9, 10, 11 and 12 (the ones with text printing). T-shirts 1, 2, 3 and 4 have the highest punctuations. On the other hand, odd T-shirts (red color) have higher rates than even T-shirts (white color). Figure 6.3 shows the means of all T-shirts having a given condition.

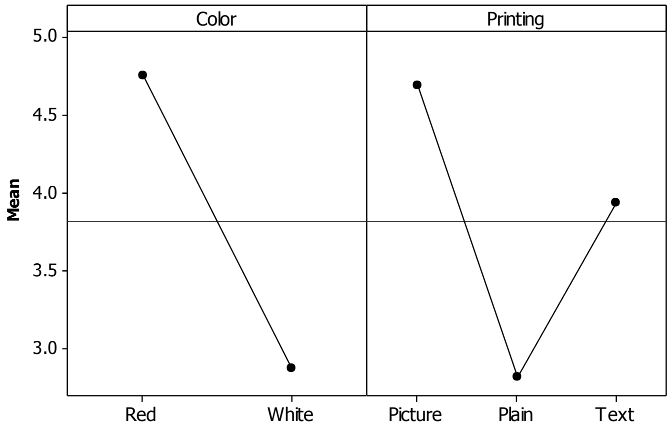


Figure 6.3. Items affecting the response: Color and Printing, with data from Table 6.2.

Data from the T-shirts example will be used to exemplify all analysis techniques described in this chapter. The only purpose of using this simple dataset is introducing each technique in the context of a Kansei Engineering study. Data is prepared so that conclusions will be the same regardless of the technique used (and no surprise will arise: all results have already been presented in Figure 6.3). In this way, attention can be focused more on the technique details than in the particular conclusions reached. Chapter 7 will offer a comparison of techniques applied to a Kansei Engineering study real dataset (the juice experiment).

6.1.2. Interesting tools that will be skipped in this dissertation

Some tools for the synthesis phase will be studied in detail in this chapter (statistical regression in Section 6.2 and rough set analysis in Section 6.3). But before, this subsection is devoted to two other tools that could be interesting for KE studies, but that have been skipped in this dissertation: neural networks (NN) and categorical regression (CatReg). A short description of these techniques will be offered and data from Table 6.2 (the T-shirts example) will be used to exemplify the procedures[81]. I

[81] SPSS is able to perform categorical regression and neural network analysis, and it is the software used in this section. Both tools can also be implemented in `R` using package `homals` (for categorical regression through optimal scaling) and `nnet` (for neural networks)

will also explain in which situations I think they can be useful, and why they have not been included in this work.

Neural networks (NN)

Neural networks (NN) are a popular tool for many predictive applications because of their versatility. In Kansei Engineering, NN are the most popular tool from the automatic learning approach, as stated in Section 5.6.2.

A large family of models can be located under this umbrella term of neural networks. These models resemble the way our brain processes knowledge; that is the reason for using the analogy with neurons and synaptic connections (however, this analogy is nowadays kept in nomenclature more for historical reasons than for a real similarity).

American scientists Warren McCulloch and Walter Pitts were the first in proposing the ideas behind neural networks in a seminal paper (McCulloch, Pitts 1943). During the 1950s and 1960s the field of neural networks became quite popular among researchers, but was later abandoned. In the middle of the 1980s, interest in NN revived, until now. Figure 6.4 shows a representation of a commonly used NN, called a multilayer perceptron (Ripley 2007). It is a feedforward network, that is, the "signal" flows forward (from left to right), without any feedback loops.

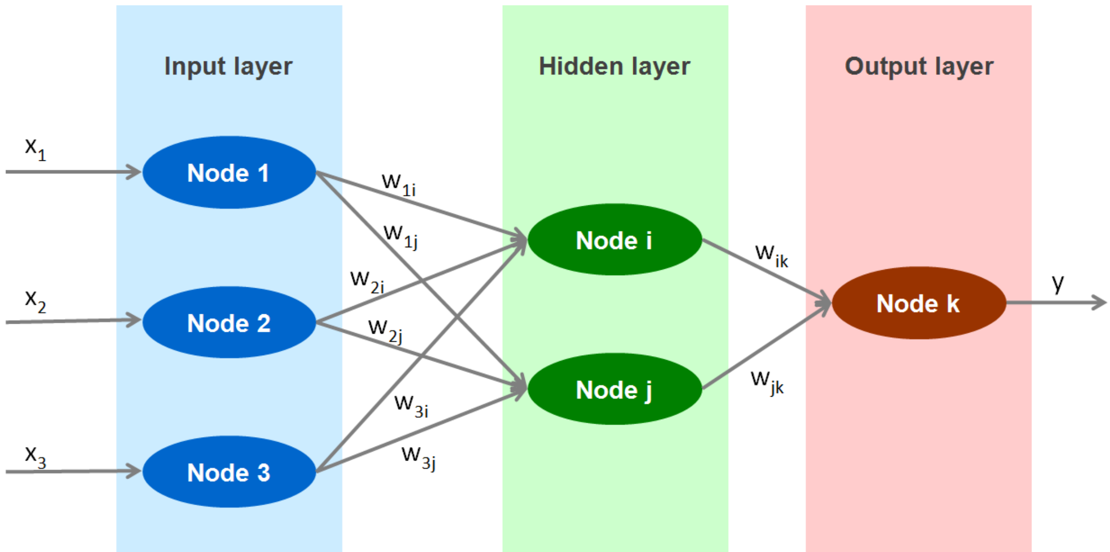


Figure 6.4. A schematic representation of a feedforward neural network.

A neural network can be used to predict a response, as in a regression model. Take the multilayer perceptron in Figure 6.4. It is composed by several layers, each one containing several nodes (called neurons). Each neuron in the input layer is fed with the independent variables from the regression model. The hidden layer contains unobservable neurons (two in this case, but we could have a different number). The output layer contains the responses (the dependent variables, just one in this example).

Each node in the hidden layer receives a weighted sum of outputs (usually called activation) from the input layer. For example, node i receives the sum $act\ i = w_{1i}\, x_1 + w_{2i}\, x_2 + w_{3i}\, x_3 + w_{0i}$ as input. w_{0i} is a constant (also called a bias). The output is a nonlinear function of its activation, $f(act\ i)$, called activation function. Nodes in the output layer work in a similar manner, but with inputs from the hidden layer.

Two matters must be solved before having a multilayer perceptron able to correctly predict the response. The first one is the architecture of the network (basically, deciding the number of neurons in the hidden layer). The second one is determining the value of the weights for each synaptic connection. Several methods exist for these purposes.

Calculating the value of the connection weights is called training the network. In the backpropagation procedure (a common method for training a neural network), the network learns by presenting inputs from an existing dataset (the training dataset), and propagating forward the signal in order to get the output. An error signal (distance between the computed output and the real output) is calculated, and propagated back through the layers. The connections weights are adjusted to reduce this error, and a new iteration starts. The procedure stops when the error is below a certain value.

A dataset is very often divided into two sets: a training set (the one used to calculate the connection weights), and a testing set (the one used to evaluate the performance of the network when faced with new inputs). It is important to assess the quality of the network with the testing set, as a real risk in neural networks is overtraining (in that case, the network predicts well when faced with inputs from the training set, but bad when fed with new inputs).

I will use the multilayer perceptron routine in SPSS (with default parameters) to design a neural network for the T-shirts example. As independent variables are factors, each category in each factor will be represented by a node in the input layer. For example, factor Color is represented by two nodes (Color=red, with input 1 if Color is red and input 0 otherwise, and Color=white, just the opposite). As the output is ordinal (with 7 possible values), we could have 7 output neurons, one for each value from 1 to 7. To simplify, I have transformed the output from Table 6.2 in the following manner: for each person, the response Colorful is 'No' if it is below the mean of responses for that person, and 'Yes' otherwise – therefore, the new response only has two categories. The diagram of this neural network is shown in Figure 6.5.

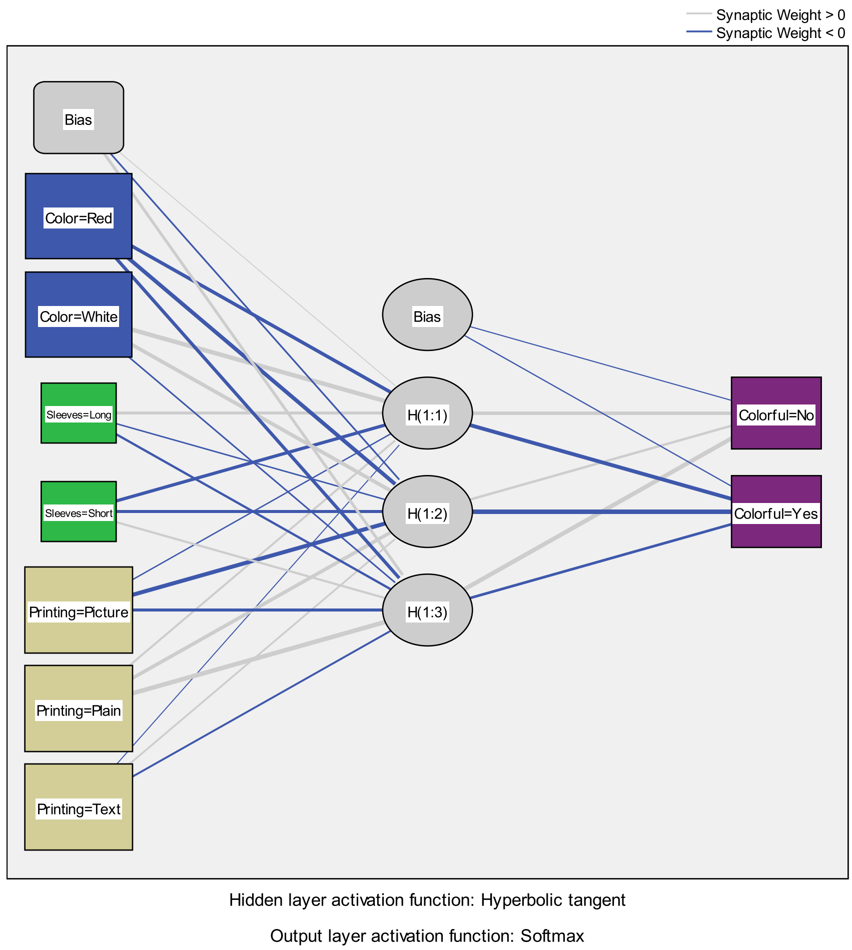


Figure 6.5. Neural network diagram for the T-shirts example.

Table 6.3 shows the SPSS output with the weights for each one of the synaptic connections in the diagram. Using these weights, when a new stimulus is used, a number is computed for each one of the output neurons (Colorful=No and Colorful=Yes), representing a probability of obtaining each response. The stimulus is attached to the response (No or Yes) with a higher probability. Using this procedure with the testing set (a randomly selected set of 30% of the complete dataset shown in Table 6.2), there is only a 18.8% of incorrect predictions.

This extremely brief revision of neural networks should be enough to illustrate its advantages, but also its difficulties when being used in KE studies. A great advantage of NN over traditional statistical methods such as regression is that neural networks require minimal assumptions on the data and – if correctly designed and trained – fit in each case the most appropriate model for the dataset (either linear or nonlinear).

They are extremely versatile (in fact, linear regressions could be considered a particular case of NN).

Table 6.3. SPSS output with the value of the synaptic weights in the neural network for the T-shirts example.

Predictor		Predicted parameter estimates				
		Hidden Layer 1			Output Layer	
		H(1:1)	H(1:2)	H(1:3)	[Colorful=No]	[Colorful=Yes]
Input Layer	(Bias)	,016	-,178	,396		
	[Color=Red]	-,558	-,600	-,479		
	[Color=White]	,879	,575	-,112		
	[Sleeves=Long]	,394	-,090	-,290		
	[Sleeves=Short]	-,447	-,361	,288		
	[Printing=Picture]	-,064	-,738	-,409		
	[Printing=Plain]	,283	,569	,929		
	[Printing=Text]	-,046	,273	-,277		
Hidden Layer 1	(Bias)				-,060	-,063
	H(1:1)				,434	-,594
	H(1:2)				,336	-,836
	H(1:3)				,844	-,384

Are there any problems with NN? Besides the risk of overfitting, the trade-off for their flexibility is that the synaptic weights of a neural network cannot be easily interpreted. Thus, if the purpose of a Kansei Engineering study is explaining how a relatively small set of factors affect a set of Kansei words, a statistical regression model would be a better option.

The kind of KE studies treated in this dissertation are those having a small set of factors (say, something between 3 and 6, for sure no more than 10), a relatively small set of stimuli (say, something between 10 and 20), and the goal is understanding how each factor affects each Kansei word. For this reason I will not delve in neural networks for Kansei Engineering studies. However, I think neural networks can be useful and adequate in some KE studies, basically when the following conditions apply:

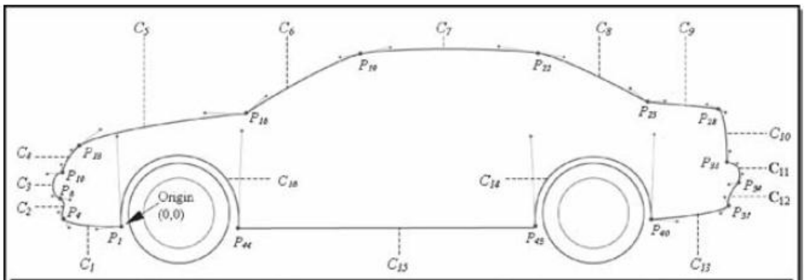
- There are lots of properties (probably dozens of them, maybe a hundred).
- The number of stimuli is high (a hundred or even more).
- The aim is predicting the emotional response and not discovering how each particular factor affects the response.

An example of an appropriate use of neural networks for Kansei Engineering can be found in Chang and Chen (2007), a case study already explained in Section 5.6.2. All three conditions just stated above apply:

- There are 92 properties (different points in an automobile profile)

- A high number of stimuli (144 vehicle profiles) are used in the data collection phase.
- The goal is not understanding how each one of the 92 points affect the response, but to globally determine if a profile proposed by a designer (by means of moving the points on a computer interface) elicits a specific emotion.

In this case study, a car profile is characterized by 92 points (as shown in Figure 6.9). The car profile is presented on a computer screen. A designer can move a little bit, using the mouse, each one of the 92 points, changing the car profile (for example, making the back of the car more rounded or more squared). After each change in the profile, the neural network predicts how the car profile is perceived (stating Yes or No for each Kansei word; for example, Powerful: No, Modern: Yes, etc.)



Curve segment designation	Point	Variables
C_1 : Under -part segment of fore bumper	$P_1 \sim P_4$	$V_{1(X)} \sim V_{8(Y)}$
C_2 : Middle-part segment of fore bumper	$P_4 \sim P_7$	$V_{7(X)} \sim V_{14(Y)}$
C_3 : Upper -part segment of fore bumper	$P_7 \sim P_{10}$	$V_{13(X)} \sim V_{20(Y)}$
C_4 : Fore lamp segment	$P_{10} \sim P_{13}$	$V_{19(X)} \sim V_{26(Y)}$
C_5 : Engine hood segment	$P_{13} \sim P_{16}$	$V_{25(X)} \sim V_{32(Y)}$
C_6 : Fore windshield segment	$P_{16} \sim P_{19}$	$V_{31(X)} \sim V_{38(Y)}$
C_7 : Car roof segment	$P_{19} \sim P_{22}$	$V_{37(X)} \sim V_{44(Y)}$
C_8 : Rear windshield segment	$P_{22} \sim P_{25}$	$V_{43(X)} \sim V_{50(Y)}$
C_9 : Trunk segment	$P_{25} \sim P_{28}$	$V_{49(X)} \sim V_{56(Y)}$
C_{10} : Rear lamp segment	$P_{28} \sim P_{31}$	$V_{55(X)} \sim V_{62(Y)}$
C_{11} : Upper-part segment of rear bumper	$P_{31} \sim P_{34}$	$V_{61(X)} \sim V_{68(Y)}$
C_{12} : Middle-part segment of rear bumper	$P_{34} \sim P_{37}$	$V_{67(X)} \sim V_{74(Y)}$
C_{13} : Upper-part segment of rear bumper	$P_{37} \sim P_{40}$	$V_{73(X)} \sim V_{80(Y)}$
C_{14} : Protection hood segment of rear wheel	$P_{40} \sim P_{43}$	$V_{79(X)} \sim V_{86(Y)}$
C_{15} : Chassis segment	$P_{43} \sim P_{44}$	$V_{85(X)} \sim V_{88(Y)}$
C_{16} : Protection hood segment of front wheel	$P_{44} \sim P_1$	$V_{87(X)} \sim V_{2(Y)}$

Figure 6.6. Specification of an automobile profile using 92 different geometrical properties, reproduced from Chang and Chen (2007)

Categorical regression (CatReg)

Unlike the relative popularity of neural networks for some kind of KE studies, categorical regression (CatReg) is scarcely used in Kansei Engineering: I am only aware of its use in a PhD thesis (Such Pérez 2003) and a paper (Lanzotti, Tarantino & Matrone 2008). The reason for including it in this chapter is that I think the rationale behind it could fit well in Kansei Engineering.

Linear multiple regression is easy and works very well when all variables are continuous. But everything is more complicated when there is the need to accommodate ordinal and nominal variables (both as regressors or as the response). When regressors are categorical variables – the common situation in KE studies – they can be converted into dummy variables, and all but one (the reference level) introduced in the linear regression model[82]. When the response is not continuous, but ordinal or nominal, it is not possible to use a linear regression anymore, and generalized linear models make its appearance.

An alternative to the use of generalized linear models is transforming all ordinal and nominal variables into continuous ones (not just the response is transformed, but also all regressors). This transformation is done using a procedure called optimal scaling using alternating least squares (Meulman 1998). A common multiple regression analysis is then performed with the transformed variables.

This idea of converting ordinal and nominal variables into continuous ones can be applied not only for regression, but also for many others multivariate methods only suitable, initially, to continuous variables (such as principal component analysis) (Meulman, Heiser 1989). The optimal scaling procedure is adjusted to the later applied method: for categorical regression, the transformation of variables is designed to maximize the relationship between each regressor and the response. The numbers assigned to the original ordinal and nominal variables are called quantifications. The alternating least squares method is an iterative procedure: a first set of quantifications are used to find a solution, quantifications are updated using that solution, the updated quantifications are used to find a new solution, which is then used to update the quantifications, and so on, until some stop criterion is reached[83].

---

[82] It is not that clear what can be done when the regressors are ordinal variables. A fast – but not very convincing – workaround is just considering the ordinal variable as continuous.

[83] The algorithms for optimal scaling using alternating least squares in categorical regression can be found in the SPSS research paper CATREG (support.spss.com/productsext/spss/documentation/statistics/algorithms/14.0/catreg.pdf, accessed June 2010)

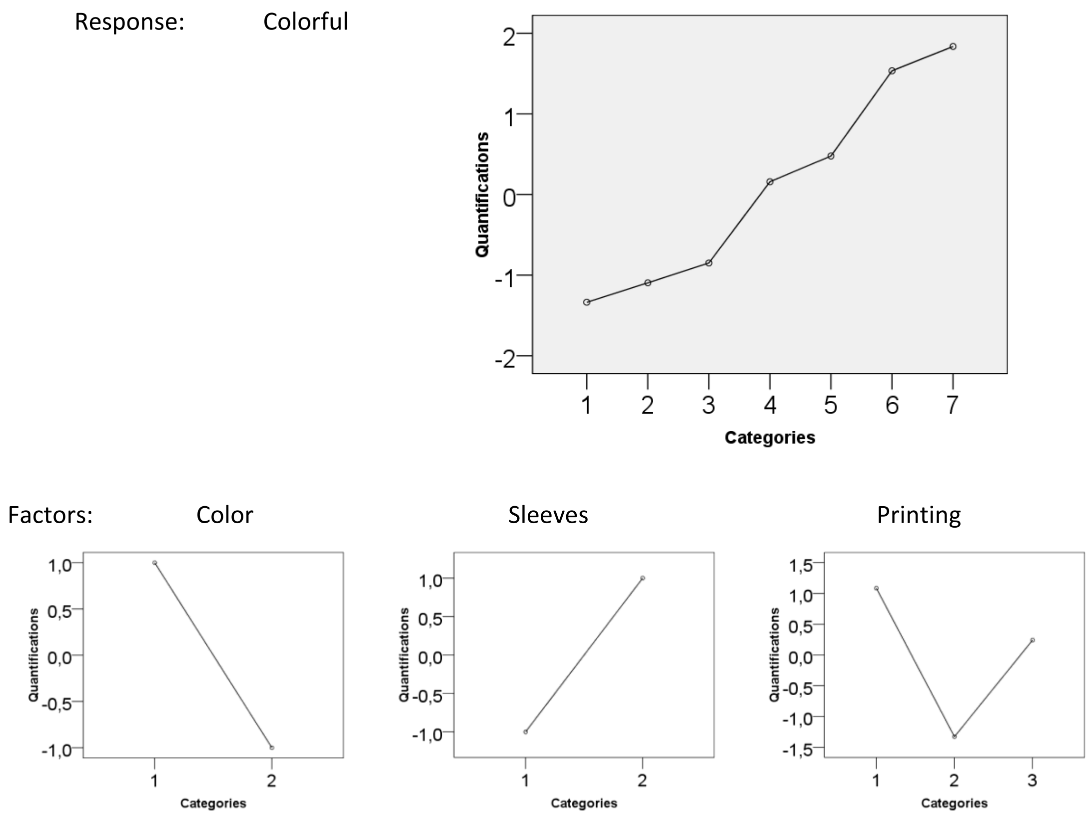


Figure 6.7. Transformation plots given by the SPSS procedure CATREG

A nice way to visualize the quantifications is using transformations plots. These plots have the quantification in the vertical axis, and each one of the categories in the horizontal axis. Figure 6.7 shows the transformation plots for data in the T-shirts example, as obtained using the CatReg procedure in SPSS. An ordinal variable, as the response Colorful in the T-shirts example, has quantifications always increasing. If two or more adjacent categories have a very similar quantification, those categories can be safely collapsed. Categories in nominal variables, as factor Printing in the T-shirts example, have quantifications with no restrictions.

Besides the transformations plots, the output of a categorical regression resembles that of a common linear regression. Table 6.4 shows a summary of the output for the T-shirts example. Significance of each factor can be assessed, as usual, with the p-values (Color and Printing are significant, but Sleeves is not). The direction of the relationship for each factor can be evaluated looking at the transformation plot for that factor and its beta coefficient. Category 1 in Color (red) has higher response than category 2 in Color (white). In Printing, categories 1 (picture) and 3 (text) have higher response than category 2 (plain). This is true because both Color and Printing have positive beta coefficients; in case of having a negative beta coefficient, the opposite direction would be true.

Table 6.4. Summary of the SPSS output for the categorical regression with the T-shirts example.

	Standardized Coefficients				
	Beta	s_{Beta}	df	F	p-value
Color	0.592	0.128	1	21.489	0.000
Sleeves	0.104	0.103	1	1.019	0.318
Printing	0.390	0.129	2	9.108	0.001

	SS	df	MS	F	p-value	Multiple R	R^2	R^2_{adj}
Regression	24.625	4	6.156	11.325	0.000	0.716	0.513	0.468
Residual	23.375	43	0.544					
Total	48.000	47						

Is categorical regression an appropriate procedure for Kansei Engineering studies? Looking at the characteristics of variables in KE (ordinal responses and usually nominal regressors), the answer would clearly be yes. The supposed advantage over using an ordinal logistic regression with dummy variables for accommodating the nominal regressors is ease of interpretation (as CatReg output mirrors conventional multiple regression).

Results in categorical regression are, however, very similar – if not identical – to those from an ordinal logistic regression, as can be seen in the study comparing both approaches done by Lanzotti, Tarantino & Matrone (2008). My election for use in KE studies is ordinal logistic regression, and not categorical regression, for the following reasons:

- Ordinal logistic regression is more popular than categorical regression, and its foundations are clear and well established.
- Categorical regression needs transforming ordinal and nominal variables into continuous variables. If possible, I prefer using original variables as they are, avoiding transformations.
- The supposedly complicated output from ordinal logistic regression (basically coefficients not directly interpretable but through odds ratios and the presence of a reference level for each categorical variable) can be clarified with graphical proposals that will be later presented in this chapter.

In my opinion, with the new graphical way to present results from an ordinal logistic regression that will be introduced in Section 6.2.2, interpretation of results from an ordinal logistic regression is easier and more appealing than that of categorical regression.

6.2. Regression models in the synthesis phase.

In statistics, regression analysis – in its broader sense – can be used to model the relationship between a dependent variable and one or more independent variables. In a KE study, each Kansei word from the semantic space acts as a response, whereas each factor from the space of properties is an independent variable. All the well known techniques from regression analysis come to our help for facing the synthesis phase from a statistical approach.

Therefore the aim is estimating (at least) the main effects and (if possible) second-order interactions[84]. In the T-shirts example, our aim is discovering the relationship between Color, Sleeves and Printing on the Kansei word Colorful.

Two facts always characterize Kansei Engineering datasets:

1. Independent variables are categorical factors, having two or more levels (quite often, some factors do have more than two levels). There are no quantitative independent variables.
2. The response is discrete, usually integers from an ordinal scale (ranging from 1 to 7, or from 1 to 5, for example).[85]

As the response is discrete, a multiple linear regression cannot be directly used for modelling the data, as a continuous response is needed. The most common way of solving this issue is working with the mean of all participants' ratings. As independent variables are all categorical, dummy variables must be used. This is basically what it is done in quantification theory type I (with some addendum, as we will see) [86].

---

[84] As it will be explained in Chapter 7, designs in Kansei Engineering studies often allow only the estimation of main effects. Second-order interactions are commonly not taken into account.

[85] As we have seen in Section 3.4.2, sometimes Kansei words are not rated on an ordinal scale, but on a visual analogue scale (VAS), having a much wider set of possibilities (usually 0 to 100). However, I already stated that using a VAS creates the illusion of having a continuous variable, but it is in fact an attempt to measure with artificial accuracy a construct that does not allow this level of precision. Data from a VAS can always be transformed into ordinal data. For example, in a VAS from 0 to 100, 27 (2.7) is converted into 3; 43 (4.3) is converted into 4; 85 (8.5) is converted into 9; and so on.

[86] In fact, it would be possible to perform a linear regression with all raw data coming from all participants, considering the ordinal responses as continuous. However, this is never done, and collapsing the responses among all participants using the mean is the preferred procedure.

Of course, information is lost when summarizing data with the mean. In particular, variability among participants in the study is not considered (I will delve deeper into this issue in Chapter 8). Why not constructing a model directly with our raw data? An ordinal logistic regression (OLR) is appropriate to this purpose. OLR is seldom used in KE studies, but it seems a very natural way to treat the data, as the response is ordinal. Dummy variables must be also used for introducing factors as regressors in the model.

A last detail must be taken into account (although it is never done): a requirement for performing the OLR is having independent responses. However, we know our responses are not independent: each participant in the study gives a rating to the whole set of stimuli. Ratings from the same person are not independent. This variability among participants should be taken into account, and the best way to do it is introducing the factor subject in the model as a random effect. In this way, properties are treated as fixed effects, whereas participants are treated as random effects. The resulting model is a mixed effects ordinal logistic regression. As far as I know, this proposal of introducing subjects as random effects in an ordinal logistic regression for KE studies is absolutely new: no references in the literature follow this approach.

The next subsections will face the three proposed regression models for Kansei Engineering studies: quantification theory type I (QT1), ordinal logistic regression (OLR) and mixed effects ordinal logistic regression (MixOLR).

6.2.1. Quantification theory type I (QT1)

Quantification theory type I is one of the most commonly used methods in KE studies (ranked first according to the literature review in Chapter 5). See, for instance, Lai et al. (2006), Ying and Yan (2006), Jindo, Hirasago & Nagamachi (1995), Matsubara and Nagamachi (1997), Okada and Castillo (2007), Fukushima et al. (1995), Nakada (1997), Tanoue, Ishizaka & Nagamachi (1997) and Jindo (1997). In fact, QT1 is also used in Japan in many other disciplines not related at all with KE (Genjo et al. 2005, Naka et al. 2007, Oba, Onoda & Komoda 2000, Matsumura 2004).

QT1 was first proposed by Chikio Hayashi in a seminal paper (Hayashi 1952), although the current name was given later. Dr. Hayashi focused on working with categorical data, and invented several methods grouped under the umbrella of quantification theories. Quantification theory type I is a multiple regression analysis with qualitative explanatory variables. Other quantification theories are similar to known methods: QT2 is discriminant analysis, QT3 is correspondence analysis and QT4 is multidimensional scaling. It is fascinating that both Chikio Hayashi in Japan and Jean Paul Benzécri (the "inventor" of correspondence analysis) in France were working in the decade of 1950 on the same topics, but in parallel, and led to the same solutions (Nishisato 2006).

QT1 attempts to quantify the relationship between the response (Kansei words) and the properties (design elements). It is simply a multiple regression model, but with some clever attributes, as we will see. Properties, usually called factors in the field of factorial designs, are here named items. Each item has several categories (levels, in factorial designs). All regressors in the equation will thus be qualitative, while the response is quantitative (the mean of all participants' ratings to a Kansei word).

As all items are qualitative, they are introduced in the regression model using dummy variables (Draper, Smith 1998). Dummy variables are built in the following way – using the notation employed in Schütte (2005):

$$\delta_{i(jk)} = \begin{cases} 1, & \text{if product i has category k in item j} \\ 0, & \text{otherwise} \end{cases}$$

$i = 1, \ldots, n$ (with n the number of products)
$j = 1, \ldots, R$ (with R the number of items)
$k = 1, \ldots, C_j$ (with C_j the number of categories in item j).

Table 6.5 gives the dummy variable conversion of the design matrix for the T-shirts example, together with the response in the last column.

Table 6.5. The design matrix of the T-shirts experiment with coded units

	Colour	Sleeves	Printing	x_{11} Red	x_{12} White	x_{21} Long	x_{22} Short	x_{31} Picture	x_{32} Plain	x_{33} Text	MEAN
1	Red	Long	Picture	1	0	1	0	1	0	0	5.75
2	White	Long	Picture	0	1	1	0	1	0	0	3.75
3	Red	Short	Picture	1	0	0	1	1	0	0	5.50
4	White	Short	Picture	0	1	0	1	1	0	0	3.75
5	Red	Long	Plain	1	0	1	0	0	1	0	2.75
6	White	Long	Plain	0	1	1	0	0	1	0	1.75
7	Red	Short	Plain	1	0	0	1	0	1	0	4.00
8	White	Short	Plain	0	1	0	1	0	1	0	3.00
9	Red	Long	Text	1	0	1	0	0	0	1	5.25
10	White	Long	Text	0	1	1	0	0	0	1	2.25
11	Red	Short	Text	1	0	0	1	0	0	1	5.50
12	White	Short	Text	0	1	0	1	0	0	1	2.75

The purpose of QT1 is estimating the coefficients β_{jk} in the following equation (together with the constant β_0, although we are only really interested in the coefficients β_{jk}):

$$Y_i = \beta_0 + \sum_{j=1}^{R} \sum_{k=1}^{C_j} \beta_{jk}\, \delta_{i(jk)} + \varepsilon_i$$

Using matrix notation (and representing matrices as letters in boldface), $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. As usual in linear regression, we assume ε_i are independent and $\varepsilon_i \sim N(0;\ \sigma)$.

Going again to the T-shirts example, we want to estimate the coefficients β_{jk} in the following equation (I am dropping the *i* sub-index from the equation for the sake of clarity):

$$Y = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{21}x_{21} + \beta_{22}x_{22} + \beta_{31}x_{31} + \beta_{32}x_{32} + \beta_{33}x_{33} + \varepsilon$$

These β_{jk} coefficients are called category scores (CS) in QT1.

Calculating Category Scores (CS) in QT1

The usual procedure to estimate these category scores is using the least squares method. The proposed coefficients are then those that minimize the sum of squared residuals, a residual being the difference between an observed value and the value given by the model. In this "easy" world of ordinary least squares, coefficients $\boldsymbol{b}$ can be found solving the normal equations (Draper, Smith 1998):

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\, \boldsymbol{X}'\boldsymbol{Y}$$

But there is a problem in this case: if we construct matrix $\boldsymbol{X}$ with all columns corresponding to our dummy variables (together with a first column of 1's for the constant b_0), the matrix $\boldsymbol{X}'\boldsymbol{X}$ is singular, and thus $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ does not exist. This happens because all columns for a given factor are not independent, and therefore some of the normal equations depend on others.

The common procedure to solve this well-known problem is constructing matrix $\boldsymbol{X}$ with all columns of a given factor but one. In the T-shirts example, we can use only column x_{11} for item Color, only column x_{21} for item Sleeves and only columns x_{31} and x_{32} for item Printing. The deleted category for each item is used as a reference level. Estimating the coefficients with some software (for example, `lm` function in `R`), the equation for the T-shirts example is the following:

$$\hat{Y} = 3.2292 + 1.9167\, x_{11} - 0.5000\, x_{21} + 0.7500\, x_{31} - 1.0625\, x_{32} \tag{6.1}$$

In fact, Eq. 6.1 is obviously the same as this one:

$$\hat{Y} = 3.2292 + 1.9167\, x_{11} + 0\, x_{12} - 0.5000\, x_{21} + 0\, x_{22} \\ + 0.7500\, x_{31} - 1.0625\, x_{32} + 0\, x_{33} \tag{6.2}$$

The variables having 0 as coefficient in Eq. 6.2 are the reference levels. Although having reference levels is something common in regression, it makes interpretation of results somewhat more complex to people not accustomed to it. As the normal equations in this case define an under determined system, with infinitely many solutions, why not choose a solution with no reference levels? This is the idea behind QT1. Specifically, we want to choose the solution that has the constant b_0 as the mean of all responses. Doing this, coefficients for the deleted categories are also computed, and each coefficient is interpreted as the difference from the mean of all responses.

Our purpose, thus, is "translating" coefficients in Eq. 6.2 so that all coefficients are present in the model (with no reference level), and the constant is the mean of all responses. Eq. 6.3 shows our desired finish point for the T-shirts example. Coefficients in the original model (Eq. 6.2) are represented by b_{jk} while coefficients in the transformed model (Eq. 6.3) are represented by $b_{jk}{}'$ (and called category scores).

$$\hat{Y} = b_0' + b_{11}'x_{11} + b_{12}'x_{12} + b_{21}{}'\, x_{21} + b_{22}{}'\, x_{22} + b_{31}{}'x_{31} \\ + b_{32}{}'\, x_{32} + b_{33}{}'\, x_{33} \tag{6.3}$$

A way to achieve this transformation is using a procedure briefly described in the short paper "A Transformation for Simplifying the Interpretation of Coefficients of Binary Variables in Regression Analysis" (Sweeney 1972). The word QT1 is never mentioned in this paper, but it is clear that its purpose is the same. The following statement can be found in this paper:

> *If a model contains several systems of dummy variables, each with a single deleted class, the interpretation of the numerous coefficients is likely to be confusing.*

This is what always happens in Kansei Engineering datasets.

How are the coefficients b_{jk}' in the transformed model obtained? One should work item by item. Consider item *j*. The coefficients in the transformed regression are calculated from the following equation:

$$b_{jk}' = b_{jk} + Q_j$$

Q_j is calculated from the formula:

$$\sum_{k=1}^{c_j} P_k\left(b_{jk} + Q_j\right) = 0$$

Each item *j* will have a different Q_j constant. P_k is the proportion of appearance of category *k* from item *j* in the sample.

An example focusing on the item Printing (*j*=3) from the T-shirts example will illustrate the procedure. The item Printing has 3 categories (Picture, Plain and Text). The coefficients in the transformed regression are calculated from the following equation:

$$b'_{3k} = b_{3k} + Q_3$$

Where Q_3 is a constant computed from

$$0.333(0.7500 + Q_3) + 0.333(-1.0625 + Q_3) + 0.333(0 + Q_3) = 0 \Rightarrow Q_3 = 0.1042$$

Table 6.6 summarizes all calculations for the item Printing.

Table 6.6. Calculation of transformed coefficients for item Printing

Category	Coefficients in original regression	Proportion in design matrix	Coefficients in transformed regression
Picture (x_{31})	b_{31} = 0.7500	4/12 = 0.333	b_{31}' = 0.7500 + 0.1042 = 0.8542
Plain (x_{32})	b_{32} = -1.0625	4/12 = 0.333	b_{32}' = –1.0625 + 0.1042 = –0.9583
Text (x_{33})	b_{33} = 0	4/12 = 0.333	b_{33}' = 0 + 0.1042 = 0.1042

Following the same procedure for items Color and Sleeves, we get the final transformed equation for the T-shirts example:

$$\hat{Y} = 3.8333 + 0.9583\,x_{11} - 0.9583\,x_{12} - 0.2500\,x_{21} + 0.2500\,x_{22} + 0.8542\,x_{31} - 0.9583\,x_{32} + 0.1042\,x_{33} \quad (6.4)$$

Notice that b'_0 is the mean of all responses. The transformed model (Eq. 6.4) gives exactly the same predictions than the original model (Eq. 6.1).

Multiple (MCC) and Partial (PCC) Correlation Coefficients in QT1

The output from QT1 is completed with multiple and partial correlation coefficients. Partial correlation coefficients (PCC) are particularly important, as they quantify the strength of the relation between items and Kansei words.

The square of the multiple correlation coefficient (MCC) is simply the coefficient of determination R^2 in multiple regression, calculated as usually:

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$$

SS_{err} is the residual sum of squares and SS_{tot} is the total sum of squares. R^2 measures the amount of variance of the response explained by the model, so the MCC gives information on the global contribution of the space of properties to each Kansei (to each response).

The partial correlation coefficient provides information about to which amount an item affects the response with all other factors in the model fixed. It can be calculated following this procedure (Tanaka 1979):

- Scores for each item are computed using the category scores (b_{jk}') from the regression model. That is $W_{i(j)} = \sum_{k=1}^{C_j} b'_{jk}\, \delta_{i(jk)}$ with j meaning the j-th item and k the k-th category in the j-th item.
- The regression of Y with W_1 , W_2 , ..., W_{j-1} , W_{j+1}, ..., W_R as regressors is performed and residuals (Res_1) are stored.
- The regression of W_j with W_1 , W_2 , ..., W_{j-1} , W_{j+1}, ..., W_R as regressors is performed and residuals (Res_2) are stored.
- The correlation of residuals Res_1 and Res_2 is the partial correlation coefficient (PCC) for item j.

Looking again at the T-shirts example, Table 6.7 has the computed scores for calculating each of the three PCCs. For example, the calculation of the PCC for item 3 (Printing) involves storing the residuals of regression $Y = f(W_1, W_2) \rightarrow Res1$ and the residuals of regression $W_3 = f(W_1, W_2) \rightarrow Res2$. The correlation between $Res1$ and $Res2$ is 0.846, so this is the PCC for item 3.

Table 6.7. Computed scores in the T-shirts example for calculating the PCC

Y	W_1	W_2	W_3
5.75	0.9583	–0.2500	0.8542
3.75	–0.9583	–0.2500	0.8542
5.50	0.9583	0.2500	0.8542
3.75	–0.9583	0.2500	0.8542
2.75	0.9583	–0.2500	–0.9583
1.75	–0.9583	–0.2500	–0.9583
4.00	0.9583	0.2500	–0.9583
3.00	–0.9583	0.2500	–0.9583
5.25	0.9583	–0.2500	0.1042
2.25	–0.9583	–0.2500	0.1042
5.50	0.9583	0.2500	0.1042
2.75	–0.9583	0.2500	0.1042

The results from QT1 (the MCC, the PCC for each item and all category scores) are normally shown in a graphic with bars. Figure 6.8 shows the result for the T-shirts example.

MCC= 0.9325

Item	PCC	Category		Category Score
Colour	0,894	Red	0,96	
		White	-0,96	
Sleeves	0,462	Long	-0,25	
		Short	0,25	
Printing	0,840	Picture	0.85	
		Plain	-0.95	
		Text	0.10	

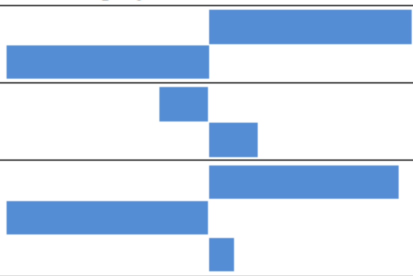


Figure 6.8. QT1 results for the T-shirts example

An improved version of QT1

QT1 is a very attractive way to present the relationship between properties and Kansei words because its results can be shown in a visual way, and it is very easy to interpret. However, looking at the bars in Figure 6.8 it is not obvious that category Sleeves from our T-shirts example do not affect the response at all. T-shirts having long sleeves get the same ratings on the word Colorful than T-shirts having short sleeves. In other words, Sleeves is not significant. How do we know this? The partial correlation coefficient for Sleeves is 0.426, too small for saying that a relationship exists. Usually, a PCC greater than 0.7 is considered to imply a significant relationship between that item and the response (Schütte, Eklund 2003).

Although this statement seems rather reasonable, it could happen that an item with a high PCC has no relation with the response. Take into consideration the invented data in Table 6.8, corresponding to 2 items (A and B). Item A has 3 categories (A1, A2 and A3) and item B has 2 categories (B1 and B2). The response Y is the mean of all Kansei ratings for each stimulus.

Table 6.8. A design matrix with 2 items (A and B)

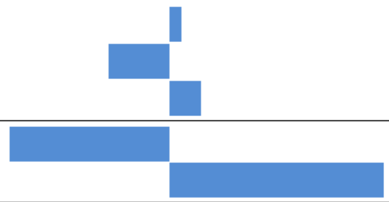
	A	**B**	x_{11} A1	x_{12} A2	x_{13} A3	x_{21} B1	x_{22} B2	**Y**
1	A1	B1	1	0	0	1	0	7.000
2	A1	B1	1	0	0	1	0	6.556
3	A2	B2	0	1	0	0	1	3.444
4	A3	B1	0	0	1	1	0	6.444
5	A2	B2	0	1	0	0	1	3.000
6	A3	B2	0	0	1	0	1	3.889
7	A1	B2	1	0	0	0	1	3.778

Figure 6.9 shows the results when performing QT1 for this data.

Figure 6.9. QT1 results for an invented set of data with 2 items

MCC = 0.9927

Item	PCC	Category	Category Score
A	0.781	A3	0.09
		A2	-0.45
		A1	0.24
B	0.987	B2	-1.20
		B1	1.61



It is clear that item B is significant, as it has a PCC of 0.987. But it seems that item A is also significant, with a PCC of 0.781. But, is it really significant? Eq. (6.5) is the original model for this data set, before transforming it with the procedure seen above to calculate the category scores:

$$Y = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{21}x_{21} \tag{6.5}$$

A3 is the reference level for item A, and B2 is the reference level for item B. We can use Minitab to estimate coefficients β_{11}, β_{12} and β_{21}. The Minitab output is the following:

```
The regression equation is
Y = 3.76 + 0.143 A1 - 0.540 A2 + 2.81 B1

Predictor      Coef   SE Coef      T       P     VIF
Constant     3.7619    0.2458  15.31   0.001
A1           0.1433    0.2692   0.53   0.631   1.469
A2          -0.5399    0.3205  -1.68   0.191   1.735
B1           2.8093    0.2692  10.43   0.002   1.469

S = 0.290812   R-Sq = 98.6%   R-Sq(adj) = 97.1%
```

The p-values for coefficients A1 and A2 are greater than the typical border of 0.05, so they can be declared not significant. Therefore item A has no effect on the response although its PCC is 0.781! The problem is that when we have few more individuals than regressors both MCC and PCC tend to increase, regardless of the significance of the items. p-value is, in this case, a more reliable measure of significance. My proposal to improve the output from QT1 is, thus, using the p-value for each item, and deciding if that item affects each Kansei word looking at this p-value[87].

However, the usual output from a regression analysis gives a p-value for each coefficient, and not for all coefficients belonging to the same item. Of course, having a p-value for each coefficient gives information on which coefficients are not significant and which are significant. We could have, for instance, a non significant coefficient A1 but a significant coefficient A2, with A3 being the reference level. That would mean that there are no differences in the response when factor A takes level A1 or level A3, but the response is different when factor A takes level A2.

But knowing if an item from the space of properties affects a specific Kansei is what really matters. If it affects the response, we can look then which categories of that item are more convenient. For having this we need just one p-value for each item. This can be achieved testing together a subset of the regression coefficients (Fox 1997, p.124), using the next procedure:

Imagine we are working with the following model that has k regressors:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q + \beta_{q+1} x_{q+1} + \cdots + \beta_k x_k + \varepsilon$$

We want to test the null hypothesis $H_0\colon \beta_1 = \beta_2 = \cdots = \beta_q = 0$ (this first q regressors could be any regressors in the equation). If the null hypothesis is correct, the "null" model would be $Y = \beta_0 + 0\,x_1 + \cdots + 0\,x_q + +\beta_{q+1} x_{q+1} + \cdots + \beta_k x_k + \varepsilon$.

Let $SS_{err\,0}$ and $SS_{reg\,0}$ represent, respectively, the residual and the regression sum of squares of the "null" model, and $SS_{err\,1}$ and $SS_{reg\,1}$ represent the residual and the regression sum of squares of the complete model. If the null hypothesis is true, this F statistic

$$F = \frac{SS_{reg\,1} - SS_{reg\,0}/q}{SS_{err\,1}/n-k-1} = \frac{n-k-1}{q} \cdot \frac{R_1^2 - R_0^2}{1 - R_1^2}$$

follows an F-distribution with q and n-k-1 degrees of freedom. R_1^2 and R_0^2 are the coefficient of determination of the complete and the "null" models, respectively. n is the number of individuals in the data set.

Going back to our last example, the p-value for item A will come out from testing $H_0\colon \beta_{11} = \beta_{12} = 0$ in Eq. (6.5). Figure 6.10 shows the results from QT1 with our

[87] At most, the partial correlation coefficient can be showed together with the p-value, but I think just the p-value is enough (and makes output more comparable with the output from other procedures, such as ordinal logistic regression).

example, but adding the p-value for each item. It is now clear that item A does not affect the response.

Figure 6.10. An improved version of the QT1 output, now with the p-value for each item

MCC = 0.9927

Item	PCC	p-value	Category	Category Score
A	0.781	0.243	A3	0.09
			A2	-0.45
			A1	0.24
B	0.987	0.002	B2	-1.20
			B1	1.61

Another way to calculate a p-value for each item is doing a test of significance on the partial correlation coefficient. This is the suggestion made by Álvarez (2009), following the proposal by Baba, Shibata & Sibuya (2004). If ρ_j is the partial correlation coefficient for item j, the following hypothesis test can be performed: $H_0\colon \rho_j = 0$ and $H_1\colon \rho_j \neq 0$. For this test, the statistic

$$t_{PCC} = PCC_j \sqrt{\frac{n-2-k}{1-PCC_j^2}}$$

approximately follows a t distribution with $n-2-k$ degrees of freedom. PCC_j is the sample partial correlation coefficient for item j, n is the number of stimuli and k is the number of conditioned items when calculating the partial correlation coefficient.

In all examples that I have tested, these p-values calculated for checking if the partial correlation coefficient is significant coincide up the fourth decimal with the p-values proposed above. The approach followed in this proposal is performing hypothesis testing on the coefficients as this is a more general approach that will be also used in ordinal logistic regression.

Finally, I propose a visual aid to quickly come up with those more significant items from a QT1 analysis. Bars from items with low p-values are drawn darker than those that have higher p-values. Figure 6.11 shows this last proposal, obtained by a R script that uses the ordinary `lm` function from the `base` package to calculate category scores, PCC and p-values.

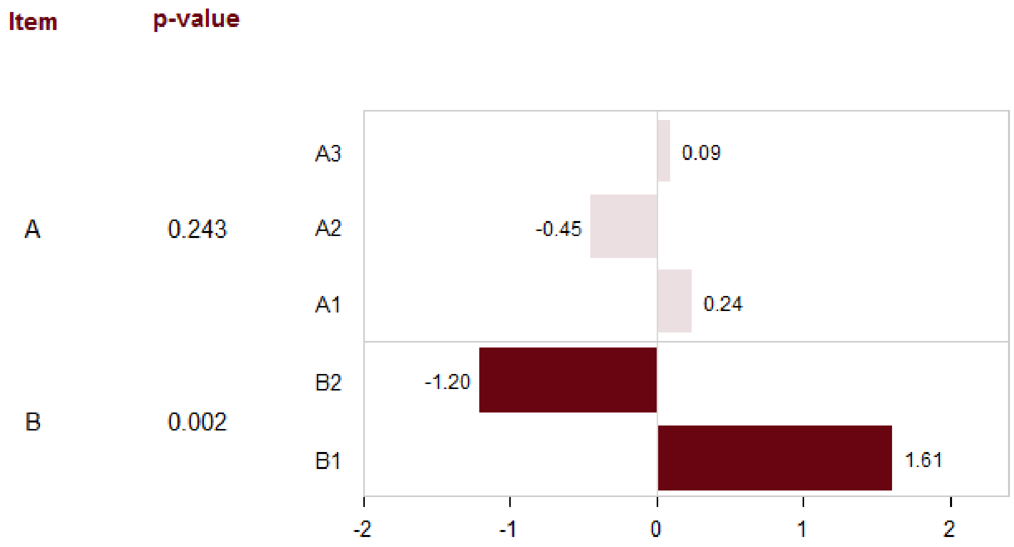


Figure 6.11. The final version of QT1, with p-values for each item and darker bars for significant items

The analysis of variance output from the general linear model command in Minitab provides, logically, the same p-values for the significance of items A and B (and the same R^2)

```
Analysis of Variance for Y. using Adjusted SS for Tests

Source  DF   Seq SS  Adj SS  Adj MS       F      P
A        2   8.0810  0.3979  0.1989    2.35  0.243
B        1   9.2074  9.2074  9.2074  108.87  0.002
Error    3   0.2537  0.2537  0.0846
Total    6  17.5421

S = 0.290812   R-Sq = 98.55%   R-Sq(adj) = 97.11%
```

In general, there is no reason to keep non-significant items in the model[88]. So models can be refitted until all items in the model are significant. Figure 6.12 shows the result of a QT1 analysis with the data from our T-shirts example using only the significant items.

[88] However, when introducing techniques as done in this chapter, I think it is sometimes didactic also showing non-significant factors in the model.

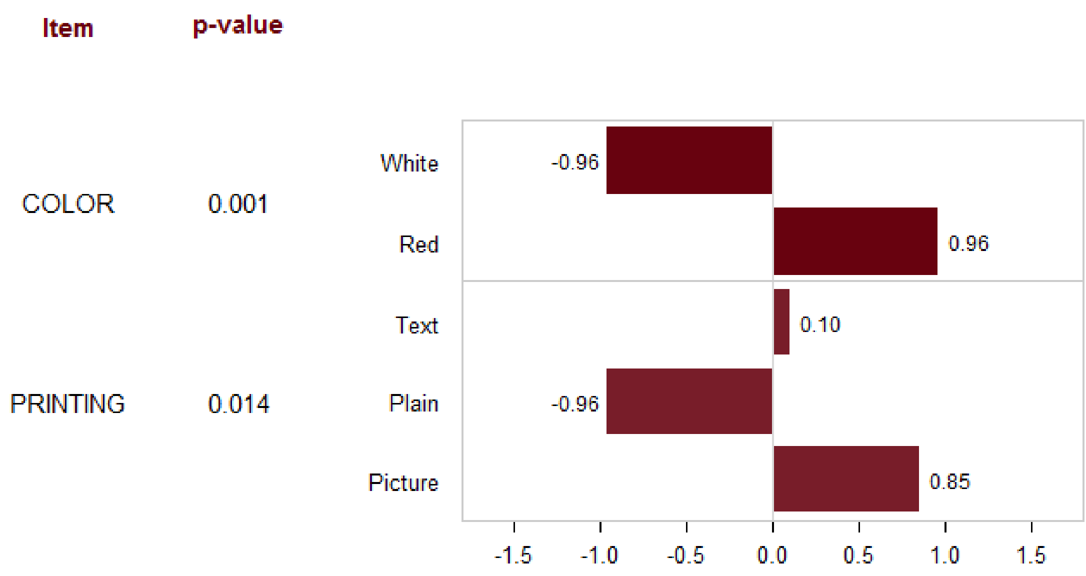


Figure 6.12. The results from QT1 analysis in the T-shirts example

Only Colour and Printing affect the Kansei colourful. Sleeves have no effect. Red color T-shirts are perceived as more colorful than white T-shirts. T-shirts with a picture on it are rated as more colorful than T-shirts with a text on it. Plain T-shirts with no printing on it are the less colorful.

Regression diagnostics are never reported in QT1. As we are performing a multiple regression analysis, I think some checking of the hypothesis of the model is recommended (at least, a plot of residuals versus fitted values, as the one in Figure 6.13).

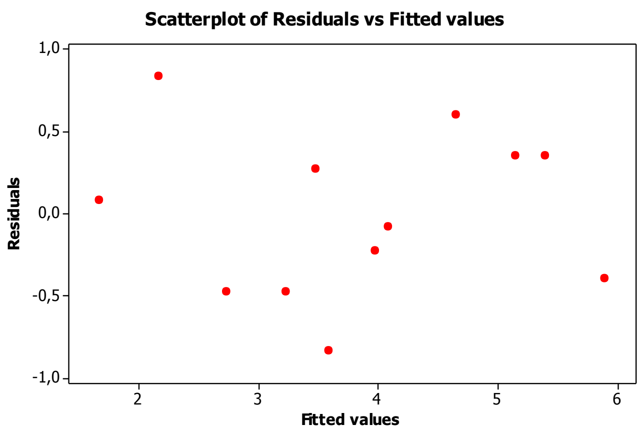


Figure 6.13. Scatterplot of residuals versus fitted values in the T-shirts example.

6.2.2. Ordinal Logistic Regression (OLR)

Raw data from KE studies is usually ordinal (integers from 1 to 7, or sometimes 1 to 5 or 1 to 9). It seems very natural analysing this response with ordinal logistic regression (OLR). The great advantage is that we avoid the need to work with means as in QT1, not losing information from the data. The disadvantage is that the output from an OLR is a bit harder to interpret than that from a linear regression (although a remedy to this inconvenience will be proposed).

Very few papers in the literature report the use of OLR as an analysis tool in the synthesis phase of a KE study. Furthermore, they all come from European researchers (van Lottum, Pearce & Coleman 2006, Barone, Lombardo & Tarantino 2007, Aktar Demirtas, Anagun & Koksal 2009). Working with logistic regression in Kansei Engineering is, thus, a rather original approach.

First step: binary logistic regression

There are some fundamental differences between a linear regression and an ordinal logistic regression. But before working with ordinal responses as the ones we have in Kansei Engineering, it is simpler to introduce logistic regression using binary responses. Imagine Y is a binary response (it can only take values 0 or 1) and X is the only predictor. $\pi(x) = E(Y|x)$ represents the conditional mean of Y given x.

$$\Pr(Y = 1|x) = \pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} \tag{6.6}$$

The conditional mean of the regression equation must be bounded between 0 and 1. $\pi(x)$ from Eq. (6.6) satisfies this restriction.

As Y can only take values 0 and 1, if we express $y = E(Y|x) + \varepsilon$, the error ε – that is, deviations from the value predicted by the model – is not normally distributed as in linear regression, but follows a binomial distribution.

Of course, our aim is estimating the parameters β_0 and β_1 in Eq. (6.6). This can be done using the regular maximum likelihood method. The idea is finding values for the β's that maximize the probability of obtaining the observed data. A likelihood function – traditionally named $L(\boldsymbol{\beta})$ – is built that expresses the probability of the observed data as a function of the unknown parameters. The values that maximize this function (called maximum likelihood estimators) are the chosen ones to estimate the β's. In fact, it is mathematically easier to work with the log of the likelihood function ($ln[L(\boldsymbol{\beta})]$). A clear deduction of the likelihood function for the case of binary logistic regression can be found in Hosmer and Lemeshow (2000). Fortunately we do not have to worry a lot about finding the maximum likelihood estimators, as software like R will promptly come in our help.

Taking Eq. (6.6), we can find a function of the response that produces a linear function of the independent variables. This function is called the link function in generalized linear models (McCullagh, Nelder 1989). In linear regression, the link function is the identity function, as the response is linear in the parameters. In logistic regression the link function is the logit transformation:

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = ln\left[\frac{\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}}{1-\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}}\right] = ln\left[e^{\beta_0+\beta_1 x}\right] = \beta_0 + \beta_1 x$$

$g(x)$ has the nice property of being lineal and it is called the linear predictor.

Imagine now the simple example described in Table 6.9. Four different persons give a dichotomous punctuation on a Kansei word (for example, they say if a watch is not modern: 0 or modern: 1). The space of properties is described by just one item (Shape), with 2 categories (Round: 0 and Rectangular: 1). There are only 2 stimuli (1 and 2)[89].

Table 6.9. A design matrix with 2 items (A and B)

Watch	Shape	x: Shape (coded)	Person	Y
1	Round	0	1	1
2	Rectangular	1	1	1
1	Round	0	2	0
2	Rectangular	1	2	1
1	Round	0	3	0
2	Rectangular	1	3	1
1	Round	0	4	0
2	Rectangular	1	4	0

The odds of rating the watch with rectangular shape (x = 1) as modern is $\frac{\pi(1)}{1-\pi(1)}$. Likewise, the odds of rating the watch with round shape (x = 0) as modern is $\frac{\pi(0)}{1-\pi(0)}$. The odds ratio (OR) is the ratio of the odds for x = 1 to the odds for x = 0:

[89] The idea of using watches as an example of KE study (something done in different parts of this work) is taken from the PhD thesis of Simon Schütte (2005).

$$OR = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}} \tag{6.7}$$

Now, going back to Eq. (6.6) and substituting it in Eq. (6.7):

$$OR = \frac{\frac{\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}{1-\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}}}{\frac{\frac{e^{\beta_0}}{1+e^{\beta_0}}}{1-\frac{e^{\beta_0}}{1+e^{\beta_0}}}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}$$

So, $OR = e^{\beta_1}$.

We can use function `glm` from the `R` package `stats` to estimate β_1 with the data from Table 6.9:

```
Call:  glm(formula = Y ~ Shape, family = binomial, data = invented.data1)

Coefficients:
     (Intercept)  ShapeRectangular
          -1.099             2.197

Degrees of Freedom: 7 Total (i.e. Null);  6 Residual
Null Deviance:      11.09
Residual Deviance: 8.997        AIC: 13
```

$\beta_1 = 2.197$ so $OR = e^{\beta_1} = e^{2.197} \cong 9$. How do we interpret this odds ratio? Watches with rectangular shape have an odds 9 times larger of being considered modern than watches with round shape. Notice that the reference level is the round shape.

Second step: ordinal logistic regression

When we have more than 2 ordered categories in the response, ordinal logistic regression is an appropriate choice. Ratings in KE are clearly ordered variables.

I will use the proportional odds model (also known with the name cumulative logit model) for modelling ratings in KE (McCullagh 1980). Suppose we have Kansei ratings from 1 to 7. We can compare category 1 to categories 2 through 7, or categories 1 and 2 to categories 3 through 7, or categories 1 through 3 to categories 4 through 7, etc. We cannot combine categories, say, 1 and 4 for comparison with categories 2, 3, 5, 6 and 7, as this would break the natural ordering of our categories (Kleinbaum, Klein 2002).

In general, if we have G ordered categories of the response Y (Y = 1, Y = 2, ..., Y = G), there are $G-1$ ways to dichotomise the response . In this situation, $odds(Y \leq g|x) = \Pr(Y \leq g|x)/\Pr(Y > g|x)$. The proportional odds model makes the assumption that the odds ratios assessing the effect of an independent variable X will be the same regardless of where the cutpoint is made. For example, $OR_{Y\leq 2} = \frac{odds(Y\leq 2|x=1)}{odds(Y\leq 2|x=0)}$ and $OR_{Y\leq 5} = \frac{odds(Y\leq 5|x=1)}{odds(Y\leq 5|x=0)}$ will be the same. The odds ratio is invariant to where the response categories are dichotomise. This implies that there is only one parameter β for each predictor (just a β_1 for x_1, a β_2 for x_2, etc.), the slope for each predictor is always the same. However, we have separate intercepts β_{0_g} for each of the $G-1$ comparisons.

That is:

$$\Pr(Y \leq g|x) = \pi_g(x) = \frac{e^{\beta_{0_g}+\beta_1 x}}{1+e^{\beta_{0_g}+\beta_1 x}}$$

and

$$\text{odds}(Y \leq g|x) = \frac{\Pr(Y \leq g|x)}{\Pr(Y > g|x)} = \frac{\frac{e^{\beta_{0_g}+\beta_1 x}}{1+e^{\beta_{0_g}+\beta_1 x}}}{1-\frac{e^{\beta_{0_g}+\beta_1 x}}{1+e^{\beta_{0_g}+\beta_1 x}}} = e^{\beta_{0_g}+\beta_1 x}$$

Therefore, the odds ratio is $OR_{Y\leq g} = \frac{odds(x=1)}{odds(x=0)} = \frac{e^{\beta_{0_g}+\beta_1}}{e^{\beta_{0_g}}} = e^{\beta_1}$, and not dependent on the cutpoint g. As we are interested in the slope β of each predictor and not in the intercepts, using a proportional odds model makes interpretation of the results more straightforward.

Table 6.10. A design matrix with 2 items (A and B)

	Shape	x: Shape (coded)	Person	Y1	Y2
1	Round	0	1	1	4
2	Rectangular	1	1	6	5
1	Round	0	2	3	3
2	Rectangular	1	2	7	4
1	Round	0	3	1	3
2	Rectangular	1	3	6	5
1	Round	0	4	1	2
2	Rectangular	1	4	2	2

Everything will hopefully become clear with the following example. Consider Table 6.10 where we have just one item (Shape) with 2 categories (Round and Rectangular). Four persons have rated a watch on the Kansei word modern, using a scale from 1 to 7. I will first concentrate the analysis in response Y1.

Function `polr` from the `R` package `MASS` (Venables, Ripley 2002) can be used to analyse response Y1 from Table 6.10:

```
Call:
polr(formula = Y ~ Shape, data = cbind(X, Y), Hess = TRUE)

Coefficients:
ShapeRound
       -3.720596

Intercepts:
      1|2       2|3       3|6       6|7
-2.714450 -1.860303 -1.006142  1.130398

Residual Deviance: 18.05278
AIC: 28.05278
```

Note that just one coefficient β_1 exists for item Shape, the reference level being Rectangular. The odds ratio is $e^{\beta_1} = e^{-3.72} = 0.024$. This result indicates that watches with round shape are 0.024 times as likely as watches with rectangular shape to have a rating of 7 versus a rating of 6 or less, and 0.024 times to have a rating of 6 or more versus a rating of 5 or less, or 0.024 times to have a rating of 5 or more versus a rating of 4 or less, etc. To summarize, round watches are about 0.024 times as likely to be rated as modern than rectangular watches (or just the opposite, rectangular watches are about 1/0.024 = 41.67 times as likely to be rated as modern than round watches).

There is no need to worry about the intercepts. Just notice that each intercept β_{0_g} is the log odds of $Y \leq g$ when all the independent variables are 0. It will always happen that $\beta_{0_g} \leq \beta_{0_{g+1}}$.

Significance of coefficients in ordinal logistic regression

As always in a statistical model, we cannot be satisfied just looking at a coefficient, but we must wonder if that coefficient is significant or not. Look at the following output from the statistical software SPSS reproducing the last analysis done in `R`:

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.
Threshold	[Y1 = 1]	1.006	1.124	.802	1	.371
	[Y1 = 2]	1.860	1.355	1.884	1	.170
	[Y1 = 3]	2.714	1.596	2.894	1	.089
	[Y1 = 6]	4.851	1.994	5.921	1	.015
Location	[Shape=Rectangular]	3.721	1.868	3.967	1	.046
	[Shape=Round]	0[a]	.	.	0	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Coefficient β_1 has a value of 3.721, the same as before, but we can now check that this coefficient is significant, as it has a p-value of 0.046 (less than the usual frontier of 0.05 to declare a coefficient as significant).

But, what happens when we go back to data in Table 6.10 and we take Y2 as response? The SPSS output is the following:

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.
Threshold	[Y2 = 2]	-.543	.999	.296	1	.587
	[Y2 = 3]	.693	1.013	.468	1	.494
	[Y2 = 4]	2.172	1.269	2.930	1	.087
Location	[Shape=Rectangular]	1.913	1.429	1.791	1	.181
	[Shape=Round]	0[a]	.	.	0	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Coefficient β_1 has now a p-value of 0.181, thus not being significant. Shape has no effect on the response; that is, shape has no effect on how modern the watch is perceived.

In this last exemple, Shape had only 2 categories. As one category is used as a reference level, only one coefficient exists (the other being 0), and just one p-value must be calculated. If the item has 3 categories, 2 coefficients must be computed thus having 2 p-values, one for each coefficient. As shown in the improved version of QT1, it is preferable and more understandable to have just one p-value for each group of coefficients corresponding to the same item. This can be achieved performing likelihood-ratio tests for the hypothesis that several coefficients are simultaneously 0 (H_0: $\beta_1 = \beta_2 = \cdots = \beta_q = 0$) (Fox 1997, p.450).

As in QT1, two models are fitted, the *full* model (model 1)

$$ln\left(\frac{\pi}{1-\pi}\right) = logit(\pi) = \beta_0 + \beta_1 x_1 + \cdots + \beta_q x_q + \beta_{q+1} x_{q+1} + \cdots + \beta_k x_k$$

and the *null* model (model 0):

$$ln\left(\frac{\pi}{1-\pi}\right) = logit(\pi) = \beta_0 + 0\, x_1 + \cdots + 0\, x_q + \beta_{q+1} x_{q+1} + \cdots + \beta_k x_k$$

In a proportional odds model, β_0 represents all the intercepts. Each model produces a maximized likelihood, L_1 for the *full* model and L_0 for the *null* model. The generalized likelihood-ratio test statistic is $G^2 = 2(\ln L_1 - \ln L_0)$. If H_0 is true, this statistics has an asymptotic chi-square distribution with q degrees of freedom.

A test for checking if all coefficients are 0 can be directly derived specifying a *null* model that only includes the intercepts. In this way, it is possible to have a general p-value for the whole model, as we had in QT1.

If desired, and in order to have a parallel output to that in QT1, some kind of R^2 of the regression can be defined. Imagine we have model *full*, the model with all predictors, and model *intercept*, the model without predictors, just with the intercepts. The quantity $G^2 = -2\ln L$, where L is the maximized likelihood, is called the deviance under the model, and is a generalization of the residual sum of squares in a linear model. McFadden's $pseudo - R^2$ is defined in the following manner:

$$R^2 = 1 - \frac{G^2_{full}}{G^2_{intercept}} = 1 - \frac{\ln L_{full}}{\ln L_{intercept}}$$

The log likelihood of the *intercept* model is treated as a total sum of squares, and the log likelihood of the *full* model is treated as the sum of squared errors. It is thus quite analogous to R^2 in a linear model. A small ratio of log likelihoods indicates that the *full* model is much better than the *intercept* model, and R^2 will then be high.

SPSS gives, beside the estimated coefficients, the global p-value checking if all slopes are 0 or not, and the value of McFadden's $pseudo - R^2$. This is part of the output for response Y1 from data in Table 6.10[90]:

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	16.164			
Final	10.310	5.854	1	.016

Pseudo R-Square

McFadden	.245

[90] All these indicators shown in the output from ordinal logistic regression in SPSS can be also calculated in `R`, taking the results from function `polr` from package `MASS`.

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	16.164			
Final	10.310	5.854	1	.016

Pseudo R-Square

McFadden	.245

A new visual version of OLR

The output from ordinal logistic regression suffers from the same difficulty of interpretation than the output of any regression model with dummy variables: the existence of reference levels. Additionally, the concept of odds ratio is somehow hard to interpret for not trained people.

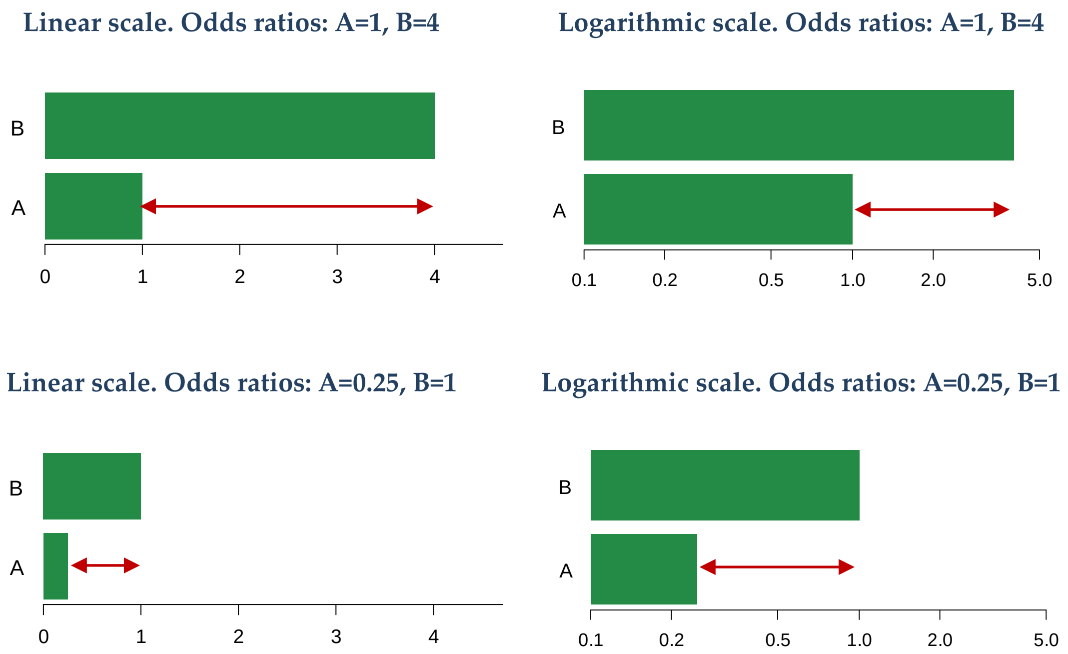


Figure 6.14. Comparison of barplots in linear and logarithmic scale

It would be nice to have a more visual output for OLR, making it more similar to the proposed version of QT1 (Figure 6.11) Odds ratios will play in OLR the role played by category scores in QT1. They will be drawn as bars in the graphic summarizing OLR. Obviously, the reference level will have an odds ratio of 1, but it will also appear in the graph. One difference with QT1 is that the horizontal axis will be logarithmic in OLR. This is necessary to avoid confusions. Imagine a fictitious situation where an item has 2 categories, A and B. A is the reference level and B has an odds ratio of 4. If B is used as reference level, the odds ratio for A would be ¼ =

0.25. If the scale is linear, the difference between level A and B could be perceived different depending on the reference level being used (Figure 6.14, left). If the scale used is logarithmic the difference between bars is the same in both cases, thus avoiding the problem (Figure 6.14, right).

For the data used in the example from Table 6.10 (watches that are rated on the Kansei word modern) using Y1 as response, the visual output from OLR is the one showed in Figure 6.15.

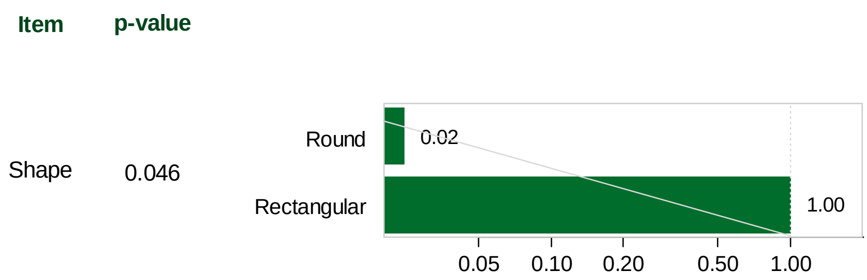


Figure 6.15. The new visual version of OLR

The graphic is very understandable and a wrong interpretation is not possible. Shape is a significant item (p-value below the usual border of 0.05). Watches with rectangular shape are perceived more modern than watches with round shape. If desired, the specific odds ratio value can be also read from the graph.

As with QT1, some kind of model fit assessment is recommended. In OLR, two commonly used goodness-of-fit tests are Pearson and Deviance. Pearson and Deviance are both types of residuals for logistic models (Hosmer, Lemeshow 2000). This is the SPSS output for the Y1 response in data from Table 6.10:

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	5.038	3	.169
Deviance	5.236	3	.155

Link function: Logit.

A low p-value indicates that the predicted probabilities deviate from the observed probabilities in a way that the model does not predict. In this case, p-values are sufficiently high.

As the visual output for OLR is based in the proportional odds model, a test of parallel lines can be performed to check if the assumption of having the same slope regardless of the category cut (and thus having constant odds ratios through all categories) is feasible. This test can be done calculating the proportional odds model and the corresponding polytomous logit model – that is, a model that does not have the constrain of having the same slope for all categories (Fox 2008, p.368). For the Y1 response in data from Table 6.10 the results are the following:

Model	Residual Deviance	Number of parameters
Proportional odds model	18.05	5
Polytomous logit model	12.82	8

The likelihood-ratio statistic for testing the assumption of proportional odds is $G_0^2 = 18.05 - 12.82 = 5.23$. A value of 5.23 provides a tail area on a chi-square distribution with 8-5=3 degrees of freedom of 0.16, leading us to keep the proportional-odds assumption for these data.

What can be done if the test of parallel lines does not support the proportional odds model? In theory, a polytomous logit model must be used. However, the interpretation of results is more complex in that case and the visual output suggested for OLR is not possible anymore. A proposed alternative is collapsing some categories and attempting the fit with a proportional odds model. For example, if ratings go from 1 to 7, categories 1 and 2 can be converted into 1, categories 3, 4 and 5 into 2, and categories 6 and 7 into 3. If all categories are collapsed into just 2, a binary logistic regression must be fitted, thus surely having just one slope for each regressor (and just one intercept for the whole model). More on collapsing categories will be explained in Section 6.2.3.

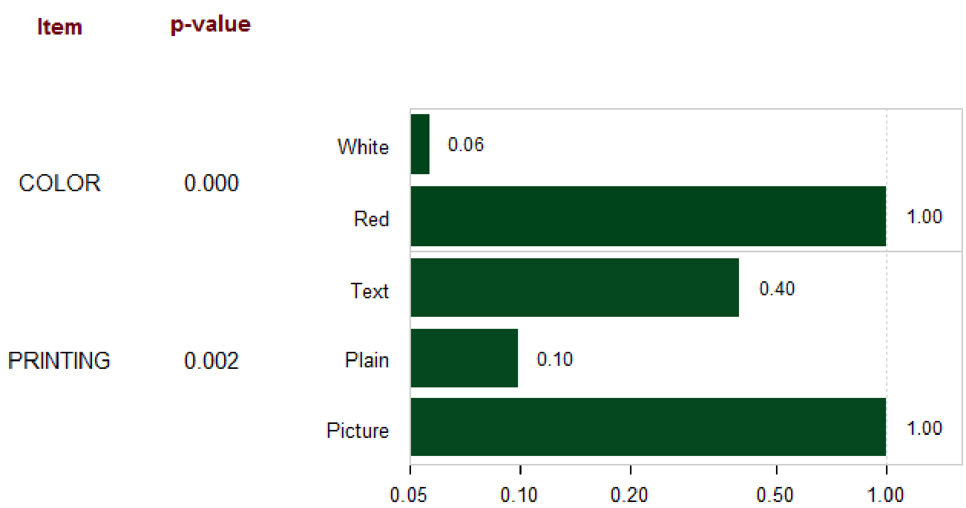


Figure 6.16. The final new version of OLR output with data from the T-shirt example

Figure 6.16 shows the final version of the output from OLR with data from the T-shirts example. Figure 6.17 compares the same output from SPSS (on the left) and the new visual output (on the right). Interpretation with the visual output is very simple, and with the guarantee that even novices in the use of OLR will come to the right conclusions.

SPSS Output

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Colorful = 1.00]	-3.994	.801	24.849	1	.000	-5.565	-2.424
	[Colorful = 2.00]	-3.173	.725	19.176	1	.000	-4.594	-1.753
	[Colorful = 3.00]	-2.422	.667	13.188	1	.000	-3.729	-1.115
	[Colorful = 4.00]	-.943	.580	2.646	1	.104	-2.079	.193
	[Colorful = 5.00]	.826	.581	2.020	1	.155	-.313	1.966
	[Colorful = 6.00]	2.181	.748	8.513	1	.004	.716	3.646
Location	[Color=Red]	-2.785	.644	18.707	1	.000	-4.046	-1.523
	[Color=White]	0[a]	.	.	0	.	.	.
	[Printing=Picture]	.879	.645	1.857	1	.173	-.386	2.144
	[Printing=Plain]	-1.384	.666	4.322	1	.038	-2.688	-.079
	[Printing=Text]	0[a]	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

The proposed visual output

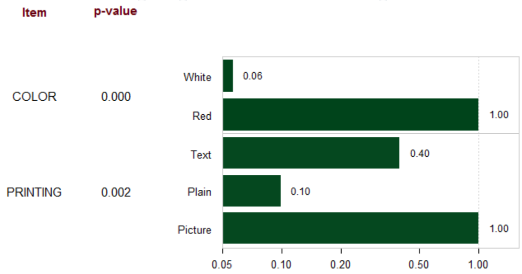


Figure 6.17. Comparison between the SPSS Output (left) and the proposed visual output (right) for OLR with data from the T-shirst example

6.2.3. Mixed effects ordinal logistic regression (mOLR)

Although an ordinal logistic regression seems more appropriate than a linear regression for the kind of response we have in KE studies, there is an important detail not yet taken into account. The response for each Kansei word is a rating, but each participant in the study give ratings for a set of stimuli. When performing an ordinal logistic regression, we suppose all responses are independent. But this supposition is clearly violated by the kind of data we have: all ratings from the same subject are somehow closer among them to ratings from another subject. Therefore, data must be treated as coming from different "clusters of variability": the subjects.

This fact is always neglected in Kansei Engineering studies: no papers from the literature review in Chapter 5 take it into account. The problem is not only unsolved, it is not even stated. How can this subject-clustered data be treated? There are two possibilities:

1. Incorporating the subject as a new categorical variable in the model. Subject is converted into several dummy variables (as many as different subjects we have, minus one; this last subject is taken as reference). Each subject has then a constant shift in the response with respect to the reference subject. In this approach, it is said that the subject is incorporated as a fixed effect.
2. Having a subject-specific term in the model that takes the same value for each rating belonging to the same subject, but treated as varying randomly among

a sample of subjects. We assume here that the subjects are a random sample of the whole population, and the aim of the inference is about the overall properties of the population. In this approach, it is said that the subject is incorporated as a random effect. In fact, it does not make sense to estimate the random effect, but to estimate the parameters that describe the distribution of this random effect.

The second approach, incorporating the subject as a random effect, seems closer to the real situation we are facing: participants in our KE study are simply a random sample from all the possible subjects that belong to the target group. Moreover, treating subjects as fixed effects implies adding many terms to the model (the number of subjects minus one), and this is very inconvenient.

Consequently, my proposal is incorporating subjects into an ordinal logistic regression as random effects. As the model combines fixed effects (the factors in the KE study) and random effects (the subjects), it is a mixed-effects ordinal logistic regression. The use of random effects when the response is continuous and thus a linear model is fitted is well established (these models are called mixed-effects linear models)[91]. However, introducing random effects for categorical responses is not so common, as it greatly complicates computation for model fitting (or perhaps I should say has not been common until now, as more and more computer routines exist – and still appear – that allow the estimation of mixed-effects generalized linear models)[92].

Generalized linear mixed models (GLMM) extend the generalized linear models (GLM) allowing the inclusion of random effects – as well as fixed effects – in the linear predictor. y_{it} denotes observation t in cluster i, x_{it} is the value of the explanatory variable for that observation, and u_i denotes the random effect for cluster i. Following the nomenclature used in Agresti (2007), and having $g(\cdot)$ as link function, $g(\mu_{it}) = \beta x_{it} + u_i$. Conditional on u_i, a GLMM resembles a GLM.

I will use ordinal logistic regressions that incorporate random effects for subjects as an intercept term. This is the most simple mixed ordinal logistic regression. u_i is supposed to have a $N(\alpha;\sigma)$ distribution. α can be explicitly entered in the linear predictor, taking u_i to have a $N(0;\sigma)$ distribution. For a given predictor values, the

[91] Menu driven statistical packages such as Minitab or SPSS allow the estimation of mixed effects linear models. Procedure MIXED in SAS can also be used. My personal choice is function `lmer` from the `R` package `lme4`.

[92] Procedure GENMOD and NLMIXED in SAS can be used for mixed-effects generalized linear models. In `R`, several functions are also available: `glmmPQL` from package `MASS`, `glmmML` from packag.e `glmmML`, `MCMCglmm` from package `MCMCglmm`. Specific for mixed-effects ordinal logistic regression, function `clmm` from package `ordinal` allows the inclusion of one random effect in a proportional odds model.

maximum likelihood fitting procedure treats the observations as independent, but conditional on the u_i. Not only fixed effects are estimated (usually the ones of more interest), but also the variability σ^2 among clusters (subjects in our case).

To obtain the maximum likelihood function in GLMM, the routine eliminates u_i by , first, forming the likelihook function as if the u_i values were known, and later averaging this function with respect to the $N(0;\sigma)$ distribution of u_i. The integrals needed in the whole process do not have a closed form, and there is a need of numerical methods that can be computationally intensive. [93]

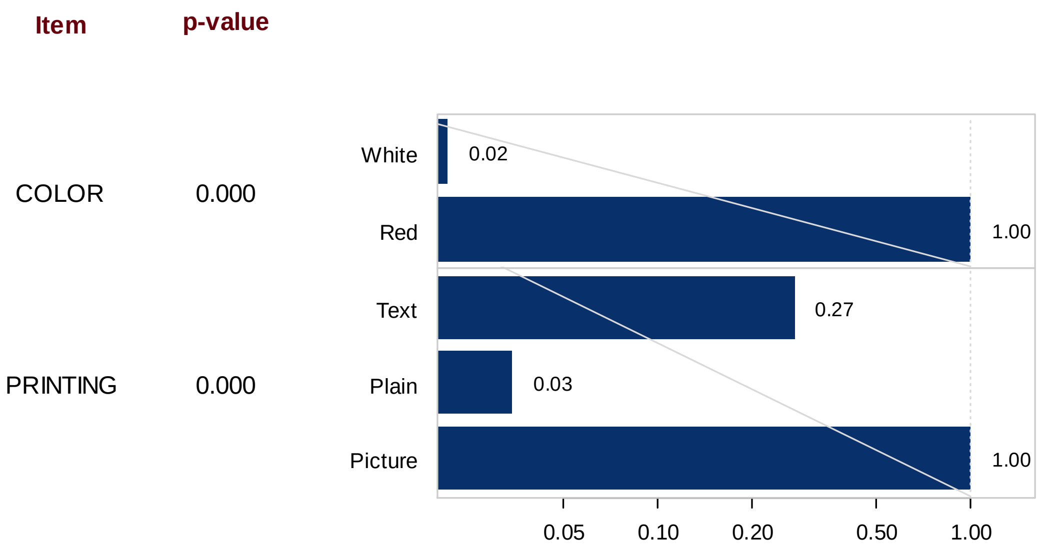


Figure 6.18. The final version of mixed OLR output with data from the T-shirt example

So my final proposal for analyzing KE data in the synthesis phase is using a proportional odds model ordinal logistic regression that incorporates subjects as random effects.[94] It is still possible to use the visual output previously suggested: odds ratios are represented as bars, and the variability among subjects is estimated, but it is of no interest.[95] As the subjects' random effect is now considered, odds ratios tend to be higher now (or lower if they are below 1), although there are usually no big differences in significance (Agresti 2002).

---

[93] However, computationally intensive only means waiting more for the result (it requires "machine" effort, but not "human" effort). So if the amount of time is reasonable, there is no need to worry.

[94] I have implemented an R function for performing this mixed ordinal logistic regression, using the excellent function `clmm` from package `ordinal`.

[95] Although I am not considering variability among subjects now, I think heterogeneity among subjects is a very important issue in Kansei Engineering, and it will be covered in detail in Chapter 8.

Figure 6.18 shows the results for the T-shirts example. The conclusions are the same than in the common OLR (Figure 6.16), but differences shown by odds ratios are accentuated (conclusions are not always exactly the same when using OLR and mOLR, as analysis of data from the juice experiment in Chapter 7 will show).

6.3. Rough sets in the synthesis phase

As explained in the previous sections, the statistical approach in the synthesis phase of KE studies implies building regression models, having Kansei words as responses and each product property (each factor) as independent variables. The output from these statistical models always includes:

- Some test of significance for each coefficient in the model, so we can check if a factor really has an effect in the response.
- An estimation of each coefficient, so we know in which direction the response moves depending on the level of each significant factor.

If we correctly obtained both issues stated above, we have successfully linked the semantic space with the space of properties. But some KE studies use another approach in this synthesis phase, as we have seen in Chapter 5: the automatic learning approach. The idea is not using any statistical model, just inferring rules from the data. This is sometimes named with the appealing term *noninvasive data analysis* (Pedrycz, Vukovich 2004). When using a regression model, a series of constrains and requirements are imposed on the data; noninvasive methods, on the contrary, assume a minimal level of constrains.

As I find the term noninvasive method a bit pompous and confusing, I will use automatic learning instead[96]. Rough sets is a commonly used automatic learning method. There are applications of rough sets in many different areas of knowledge: medicine (Czerniak, Zarzycki 2003, Tsumoto, Ziarko 1996), economy (Tay, Shen 2002), linguistics (Rapp, Jessen & Dogil 1994), computer science (Chen-Jimenez, Kornecki & Zalewski 1998), and many others. In fact, almost every time a problem can be faced building a regression model, rough sets can also be used. In KE studies, rough sets have gained popularity after its introduction by the team lead by Prof. Nagamachi in Japan – and especially due to the research of Tatsuo Nishino (Nishino, Nagamachi & Tanaka 2005, 2006, Nishino et al. 2007, 2008)).

[96] In fact, I am not totally happy with the term automatic learning. Anyway, many techniques commonly used in data mining could be located under this term (neural networks, clustering, genetic algorithms, decision trees…)

As rough sets is a rule induction system, discretization of data is required. Most inductive learning methods require a small number of attribute values. When we have a dataset with continuous variables, it is necessary to convert these continuous variables into discrete attributes. So discretization is a technique used in many areas besides rough sets, such as bayesian networks and classification and regression trees (CART) (Grzymala-Busse, Ziarko 2000).

In Kansei Engineering, factors (items in QT1 nomenclature) are already discrete. But the response can be either continuous, when using the average of all ratings, or discrete (but probably with many categories), when using the direct rating from each participant. Although several algorithms can be used to discretize data, my proposal for rough sets in KE studies is one the following:

1. **Discretization 1 (Disc1)**: converting the numbers into 3 categories (or perhaps 2 or 4, but clearly no more than 4) according to some equivalence table. If we decide to discretize numbers into 3 categories, and ratings have been done on a 7-point scale, the following conversions could be used:
    - When using the mean of all ratings from each participant:
      $1 \leq \text{mean} \leq 2.5 \rightarrow 1$ / $2.5 < \text{mean} < 5.5 \rightarrow 2$ / $5.5 \leq \text{mean} \leq 7 \rightarrow 3$
    - When directly using the ordinal responses from each participant:
      1 or 2 → 1 / 3, 4 or 5 → 2 / 6 or 7 → 3

    If 3 categories are used, I find it reasonable using quite extreme values for category 1 and 3[97].

2. **Discretization 2 (Disc2)**: converting the numbers into just 2 categories: category 0 for numbers below the mean of all values and category 1 for numbers above this mean.

    Some numbers fall in category 0 and some others in category 1 (probably half and half in each category, or something not very far away from this). In this manner, the probability of finding significant effects is maximized.

[97] This is, in fact, the idea conveyed by business author Fred Reichheld in his book "The ultimate question" (Reichheld 2006). In this book about customer loyalty the proposal is asking customers a single question on a 0 to 10 rating scale: "How likely is it that you would recommend our company to a friend or colleague?" Based on their responses, customers are categorized into one of three groups: Promoters (9-10 rating), Passives (7-8 rating), and Detractors (0-6 rating). Only the two highest categories (from 11 available) are used for labelling a customer as a promoter. If modern is a Kansei word, only lowest categories should be used to consider the stimulus as not modern (1) and highest categories to consider the stimulus as modern (3). Categories in the middle are assigned to category 2 (something neutral)

Table 6.11 presents three different possible ratings on a set of 8 stimuli, and will be used to exemplify both discretizations just suggested. The mean of all responses from participants is used (the same could be done if using the direct ordinal ratings from participants).

In situation 1, some stimuli get a low rating and some others a high rating. Thus categories 1, 2 and 3 appear in Disc1. Category 3 in Disc1 is converted into category 1 in Disc2, whereas category 1 in Disc1 is converted into category 0 in Disc2.

In situation 2, all stimuli get a middle rating but one (stimulus 6), that gets a high rating. All stimuli fall in category 2 in Disc1 but stimulus 6, which falls in category 3. On the contrary, the proportion of category 0 and 1 in Disc2 is quite balanced.

In situation 3, all stimuli get low ratings. Therefore, all stimuli fall in category 1 in Disc1, but both categories 0 and 1 are present in Disc2.

Table 6.11. A design matrix with 2 items (A and B)

	Situation 1			Situation 2			Situation 3		
	Mean	Discr1	Discr2	Mean	Discr1	Discr2	Mean	Discr1	Discr2
Stimulus 1	5.6	3	1	3.8	2	0	1.2	1	0
Stimulus 2	3.2	2	0	2.6	2	0	2.4	1	1
Stimulus 3	6.4	3	1	3.5	2	0	1.3	1	0
Stimulus 4	1.9	1	0	4.1	2	1	2.0	1	1
Stimulus 5	4.1	2	1	2.9	2	0	1.7	1	0
Stimulus 6	2.3	1	0	6.3	3	1	2.4	1	1
Stimulus 7	5.2	2	1	4.5	2	1	2.3	1	1
Stimulus 8	3.7	2	0	3.6	2	0	1.5	1	0
Global mean	4.0500			3.9125			1.8500		

When there are significant differences among stimuli, although being small differences, Disc2 will produce stimuli in different categories (as in situation 3 in Table 6.11). This is similar to what happens when using the statistical approach. As our purpose is having an output from rough sets that resembles the conclusions from the regression approach, Disc2 will be used when working with rough sets.

Two flavors of rough sets commonly used in KE studies will be studied in this section:

- Original rough sets (OrigRS) is the usual rough set procedure. It will be used when working with the average of all participants' ratings.
- Variable precision bayesian rough sets (VPBRS) is a modified version of rough sets used when working with the direct ratings of each participant.

6.3.1. Original rough sets (OrigRS)

Although I use the term original rough sets (OrigRS) to distinguish this procedure from variable precision bayesian rough sets (VPBRS) in the scope of this dissertation, OrigRS is simply called rough sets in literature. Rough sets were introduced by Pawlak Zdzisław in 1982 (Pawlak 1982). Although its mathematical formulation can be a bit complex in the beginning, the idea is simple and can be easily understood with an example. So let's start with an example...

Imagine a small KE study on watches with only 4 stimuli (the 4 rows in Table 6.12). Two factors are used in the study: shape of the watch and color of the face. The Kansei word modern is evaluated, having just two categories at the end: not modern and modern.

	Shape	Color	Modern
Watch 1	Round	White	No
Watch 2	Rectangular	White	Yes
Watch 3	Round	Brown	Yes
Watch 4	Rectangular	Brown	Yes

Table 6.12. A simple KE study on four watches, evaluating the Kansei word modern.

Which rules can be extracted from this table? Just looking at this simple dataset, we can see that:

- When watches have a rectangular shape, they are perceived as modern. When watches have a round shape, they can be either perceived as modern or as not modern.
- When watches have brown face, they are perceived as modern. When they have white face, they can be perceived either as modern or as not modern.

So the following two exact rules can be extracted:

```
SHAPE: Rectangular → Modern
COLOR: Brown → Modern
```

Watches simultaneously having round shape and white color are perceived as not modern. So the last exact rule is:

```
SHAPE: Round / COLOR: White → Not Modern
```

The purpose of rough sets is extracting these rules automatically, using algorithms. The output obtained is radically different from statistical regression models, but it has the advantage that no requirements on the data are made.

Basic mathematical formulation of rough sets[98]

Consider an information system $S = \{U, A, V, f\}$ where U is a non-empty finite set of N objects $\{x_1, x_2, \dots, x_N\}$, A is a non-empty finite set of r attributes $\{a_1, a_2, \dots, a_r\}$ used to describe the objects, V is a non-empty finite set of all attributes' values $V = \bigcup_{a \in A} V_a$ with V_a the possible values of attribute a, and $f: U \times A \to V$ is a mapping function that gives the value of an attribute to an object: $f(x, a) \in V_a, \forall a \in A, \forall x \in U$. Objects in rough sets are stimuli in KE, attributes in rough sets are factors or properties in KE and attributes' values are levels of a factor.

Using data from the first two columns of Table 6.12 (the columns in red, named Shape and Color) as an example, we have:

$$
\begin{aligned}
&U = \{Watch\ 1, Watch\ 2, Watch\ 3, Watch\ 4\} \\
&A = \{Shape, Color\} \\
&V_{Shape} = \{Round, Rectangular\} \\
&V_{Colour} = \{White, Brown\} \\
&V = \{Round, Rectangular, White, Brown\}
\end{aligned}
$$

In an information system, consider $B \subseteq A$ a subset of attributes and objects $x, y \in U$. A binary relation can be defined which will be called B-indiscernibility relation:

$$IND(B) = \{(x, y) \in U \times U \mid \forall a \in B, f(x, a) = f(y, a)\}$$

For example, with data from the first two columns of Table 6.12, watch 1 and watch 3 are indiscernible respect to the attribute Shape (they are Shape-indiscernible), because both of them have round shape. The same happens with watch 2 and watch 4, both having rectangular shape. So an indiscernibility relation splits the universe U of stimuli in a family of equivalence classes $\{X_1, X_2, \dots, X_k\}$. In our example, $IND(Shape)$ defines two equivalence classes, $X_1 = \{Watch\ 1, Watch\ 3\}$ and $X_2 = \{Watch\ 2, Watch\ 4\}$.

So $IND(Shape) = \{X_1, X_2\} = \big\{\{Watch\ 1, Watch\ 3\}, \{Watch\ 2, Watch\ 4\}\big\}$.

Similarly, $IND(Color) = \{X_3, X_4\} = \big\{\{Watch\ 1, Watch\ 2\}, \{Watch\ 3, Watch\ 4\}\big\}$.

Classes of equivalence $\{X_1, X_2, X_3, X_4\}$ are the elementary concepts in our information system. Taking $B = A$,

$$IND(\{Shape, Color\}) = \{X_5, X_6, X_7, X_8\} = \big\{\{Watch\ 1\}, \{Watch\ 2\}, \{Watch\ 3\}, \{Watch\ 4\}\big\}$$

---

[98] This subsection is based on information and examples extracted from Álvarez (2009) and Álvarez et al. (2001)

and classes of equivalence $\{X_5, X_6, X_7, X_8\}$ are the basic concepts in our information system.

Some information systems can be defined as decision tables. The set of attributes A is divided in two different disjoint subsets $A = C \cup D$. C is the set of conditional attributes and D the set of decision attributes, $C \cap D = \emptyset$. In our example, $C = \{\{Shape\}, \{Color\}\}$ and $D = \{\{Modern\}\}$. In KE studies, data is represented in decision tables $DT = \{U, C \cup D, V, f\}$, where U are all the stimuli, C the factors (Shape and Color, for example) and D the response from a Kansei word (Modern, for example). Although in rough sets D can include many decision attributes, in KE only one decision attribute (one Kansei word) is considered.

Why do rough sets have this name? I will explain the idea with a slight modification of the watches example being used in this section. Table 6.13 is the same as Table 6.12 but with an added watch. This last watch (Watch 5) have the same shape and color than Watch 3.

	Shape	Color	Modern
Watch 1	Round	White	No
Watch 2	Rectangular	White	Yes
Watch 3	Round	Brown	Yes
Watch 4	Rectangular	Brown	Yes
Watch 5	Round	Brown	No

Table 6.13. A simple KE study on five watches (last one added), evaluating the Kansei word modern.

We can calculate all equivalence relations considering attributes in C (Shape and Color):

$$IND(Shape) = \{\{Watch\ 1, Watch\ 3, Watch\ 5\}, \{Watch\ 2, Watch\ 4\}\}$$
$$IND(Color) = \{\{Watch\ 1, Watch\ 2\}, \{Watch\ 3, Watch\ 4, Watch\ 5\}\}$$
$$IND(\{Shape, Color\}) = \{\{Watch\ 1\}, \{Watch\ 2\}, \{Watch\ 3, Watch\ 5\}, \{Watch\ 4\}\}$$

Let $S = \{U, C \cup D, V, f\}$ be an information system, $X \subseteq U$ a subset of stimuli and $R = IND(B)$ an equivalence class, with $B \subseteq C$. X is R-exact if X is the union of some R-basic concepts, otherwise it is R-rough.

In the example from Table 6.13 we are interested in describing watches being perceived as not modern, that is, the set $X = \{Watch\ 1, Watch\ 5\}$. Consider $R = IND(\{Shape, Color\})$. The union of some basic concepts from R (Figure 6.19, a) cannot form the set X, so X is a rough set (Figure 6.19, b)

A rough set can be approximated using the following sets:

- The lower approximation of X: $\underline{R}X = \bigcup\{Y \in R \mid Y \subseteq X\}$ (Figure 6.19, c)

- The upper approximation of X: $\overline{R}X = \bigcup\{Y \in R \mid Y \cap X \neq \emptyset\}$ (Figure 6.19, d)
- The boundary of X: $BND_R(X) = \overline{R}X - \underline{R}X$ (Figure 6.19, e)

Figure 6.19 summarizes the example just described. The rough set $X = \{Watch\ 1, Watch\ 5\}$ to approximate is painted in red. Lower and upper approximations, and the boundary, are painted in green.

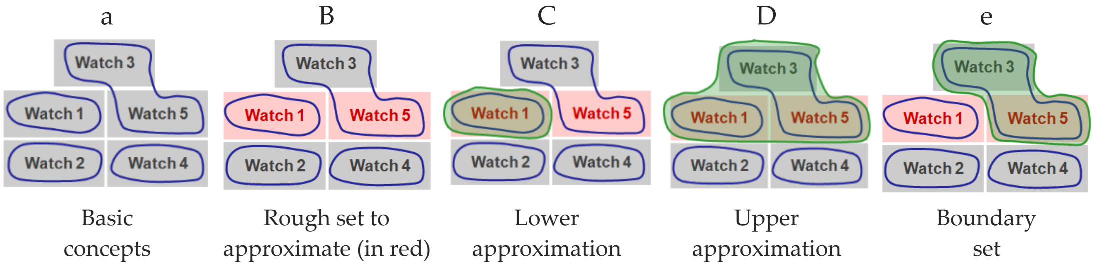


Figure 6.19. Rough sets approximations for the example using data in Table 6.13

$\underline{R}X = \{Watch\ 1\}$ is the R-positive region $POS_R(X)$ of X in S. It includes all watches in U that can be classified in decision attribute not modern (X) using the information in condition attribute set R. Watch 1 is surely perceived as not modern.

$U - \overline{R}X = \{Watch\ 2, Watch\ 4\}$ is the R-negative region $NEG_R(X)$ of X in S. It includes all watches in U that cannot be classified in decision attribute not modern (X) using the information in condition attribute set R. That is, Watch 2 and Watch 4 are surely perceived as modern.

X is R-exact when $\overline{R}X = \underline{R}X$. As design matrices in KE studies (conditional attributes C in rough set terminology) do not usually repeat any row, response categories in the Kansei word can be usually expressed as the union of basic concepts, thus being exact.

The rough sets theory includes some other concepts, such as the definition of reducts and core, which will not be used in this dissertation.

Rule extraction and calculation of indexes

The interest rough sets has attracted in the field of KE is due to the fact that the technique can be used for extracting rules with no further requirements on the data. Using the example from Table 6.13, the process of rule extraction can be summarized in the following manner:

1. Considering all attributes (all factors) C, all possible equivalence classes $R = IND(B)$ with $B \subseteq C$, are calculated. With the example in Table 6.13, $IND(Shape)$, $IND(Color)$ and $IND(\{Shape, Color\})$ are calculated.

2. For each equivalence class, rules are extracted looking at the decision attribute (the response). Consider for example $IND(Shape)$:
   $IND(Shape) = \{\{Watch\ 1, Watch\ 3, Watch\ 5\}, \{Watch\ 2, Watch\ 4\}\}$

   When $X \in \{Watch\ 1, Watch\ 3, Watch\ 5\}$, Shape = round, but the response can be either Modern = yes or Modern = no. The extracted rule is, thus:
   `Shape = round → Modern = no OR Modern = yes`
   This is a non-deterministic rule.

   When $X \in \{Watch\ 2, Watch\ 4\}$, Shape = rectangular, and the response is always Modern = yes. The extracted rule is, thus:
   `Shape = rectangular → Modern = yes`
   This is a deterministic rule.

3. The process of rule extraction continues for all equivalence classes.
   Rules extracted from equivalence classes with more attributes already considered by rules extracted from equivalence classes with less attributes are removed. For example, when considering equivalence class $IND(\{Shape, Colour\})$ the following two rules are extracted:
   `Shape = rectangular AND Color = white → Modern = yes`
   `Shape = rectangular AND Color = brown → Modern = yes`
   These rules are removed because they are already considered by the more general rule `Shape = rectangular → Modern = yes`

There are several algorithms for rule extraction in rough sets, LEM2, MODLEM or EXPLORE are among them. Software such as ROSE2, developed in the Institute of Computing Science in Poznan University of Technology, implements rough set theory and rule discovery techniques (Predki et al. 2009).

Final rules from the example in Table 6.13 are:

- Deterministic rules:
  `Shape = rectangular → Modern = yes`
  `Shape = round AND Color = white → Modern = no`
- Non-deterministic rules:
  `Shape = round AND Color = brown → Modern = no OR Modern = yes`

As non-deterministic rules do not help a lot in discovering how factors affect the response in Kansei Engineering, only deterministic rules will be considered. An `R` function has been developed that extracts deterministic rules from KE datasets.

There are no goodness-of-fit indicators in rough sets. However, three indexes are defined that give information for each rule from a rough sets analysis. Imagine that we have a list of m rules. The rules have the structure: $cond_k \rightarrow D_j$, with $k = 1, \ldots, m$

and $j = 1, \dots, l$ (l is the number of different categories the response has). The decision table has a set U of N stimuli $\{x_1, x_2, \dots, x_N\}$

Consider a rule $cond_k \rightarrow D_j$. The following three indexes are defined ($|\cdot|$ denotes cardinality):

Certainty	Coverage	Strength
$cer(cond_k; D_j) = \dfrac{\lvert cond_k \cap D_j \rvert}{\lvert cond_k \rvert}$	$cov(cond_k; D_j) = \dfrac{\lvert cond_k \cap D_j \rvert}{\lvert D_j \rvert}$	$str(cond_k; D_j) = \dfrac{\lvert cond_k \cap D_j \rvert}{\lvert U \rvert}$

Taking again the example in Table 6.13, these are the three rules with its indexes:

		cer	*cov*	*str*
1	Shape=rectangular → Modern=Yes	2/2 = 1.00	2/3 = 0.67	2/5 = 0.4
2	Shape=round AND Color=white → Modern=No	1/1 = 1.00	1/2 = 0.50	1/5 = 0.2
3	Shape=round AND Color=brown → Modern=No	1/2 = 0.50	1/2 = 0.50	1/5 = 0.2
	Shape=round AND Color=brown → Modern=Yes	1/2 = 0.50	1/3 = 0.67	1/5 = 0.2

Deterministic rules have an index $cer = 1$. When $cer < 1$, the rule is non–deterministic. Only looking at rules with $cer = 1$ implies only considering the lower approximation when the set is rough.

Table 6.14 shows the extracted rules using rough sets with the dataset in Table 6.2 (the T-shirts example). The response (colorful) has been coded according to **Disc2** (just 2 categories: below average and above average).

Table 6.14. Extracted rules using original rough sets for data in Table 6.2

		cer	*cov*	*str*
COLOR: White	0	1.000	0.857	0.500
COLOR: Red / SLEEVES: Short	1	1.000	0.600	0.250
COLOR: Red / PRINTING: Picture	1	1.000	0.400	0.167
COLOR: Red / PRINTING: Text	1	1.000	0.400	0.167
SLEEVES: Long / PRINTING: Plain	0	1.000	0.286	0.167

Rules are ordered from highest to lowest strength. Rules with highest strength correctly capture the effect of Color (setting aside the factor Sleeves): white gives a perception of not colorful, whereas red gives a perception of colorful. The effect of printing is somehow more difficult to capture, although the general trend is correctly shown: T-shirts with pictures and T-shirts with text are perceived as more colorful than plain T-shirts.

Factor Sleeves appears in some rules, although it is an inert factor. This sometimes happens in rough sets, making interpretation of results more complex and somehow confusing.

A small enhancement in the presentation of rules from rough sets is achieved with the use of color: rules with highest strength indexes have darker colors than rules with lower strength indexes. This facilitates reading the results and only focusing in the most important rules.

Adding tests of significance for rules in rough sets

Although just 5 rules appeared when using rough sets with the T-shirts example dataset, the list of rules is usually much longer when working in KE studies. This frequently huge list of rules makes interpretation of results difficult, to say the least. The common procedure is sorting rules according to some criteria (normally strength), and just looking at the first rules. The cutpoint is, however, arbitrary.

A method is here proposed to perform tests of significance on each rule, so that a p-value can guide in the decision of considering if a rule is unlikely to have occurred by chance[99]. The method is illustrated using the invented data from Table 6.12. Remember that this dataset only has 4 stimuli (watches), and two factors: Shape (with levels Round and Rectangular) and Color (with levels White and Brown). The response is modern (1) or not modern (0). The 4 left-most columns of Table 6.15 reproduce the dataset.

Table 6.15. A simple KE study on five watches (last one added), evaluating the Kansei word modern.

	S	C	Y	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16
W1	Ro	Wh	0	0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
W2	Re	Wh	1	0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
W3	Ro	Br	1	0	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
W4	Re	Br	1	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1
strength rule: S = Re →1			0.50	0	0	0.25	0	0.25	0.25	0	0.25	0.25	0.50	0.25	0.25	0.50	0.25	0.50	0.50

The response vector {0,1,1,1} (in red in Table 6.15) is just one of the 16 possible responses. These 16 possible responses are shown in the 16 right-most columns of

[99] This idea of performing tests of significance for extracted rules in rough sets is also outlined in the book "Noninvasive data analysis: a web of information granules" (Pedrycz, Vukovich 2004). The method proposed, however, is different from the method created in this dissertation.

Table 6.15. Consider now the first rule that can be extracted from this example, Shape=rectangular → Modern=Yes (S = Re → 1). This rule has a strength of 0.50.

We can calculate the strength of this rule (S = Re → 1) for all 16 possible responses (response Y15 is the actual response in this case). How likely is it that we get this rule with a strength of 0.50 or more by chance? If it is very unlikely to obtain this strength index by chance, the rule is significant. On the contrary, if it is very likely, the rule is non-significant. Strength acts as the test statistics, and it is compared to an empirical distribution of strengths.

In our example, only 4 responses out of 16 get a strength of 0.50 or higher (last row of Table 6.15). The empirical p-value for rule Shape = Rectangular → 1 is, therefore, 4/16 = 0.25. The same procedure can be used to calculate p-values for the rest of the rules, and they are shown in Table 6.16.

Table 6.16. Extracted rules using original rough sets for data in Table 6.12

		cer	cov	str	p-value
SHAPE: Rectangular	1	1	0.667	0.500	0.250
COLOR: Brown	1	1	0.667	0.500	0.250
SHAPE: Round & COLOR: White	0	1	1.00	0.250	0.500

Consider now the T-shirts example. Table 6.14 can be now updated with a p-value for each rule (Table 6.17).

Table 6.17. Extracted rules using OrigRS for data in Table 6.2 (the T-shirts example)

		cer	*cov*	*str*	*p-value*
COLOR: White	0	1.000	0.857	0.500	0.0156
COLOR: Red / SLEEVES: Short	1	1.000	0.600	0.250	0.1250
COLOR: Red / PRINTING: Picture	1	1.000	0.400	0.167	0.2500
COLOR: Red / PRINTING: Text	1	1.000	0.400	0.167	0.2500
SLEEVES: Long / PRINTING: Plain	0	1.000	0.286	0.167	0.2500

The first rule, Color=white → 0, is significant at the 5% level, as its p-value is 0.0156. This p-value comes from calculating the strength of rule Color=white → 0 considering all possible responses (in this case, that means analyzing 2^{12} = 4096 possibilities). Only some values of strength are possible (in this case, 0, 0.083, 0.167, 0.250, 0.333, 0.417, 0.500), so the histogram in Figure 6.20 only has 7 bars. It is quite difficult to obtain a strength of 0.5 in rule Color=white → 0 just by chance, so we can consider this rule as significant (with a p-value of 64/4096 = 0.0156).

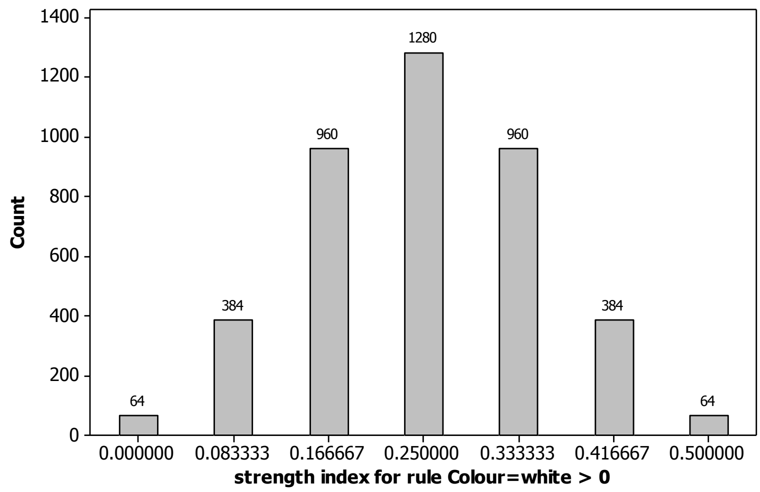


Figure 6.20. Barchart with the empirical distribution of strengths for the rule Color=white → 0 in the T-shirts example.

Although the idea of assigning a test of significance to each rule seems promising, there are several pitfalls with the procedure:

- As the number of stimuli in the design matrix increases, the computational time needed for searching the empirical distribution of strengths goes up rapidly.
- The distribution of strengths is quite "discrete" (Figure 6.20), as strength can only take some values (the more stimuli we have, the more different values of strength we have, and lower p-values can be obtained).
- p-values and strength are inversely related, so for a given set of stimuli, rules with higher strength have lower p-values. So, in fact, strength for each rule gives very similar information to p-values.

So in practice, I suggest the following procedure for deciding which rules to keep in a rough sets analysis:

1. Sorting rules according to its strength, descending.
2. Drawing the values of strength in a graph similar to scree plots from principal component analysis.
3. Consider rules as far as the rule where "the mountain" ends.

6.3.2. Variable precision bayesian rough set (VPBRS)

Applying original rough sets (OrigRS) for extracting rules in KE studies implies using the average of all participants' ratings as response. In the same manner as ordinal logistic regression was introduced in this dissertation for working with each individual rating instead of the average, variable precision bayesian rough sets (VPBRS) can be used as a "rough sets version" that allows working with each individual rating.

After the introduction of rough sets by Prof. Zdzisław Pawlak in 1982 several modifications of the original method were suggested. The modifications are based in the use of the probabilistic structure of rough sets (Pawlak 1999, Ziarko 1993, Ślezak, Ziarko 2005). The newest version, variable precision bayesian rough sets (Slezak, Ziarko 2003), has been proposed for use in Kansei Engineering by Tatsuo Nishino (Nishino, Nagamachi & Tanaka 2005, 2006), and has been applied in some case studies (Nagamachi et al. 2007).

Consider the example shown in Table 6.18, with the same design matrix used for illustrating OrigRS (a 2^2 factorial design with factors Shape and Color). Three people give a rating on each of the four stimuli.

	Shape	Color	P1	P2	P3
Watch 1	Round	White	0	0	0
Watch 2	Rectangular	White	1	1	1
Watch 3	Round	Brown	1	1	0
Watch 4	Rectangular	Brown	1	0	1

Table 6.18. A simple KE study on four watches, evaluating the Kansei word modern.

An alternative presentation of Table 6.18 is shown in Table 6.19. All ratings are now in a single column. Watch 1 is unanimously perceived as not modern, whereas watch 2 is unanimously perceived as modern. On the contrary, there are different opinions when rating watches 3 and 4.

Just looking at rectangular watches, 5 out of 6 responses (from 3 different persons) were "modern (1)". Can we then extract the rule Shape=Rectangular → Modern? Using VPBRS gives the answer.

The notation used for describing VPBRS is the one suggested in Nishino, Nagamachi & Tanaka (2005). m products (stimuli) are evaluated. In the example from Table 6.19, m=4 and the set of products is $E = \{E_1, E_2, E_3, E_4\}$. n people rate the stimuli (n=3 in our example). So we have $n \cdot m$ ratings, in the example, 3·4 = 12 ratings, the universe of ratings, $U = \{x_{11}, \dots, x_{ji}, \dots, x_{nm}\}$. There are r attributes, $A = \{a_1, \dots, a_k, \dots, a_r\}$, in our example, $A = \{a_1, a_2\} = \{Shape, Color\}$; and $D = \{D_1, \dots, D_s, \dots, D_p\}$ for p decision classes, where $D_s = \{x \mid d(x) = s\}$. In the example, only two classes are used: 0 and 1.

Product (E)	Event (U)	Shape (a_1)	Color (a_2)	Modern (d)
Watch 1 E_1	x_{11}	Round	White	0
	x_{21}			0
	x_{31}			0
Watch 2 E_2	x_{12}	Rectangular	White	1
	x_{22}			1
	x_{32}			1
Watch 3 E_3	x_{13}	Round	Brown	1
	x_{23}			1
	x_{33}			0
Watch 4 E_4	x_{14}	Rectangular	Brown	1
	x_{24}			0
	x_{34}			1

Table 6.19. A simple KE study on four watches, evaluating the Kansei word moder⚜.

Two different probabilities can be calculated from data such as that in Table 6.19:

The prior probability of decision	$P(D_s) = \dfrac{\lvert D_s \rvert}{\lvert U \rvert}$
The conditional probability (dependent on the conditional attributes)	$P(D_s \mid E_i) = \dfrac{\lvert D_s \cap E_i \rvert}{\lvert E_i \rvert}$

The calculated probabilities for data in Table 6.19 are shown in Table 6.20.

$P(D_0) = {}^{5}/_{12} = 0.4167$	$P(D_1) = {}^{7}/_{12} = 0.5833$
$P(D_0 \vert E_1) = {}^{3}/_{3} = 1$	$P(D_1 \vert E_1) = {}^{0}/_{3} = 0$
$P(D_0 \vert E_2) = {}^{0}/_{3} = 0$	$P(D_1 \vert E_2) = {}^{3}/_{3} = 1$
$P(D_0 \vert E_3) = {}^{1}/_{3} = 0.3333$	$P(D_0 \vert E_3) = {}^{2}/_{3} = 0.6667$
$P(D_0 \vert E_4) = {}^{1}/_{3} = 0.3333$	$P(D_0 \vert E_4) = {}^{2}/_{3} = 0.6667$

Table 6.20. The prior and conditional probabilities in the watches example.

A fast review of rough sets versions until VPBRS, and its use in KE

A summary of rough sets versions before the last step, VPBRS, can be found in Slezak and Ziarko (2003).

In original rough sets, the approximation regions are defined by:

$$POS(D_s) = \bigcup \{E_i \mid P(D_s|E_i) = 1\}$$
$$NEG(D_s) = \bigcup \{E_i \mid P(D_s|E_i) = 0\}$$
$$BND(D_s) = \bigcup \{E_i \mid P(D_s|E_i) \in (0,1)\}$$

These regions represent the stimuli which have, respectively, a rating E_i certain, unlikely, and possible but not certain.

In variable precision rough sets (VPRS), the original RS model is slightly modified to increase its discriminatory capabilities. A lower and upper certainty threshold are used to define the *u*-positive, *l*-negative and *(l,u)*-boundary regions:

$$POS_u(D_s) = \bigcup \{E_i \mid P(D_s|E_i) \geq u\}$$
$$NEG_l(D_s) = \bigcup \{E_i \mid P(D_s|E_i) \leq l\}$$
$$BND_{l,u}(D_s) = \bigcup \{E_i \mid P(D_s|E_i) \in (l,u)\}$$

OrigRS is a particular case of VPRS with l=0 and u=1.

In bayesian rough set model (BRS), the certainty threshold is not predefined, but taken from the available information – through the use of the prior probability $P(D_s)$. Positive and negative regions are defined respectively as areas of certainty improvement and loss with respect to predicting the rating D_j.

$$POS^*(D_s) = \bigcup \{E_i \mid P(D_s|E_i) > P(D_s)\}$$
$$NEG^*(D_s) = \bigcup \{E_i \mid P(D_s|E_i) < P(D_s)\}$$
$$BND^*(D_s) = \bigcup \{E_i \mid P(D_s|E_i) = P(D_s)\}$$

Variable precision bayesian rough set models join both ideas used in VPRS and BRS. They are similar to variable precision rough set in the sense that they use a certainty threshold to define the approximate regions, but this threshold is not a predetermined number, but also incorporates the available information using the prior probability $P(D_s)$.

The VPBRS adaptation to Kansei Engineering suggested in Nishino, Nagamachi & Tanaka (2005, 2006) uses a parameter β to define the approximate regions:

$$POS^\beta(D_s) = \bigcup \left\{E_i \mid P(D_s|E_i) \geq \frac{P(D_s)}{1-\beta}\right\}$$

$$NEG^\beta(D_s) = \bigcup \left\{E_i \mid P(D_s|E_i) \leq \frac{P(D_s)-\beta}{1-\beta}\right\}$$

$$BND^\beta(D_s) = \bigcup \left\{E_i \mid P(D_s|E_i) \in \left(\frac{P(D_s)-\beta}{1-\beta}, \frac{P(D_s)}{1-\beta}\right)\right\}$$

This is the method that I will use for VPBRS in KE studies. β is commonly 0.2, but its value can be changed depending on the dataset.

Using the data in Table 6.19 and β=0.2, the positive, negative and boundary regions for defining D_0 (responses with 0, not modern) and D_1 (responses with 1, modern) are the following:

$$D_0 \begin{cases} POS^{0.2}(D_0) = \bigcup \left\{ E_i \mid P(D_0|E_i) \geq \dfrac{P(D_0)}{1-0.2} = 0.5209 \right\} = E_1 \\ NEG^{0.2}(D_0) = \bigcup \left\{ E_i \mid P(D_0|E_i) \leq \dfrac{P(D_0)-0.2}{1-0.2} = 0.2709 \right\} = E_2 \\ BND^{0.2}(D_0) = \bigcup \{ E_i \mid P(D_0|E_i) \in (0.2709, 0.5209) \} = E_3 \cup E_4 \end{cases}$$

$$D_1 \begin{cases} POS^{0.2}(D_1) = \bigcup \left\{ E_i \mid P(D_1|E_i) \geq \dfrac{P(D_1)}{1-0.2} = 0.7291 \right\} = E_2 \\ NEG^{0.2}(D_1) = \bigcup \left\{ E_i \mid P(D_1|E_i) \leq \dfrac{P(D_1)-0.2}{1-0.2} = 0.4791 \right\} = E_1 \\ BND^{0.2}(D_1) = \bigcup \{ E_i \mid P(D_1|E_i) \in (0.4791, 0.7291) \} = E_3 \cup E_4 \end{cases}$$

It always happens that:

$$U = POS^{\beta}(D_s) \cup NEG^{\beta}(D_s) \cup BND^{\beta}(D_s) \tag{6.8}$$

Extraction method of decision rules from approximate regions

Approximate regions are exclusive one from the other, as can be seen from Eq. (6.8). Therefore, a consistent decision matrix can be constructed as the one in Table 6.21.

$$\begin{array}{cc|c} & & NEG^{\beta}(D_j) \\ & & E_{N1} \ \dots \ E_j \ \dots \\ \hline POS^{\beta}(D_j) & \begin{matrix} E_{P1} \\ \vdots \\ E_i \\ \vdots \end{matrix} & M_{ij}^{\beta}(D_j) \end{array}$$

Table 6.21. The decision matrix for the approximate regions.

Elements of the decision matrix are defined:

$$M_{ij}^{\beta}(D_j) = \left\{ \bigvee a_k = v_{ik} \mid a_k(E_i) \neq a_k(E_j), \forall a_k \in A \right\} \tag{6.9}$$

where $\bigvee a_k = v_{ik}$ is a disjunction of attribute elements to discern E_i and E_j (that is, attributes in E_i that are different from those in E_j).

From $POS^{\beta}(D_j)$ we can derive decision rules using the following decision function:

$$POS^{\beta}(D_j) = \bigvee_{E_i \in POS^{\beta}(D_j)} \bigwedge_{E_j \notin POS^{\beta}(D_j)} M_{ij}^{\beta}(D_j)$$

Everything will hopefully be clarified with the watches example. Table 6.22 reproduces the design matrix of the example.

	Shape	Color
E_1	Round	White
E_2	Rectangular	White
E_3	Round	Brown
E_4	Rectangular	Brown

Table 6.22. The design matrix in the watches example for illustrating the rule extraction process in VPBRS

Decision matrices for D_0 and D_1 are shown in Table 6.23. Notice how attributes are selected according to Eq. (6.9)

Table 6.23. Decision matrices for D_0 and D_1 for the watches example

		$NEG^{0.2}(D_0)$
		E_2
$POS^{0.2}(D_0)$	E_1	Shape=Round

		$NEG^{0.2}(D_1)$
		E_1
$POS^{0.2}(D_1)$	E_2	Shape=Rectangular

The extracted rules – derived from $POS^{0.2}(D_s)$ – are:
Shape=Round → 0 (not modern)
Shape=Rectangular → 1 (modern)

Calculation of indexes in VPBRS

As we did in OrigRS, three indexes can be calculated for each extracted rule. The indexes are the same as in OrigRS, certainty, coverage and strength ($|\cdot|$ denotes cardinality)

Certainty $$cer(cond_k; D_s) = \frac{|cond_k \cap D_s|}{|cond_k|} = \frac{\sum_{E_i \in cond_k} |E_i| P(D_s|E_i)}{\sum_{E_i \in cond_k} |E_i|}$$

Coverage $$cov(cond_k; D_s) = \frac{|cond_k \cap D_s|}{|D_s|} = \frac{\sum_{E_i \in cond_k} |E_i| P(D_s|E_i)}{|D_s|}$$

Strength $$str(cond_k; D_s) = \frac{|cond_s \cap D_j|}{|U|} = \frac{\sum_{E_i \in cond_k} |E_i| P(D_s|E_i)}{|U|}$$

Certainty is the ratio of the number of events that satisfy the rule $cond_k \rightarrow D_s$ to the number of events that satisfy the condition part of the rule $cond_k$. It measures the degree to which the rule $cond_k \rightarrow D_s$ holds.

Coverage is the ratio of the number of events that satisfy the rule $cond_k \rightarrow D_s$ to the number of events that have response category D_s. It measures the degree to which the inverse rule $D_s \rightarrow cond_k$ holds.

Strength is the ratio of the number of events that satisfy the rule $cond_k \rightarrow D_s$ to the total number of events, so it measures, of course, the strength of rule $cond_k \rightarrow D_s$.

In contradistinction to what happened in OrigRS, where only deterministic rules were shown, not all extracted rules in VPBRS have certainty equal to 1. Therefore, not only strength, but also certainty, can be used in VPBRS to select rules.

The two extracted rules for the watches example, now with its indexes, are shown in Table 6.24.

		cer	*cov*	*str*
SHAPE: Rectangular	1	0.8333	0.7143	0.4167
SHAPE: Round	0	0.6667	0.8000	0.3333

Table 6.24. Extracted rules using VPBRS for data in Table 6.18

Table 6.25 shows the extracted rules for the T-shirts example when using VPBRS with parameter β=0.2. The list of rules is long. They are sorted according to its strength, descending. The graph in Figure 6.21 helps in the decision of how many rules to keep: only rules "in the mountain" will be used. Color codification in Table 6.25 shows discarded rules in very pale blue, implicitly implying that there is no need to consider those rules.

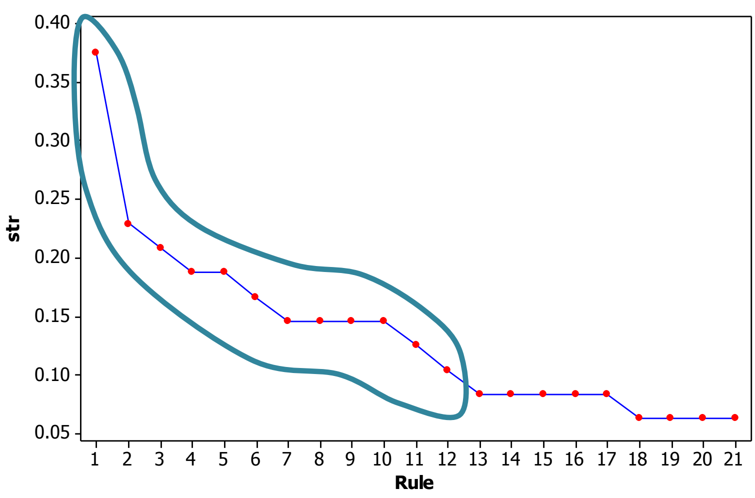


Figure 6.21. Scree plot of strength of rules from the VPBRS analysis with the T-shirts data

We know that Sleeves is an inert factor. However, many rules have Sleeves on it. This is, in fact, a consequence of Sleeves being inert: as Sleeves has no effect on the response, it usually appears with other factors that do have an effect. For example, the rules SLEEVES: Short & COLOR: Red → 1, COLOR: White & SLEEVES: Long → 0 and SLEEVES: Short & COLOR: White → 0, are in fact reinforcing the rule COLOR: White → 0 and implying the rule COLOR: Red → 1.

Something similar happens with Printing, the other active factor. Clearly, PRINTING: Picture → 1. Rule COLOR: White & PRINTING: Plain → 0 goes in the good direction of pointing that plain printing gives a low rating on the response. Interestingly, PRINTING: Text & COLOR: Red → 1 but PRINTING: Text & COLOR: White → 0. We already know that T-shirts with text have a middle rating on the response (more than plain T-shirts, but less than T-shirts with a picture). This is why, when combined with red color, response is 1 (as red color increases the response), but when combined with white color, the response is 0 (as white color decreases the response).

In my opinion, the complex output from rough sets (either OrigRS or VPBRS) makes it unsuitable to recommend its general use in the synthesis phase. However, it can complement well the results from regression models.

Table 6.25. Extracted rules using VPBRS for data in Table 6.2 (the T-shirts example)

		cer	*cov*	*str*
COLOR: White	0	0.7500	0.7826	0.3750
PRINTING: Picture	1	0.6875	0.4400	0.2292
SLEEVES: Short & COLOR: Red	1	0.8333	0.4000	0.2083
COLOR: White & SLEEVES: Long	0	0.7500	0.3913	0.1875
SLEEVES: Short & COLOR: White	0	0.7500	0.3913	0.1875
PRINTING: Text & COLOR: Red	1	1.0000	0.3200	0.1667
PRINTING: Plain & SLEEVES: Long	0	0.8750	0.3043	0.1458
COLOR: White & PRINTING: Plain	0	0.8750	0.3043	0.1458
COLOR: White & PRINTING: Text	0	0.8750	0.3043	0.1458
PRINTING: Picture & COLOR: Red	1	0.8750	0.2800	0.1458
PRINTING: Picture & SLEEVES: Long	1	0.7500	0.2400	0.1250
PRINTING: Picture & SLEEVES: Short	1	0.6250	0.2000	0.1042
PRINTING: Plain & COLOR: White & SLEEVES: Long	0	1.0000	0.1739	0.0833
SLEEVES: Short & PRINTING: Text & COLOR: White	0	1.0000	0.1739	0.0833
PRINTING: Picture & COLOR: Red & SLEEVES: Long	1	1.0000	0.1600	0.0833
PRINTING: Text & SLEEVES: Long & COLOR: Red	1	1.0000	0.1600	0.0833
PRINTING: Text & SLEEVES: Short & COLOR: Red	1	1.0000	0.1600	0.0833
SLEEVES: Short & PRINTING: Plain & COLOR: White	0	0.7500	0.1304	0.0625
PRINTING: Text & SLEEVES: Long & COLOR: White	0	0.7500	0.1304	0.0625
SLEEVES: Short & PRINTING: Picture & COLOR: Red	1	0.7500	0.1200	0.0625
SLEEVES: Short & COLOR: Red & PRINTING: Plain	1	0.7500	0.1200	0.0625

7 The Tools for the Synthesis Phase in Action

This chapter uses simulation to compare among the tools for the synthesis phase studied in Chapter 6. A procedure is introduced to help decide the design matrix for a KE study. Interactions are also introduced, together with a visual representation that facilitates interpretation. Finally, data from the fruit juice experiment is reanalyzed.

7.1. Comparing Methods for the Synthesis Phase Using Simulations

Ma and Nakamori (2007) reproduce on a paper about Kansei Engineering some stimulating words about simulation first written by Axelrod (1997):

> *Simulation is a third way of doing science. Like deduction, it starts with a set of explicit assumptions. But unlike deduction, it does not prove theorems. Instead a simulation generates data that can be analyzed inductively. Unlike typical induction, however, the simulated data comes from a rigorously specified set of rules rather than directly measurement of the real world. While induction can be used to find patterns in data, and deduction can be used to find consequences of assumptions, simulation modeling can be used to aid intuition*

Last chapter introduced a set of useful tools for analyzing data from Kansei Engineering studies in the synthesis phase. The aim of this chapter is comparing the behavior of these tools and ending up with a proposal of the best way to proceed. Simulated data will be used to perform the comparisons.

After analyzing real data (as that from the fruit juice experiment), one does not know if the conclusions are right or wrong (we just believe they are right). Using simulated data we know the truth, so we can judge if a particular tool works well or not. In fact, data from the T-shirts experiment in Chapter 6 was simulated data (Table 6.2). That dataset was produced going through the following steps:

1. A model was written that included the significant factors (Color and Printing, in that case) in the desired way (so that the effects were those that can be seen in Figure 6.3).
2. Noise was added to simulate intra and inter variability, that is, variability among people ratings (inter) and among the ratings given by each person (intra).
3. Data was discretized so that each rating was a number on a 7-point scale.

The purpose of that dataset was simply illustrating all the introduced techniques for the synthesis phase. I thought it was pedagogical using a simulated dataset that produced the same results with all techniques. This was achieved introducing a low amount of variability, thus having very clear significant effects. This first set of simulated data will be called from now on simulation set 1.

In this section, two experiences will be presented:

- Experience 1: What happens when we are faced with a dataset with more variability – so close to what happens in real life (simulation set 2).
- Experience 2: What happens when we start removing stimuli from a full factorial design (simulation set 3, 4 and 5).

7.1.1. Experience 1 (simulation set 2)

Simulation set 2 has been created exactly in the same way as simulation set 1, but adding more noise. The significant factors are the same: Color and Printing (and we know it for sure because our model was written to satisfy this). Table 7.1 shows the results from simulation set 2.

Table 7.1. Results for the T-shirts experiment (simulation set 2)

	Color	Sleeves	Printing	Carla	Joan	Marc	Maria	MEAN	
1	Red	Long	Picture	5	3	7	7	5.50	1
2	White	Long	Picture	4	5	2	3	3.50	2
3	Red	Short	Picture	7	5	6	2	5.00	3
4	White	Short	Picture	6	4	6	6	5.50	4
5	Red	Long	Plain	4	3	3	1	2.75	5
6	White	Long	Plain	4	1	1	4	2.50	6
7	Red	Short	Plain	6	4	7	7	6.00	7
8	White	Short	Plain	4	1	2	3	2.50	8
9	Red	Long	Text	7	5	5	4	5.25	9
10	White	Long	Text	3	7	4	6	5.00	10
11	Red	Short	Text	6	3	5	5	4.75	11
12	White	Short	Text	4	4	2	5	3.75	12

Table 7.2 shows the p-values for each main effect using QT1, ordinal logistic regression (OLR) and mixed effects ordinal logistic regression (mOLR). Significant p-values are shadowed in light red.

Table 7.2. p-values for simulation set 2 using QT1, OLR and mOLR

	Color	Sleeves	Printing
QT1	0.1326	0.4580	0.2044
OLR	**0.01868**	0.24454	**0.02606**
mOLR	**0.01568**	0.22687	**0.02251**

QT1 fails in the endeavor of detecting the significant effects: all factors wrongly appear as inert factors. Instead, both OLR and mOLR correctly identify Color and Printing as significant. Ordinal logistic regression is able to extract the variability among participants in the study, and thus is more powerful for detecting significance. In this small example (only 4 participants) there are no big differences between OLR and mOLR, although p-values from mOLR are slightly smaller.

Table 7.3 shows the rules involving two factors extracted using original rough sets (OrigRS, left) and variable precision bayesian rough sets (VPBRS, right). Nothing is added to our conclusions coming from logistic regression – or even worse, confusion is added. It is never very clear, when using rough sets, which factors are significant and which are inert. In my opinion, this fact makes directly extracting conclusions from rough sets analysis very hard, if not impossible.

However, rough sets can be used as a check after performing a logistic regression. Looking at the left part of Table 7.2 we can ascertain that red color always gives 1 (a high response rating) as a result, whereas white color gives 0 (a low response rating). Picture and text in Printing give 1 and plain gives 0. These results agree with those from the logistic regression. Sleeves, an inert factor, also appears in the rules, but with level long both giving 0 and 1 (an implicit confirmation of its inertness). Rules from VPBRS are quite similar than those from OrigRS.

Table 7.3. Rules from original rough sets (left) and variable precision bayesian rough sets (right) for simulation set 2 data

COLOR: Red	SLEEVES: Short	→	1		PRINTING: Text	→	1
COLOR: Red	PRINTING: Picture	→	1	COLOR: Red	SLEEVES: Short	→	1
COLOR: Red	PRINTING: Text	→	1	COLOR: Red	PRINTING: Text	→	1
SLEEVES: Short	PRINTING: Picture	→	1	SLEEVES: Short	PRINTING: Picture	→	1
SLEEVES: Long	PRINTING: Tex	→	1	SLEEVES: Long	PRINTING: Text	→	1
SLEEVES:◎Long	PRINTING: Plain	→	0	SLEEVES: Short	PRINTING: Text	→	1
COLOR: White	PRINTING: Plain	→	0	COLOR: White	SLEEVES: Long	→	0
				COLOR: White	PRINTING: Plain	→	0
				SLEEVES: Long	PRINTING: Plain	→	0
				SLEEVES: Long	PRINTING: Picture	→	0

7.1.2. Experience 2 (simulation sets 3, 4 and 5)

Recall simulation set 1, the T-shirts dataset originally used in Chapter 6 (Table 6.2). The design matrix of that set has 12 runs, all combinations of levels for our 3 factors (a full factorial design). What happens when we start erasing runs from simulation set 1? Answering this question is the purpose of this subsection. Three simulation sets will be used:

- Simulation set 3 has exactly the same results than simulation set 1, but 4 runs have been erased (runs 6, 7, 9 and 12). The 8 remaining runs constitute the minimum permitted runs for keeping the matrix orthogonal and allowing the estimation of main effects free of confusion among them.
- Simulation set 4 is build erasing run 2 from simulation set 3; it only has 7 runs. Simulation set 5 is build erasing run 1 from simulation set 4; it only has 6 runs. Orthogonality is lost in simulations sets 4 and 5. Our interest is checking if these non-orthogonal design matrices are able to correctly estimate all main effects.

This process of erasing runs for obtaining simulation sets 3, 4 and 5 is summarized in Table 7.4.

I have performed QT1, OLR and mOLR on the three simulation sets. Table 7.5 shows the p-values for each main effect: Color, Sleeves and Printing. Significant p-values (considering a level of significance of 5%) are printed bold.

Table 7.4. Results from the T-shirts experiment (simulation sets 3, 4 and 5)

Runs used in simulation set... 3	4	5		Color	Sleeves	Printing	Carla	Joan	Marc	Maria	MEAN
•	•		1	Red	Long	Picture	5	6	7	5	5.75
•			2	White	Long	Picture	3	4	5	3	3.75
•	•	•	3	Red	Short	Picture	7	4	5	6	5.50
•	•	•	4	White	Short	Picture	4	4	4	3	3.75
•	•	•	5	Red	Long	Plain	1	4	5	1	2.75
			6	White	Long	Plain	1	3	2	1	1.75
			7	Red	Short	Plain	4	5	5	2	4.00
•	•	•	8	White	Short	Plain	2	4	5	1	3.00
			9	Red	Long	Text	5	6	6	4	5.25
•	•	•	10	White	Long	Text	1	2	5	1	2.25
•	•	•	11	Red	Short	Text	5	6	7	4	5.50
			12	White	Short	Text	2	4	3	2	2.75
8	7	6	← Total number of runs								

Table 7.5. p-values for simulation set 3 using QT1, OLR and mOLR

	Simulation set 3			Simulation set 4			Simulation set 5		
	Color	Sleeves	Priting	Color	Sleeves	Priting	Color	Sleeves	Priting
QT1	0.0557	0.2390	0.1587	0.1240	0.3096	0.3443	**0.0473**	0.0524	0.1339
OLR	**0.0006**	0.1757	**0.0457**	**0.0019**	0.1789	0.1334	**0.0100**	**0.0256**	0.4542
mOLR	**0.0000**	0.0633	**0.0057**	**0.0004**	0.0909	**0.0418**	**0.0036**	**0.0069**	0.2682

The behavior of each tool with each simulation set is the following:

- In simulation set 3, QT1 fails to detect any significant factor. Both OLR and mOLR correctly detect Color and Printing as the significant factors. mOLR produces lower p-values than OLR.
- In simulation set 4, QT1 is again unable to detect anything significant. Results for this simulation set differ between OLR and mOLR. OLR only finds Color as significant, whereas mOLR correctly shows low p-values for both significant factors – Color and Printing.
- In simulation set 5, none of the tools are able to correctly find the significant factors. OLR and mOLR detect Sleeves as significant, but we know – thanks to the model used to generate the data – that Sleeves is an inert factor.

The results from using rough sets with simulation sets 3 and 4 are the same as in experience 1: they can be used as a check of the obtained results from the regression model. In simulation set 5, original rough sets gives the rules shown in Table 7.6. The influence of the significant factors is correctly captured: Color white gives 0 and color red gives 1, whereas Printing plain gives 0 and Printing text and picture give 1. However, and at the same level of importance, Sleeves long give 0 and Sleeves short give 1. And as always with rough sets, there is no clear way to discover which effects are really significant. It seems there is no tool that works well in simulation set 5.

Table 7.6. Extracted rules using OrigRS with simulation set 5

	COLOR: White	→	0
	SLEEVES: Long	→	0
	PRINTING: Plain	→	0
COLOR: Red	SLEEVES: Short	→	1
COLOR: Red	PRINTING: Picture	→	1
COLOR: Red	PRINTING: Text	→	1
SLEEVES: Short	PRINTING: Text	→	1

A conclusion from the results in experience 1 and 2 is that mOLR works better than OLR, and OLR works better than QT1. The difference in behavior between QT1 and both OLR and mOLR is bigger than the difference between both versions of the logistic regression. So my recommendation is avoiding the use of QT1, and using instead a logistic regression. If possible, a mixed effects ordinal logistic regression must be preferred over a fixed effects ordinal logistic regression.

Another conclusion is that, in some occasions, none of the studied tools give the right result (simulation set 5). Why does this happen? Is it possible to use a new tool that works well for simulation set 5? These questions are the topic of next section.

7.2. Assessing the suitability of design matrices

In Kansei Engineering studies, the definition of the design matrix (that is, the combinations of levels for each factor that constitutes each run) is done when spanning the space of properties. The statistical properties of this design matrix are very important for the posterior analysis, once the data collection phase is finished. However, this aspect is often neglected. One reason might be the fact that, as we have seen in Section 5.4.2, around 80% of the reviewed papers use a selection of already existing products as a set of stimuli instead of building prototypes. In this situation, it could be difficult – if not impossible – to find all the needed stimuli.

When the statistical issues regarding the design matrix are faced, the recommendation is using factorial designs (Barone, Lombardo & Tarantino 2007, Bahn et al. 2009) or Taguchi orthogonal arrays (Ogawa, Nagano & Yukawa 2009, Chen, Chuang 2008, Lai, Chang & Chang 2005). Even for non statisticians, it is quite obvious that the fact of having a matrix as balanced as possible seems a good option: probably things will go better if all levels of factors are present in a similar number.

When is our design matrix good enough to be able to correctly perform the synthesis phase? The purpose of this section is answering this question, and quantifying "how good" our design matrix is. The design matrix in simulation set 5 was not good enough: this is why all tools were unable to detect the right significant factors, no matter what we did in the synthesis phase. Notice that we were aware of the problem in simulation set 5 because we knew what the correct answer was (as we were using simulated data). With real data, we could be happy wrongly thinking that Color and Sleeves were the significant factors.

It could perfectly be the case that some Kansei Engineering studies reached wrong conclusions because of a bad-defined design matrix. However, nobody ever noticed, because the right answer was unknown. Furthermore, it is easy to find explanations that reinforce the reached conclusion, whatever it is.

A desired condition for the design matrix is orthogonality. Orthogonality plays an important role in the theory of design of experiments. When having an orthogonal design matrix, each factor effect can be calculated independently of other factors. The standard error of the estimates is the same for all of them, and the interpretation of the effects can be also made independently of other factors. In addition, some techniques that require iterative calculations become more unstable when working with matrices that depart a lot from orthogonality.

The best situation is, undoubtedly, having full and fractional factorial designs. These designs are not only orthogonal, but also regular (with no partial aliasing). The popular use of some handy graphical tools for choosing significant effects, such as representing the effects on a normal probability plot – or on a half normal probability plot – (Daniel 1959) is possible thanks to having orthogonal and regular matrices. However, there are designs that, being orthogonal, are also non-regular (having partial aliasing). That is the case of Plackett and Burman designs, for instance. At least, Plackett and Burman designs do not confuse main effects among them (something done in supersaturated designs, which are not only non-regular but also non-orthogonal).

So, when deciding on the design matrix for a KE study in the spanning the space of properties phase, both orthogonality and level of confusion must be taken into account. I will define two suitability indexes for quantifying the quality of a design matrix, both defined as percentages:

- Orthogonality index: an index that controls if a matrix is orthogonal (100%), or the amount of deviation from orthogonality (something between 0% and 100%).
- Confusion index: an index that checks if there is no confusion among main effects (100%), and that goes closer to 0% when the correlation between main effects increases.

7.2.1. Orthogonality index (OI)

Although orthogonality is a desired property, it cannot be always achieved. How can we measure to which degree a matrix departs from orthogonality? This question is particularly relevant in the field of supersaturated designs[100], and several indicators have been developed under the umbrella of these designs. For instance, Wang and Wu (1992) proposed nearly orthogonal arrays using a criterion of efficiency based on the idea of minimizing the number of non-orthogonal pairs. This idea was later expanded by Ma, Fang & Liski (2000) and Jang (2002), who proposed several measures for non-orthogonality. I will derive an orthogonality index (OI) for KE studies from the $D(A)$ measure defined in Jang (2002).

An orthogonal design is a matrix that fulfils the following two conditions:

- Condition 1: Each level in each column appears the same number of times.
- Condition 2: In any two columns all the level combinations appear the same number of times.

Let $M = (\boldsymbol{c}_1, \boldsymbol{c}_1, \dots, \boldsymbol{c}_r)$ be an $N \times r$ matrix. Each column $\boldsymbol{c}_i$ corresponds to a factor. Each factor i has q_i levels. Jang (2002) suggests the following expressions – slightly modified here – to measure the non-orthogonality of a matrix:

---

[100] Supersaturated designs are not orthogonal, but efforts are made to keep them "as orthogonal as possible". They confound main effects among them and have complex partial aliasing structures. There are different ways to analyze its results, none of them really satisfactory. In my opinion, supersaturated designs is a murky field of research. On the one hand, they are attractive, because it seems they allow the estimation of main effects with less runs than the number of effects to estimate. But on the other hand, the analysis methods do not work very well, probably because they try to achieve something impossible. The fact that different methods of analysis exist, none of them being "the good one", points to my impression that supersaturated designs is "asking the impossible".

$$f_1(\boldsymbol{c}_i) = \frac{1}{q_i}\sum_{k=1}^{q_i}\left|N_{\boldsymbol{c}_i}(k) - \frac{N}{q_i}\right| \tag{7.1}$$

$$D_1(M) = \frac{1}{r}\sum_{1\le i\le r} f(\boldsymbol{c}_i) \tag{7.2}$$

$$f_2(\boldsymbol{c}_i,\boldsymbol{c}_j) = \frac{1}{q_iq_j}\sum_{k=1}^{q_i}\sum_{l=1}^{q_j}\left|N_{\boldsymbol{c}_i,\boldsymbol{c}_j}(k,l) - \frac{N}{q_iq_j}\right| \tag{7.3}$$

$$D_2(M) = \frac{1}{N_r}\sum_{1\le i\le j\le r} f(\boldsymbol{c}_i,\boldsymbol{c}_j) \tag{7.4}$$

$$D(M) = \frac{1}{N}\left(D_1(M) + D_2(M)\right) \tag{7.5}$$

where N is the number of rows in the matrix, $\boldsymbol{c}_i, \boldsymbol{c}_j$ are two columns of M, $N_{\boldsymbol{c}_i}(k)$ is the number of k-th level in a column $\boldsymbol{c}_i$, N/q_i is the average number of levels in column $\boldsymbol{c}_i$, $N_{\boldsymbol{c}_i,\boldsymbol{c}_j}(k,l)$ is the number of (k,l) pairs in $(\boldsymbol{c}_i, \boldsymbol{c}_j)$ and $N/q_i\,q_j$ is the average number of level combinations in each pair of two columns $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$.

$D(M)$ is a non-orthogonality value, comprised of two components, $D_1(M)$ and $D_2(M)$. $D_1(M)$ evaluates condition 1 for being an orthogonal design: if this value is 0, condition 1 is satisfied. $D_2(M)$ evaluates condition 2 for being an orthogonal design: if this value is 0, condition 2 is satisfied. If $D(M)$ is 0, the matrix is an orthogonal design. The larger the value $D(M)$ is, the further matrix M departs from orthogonality.

Calculation of $D(M)$ will be exemplified with matrices M_1(a $2^{3\text{-}1}$ fractional factorial design) and M_2 (four randomly selected runs from a $2^1 \cdot 3^1$ factorial design), shown in Table 7.7.

Table 7.7. Two simple matrices to illustrate the calculation of value $D(M)$: M_1 (left) and M_2 (middle). M_2^w (right) is the "worst" possibility for matrix M_2

Matrix M_1			Matrix M_2		"Worst" matrix M_2 = M_2^w	
A	B	C	A	B	A	B
A1	B1	C1	A1	B1	A1	B1
A2	B1	C2	A2	B1	A1	B1
A1	B2	C2	A3	B2	A1	B1
A2	B2	C1	A1	B2	A1	B1

All full and fractional factorial designs are orthogonal matrices, so matrix M_1 must have a value $D(M_1)$ of 0:

$$f_1(\boldsymbol{c}_1) = \frac{1}{2}\left(\left|2-\frac{4}{2}\right| + \left|2-\frac{4}{2}\right|\right) = 0$$
$$f_1(\boldsymbol{c}_2) = 0$$
$$f_1(\boldsymbol{c}_3) = 0$$

$$f_2(\boldsymbol{c}_1,\boldsymbol{c}_2) = \frac{1}{2\cdot 2}\left(\left|1-\frac{4}{2\cdot 2}\right| + \left|1-\frac{4}{2\cdot 2}\right| + \left|1-\frac{4}{2\cdot 2}\right| + \left|1-\frac{4}{2\cdot 2}\right|\right) = 0$$
$$f_2(\boldsymbol{c}_1,\boldsymbol{c}_3) = 0$$
$$f_2(\boldsymbol{c}_2,\boldsymbol{c}_3) = 0$$

$$D_1(M_1) = \frac{1}{3}(0+0+0) = 0 \qquad D_2(M_1) = \frac{1}{3}(0+0+0) = 0$$
$$D(M_1) = \frac{1}{N}\left(D_1(M_1) + D_2(M_1)\right) = 0$$

Matrix M_2, on the contrary, does not fulfill neither condition 1 (level A1 appears twice in the column, whereas levels A2 and A3 appear only once) nor condition 2 (some combinations of levels of factors A and B appear once, while other do not appear at all) to become an orthogonal design. For this reason, $D(M_2)$ is different from 0.

$$f_1(\boldsymbol{c}_1) = \frac{1}{3}\left(\left|2-\frac{4}{3}\right| + \left|1-\frac{4}{3}\right| + \left|1-\frac{4}{3}\right|\right) = \frac{4}{9}$$
$$f_1(\boldsymbol{c}_2) = \frac{1}{2}\left(\left|2-\frac{4}{2}\right| + \left|2-\frac{4}{2}\right|\right) = 0$$

$$f_2(\boldsymbol{c}_1,\boldsymbol{c}_2) = \frac{1}{3\cdot 2}\left(\left|1-\frac{4}{3\cdot 2}\right| + \left|1-\frac{4}{3\cdot 2}\right| + \left|1-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right| + \left|1-\frac{4}{3\cdot 2}\right|\right) = \frac{8}{3}$$

$$D_1(M_2) = \frac{1}{2}\left(\frac{4}{9}+0\right) = \frac{2}{9} \qquad D_2(M_2) = \frac{1}{6}\left(\frac{8}{3}\right) = \frac{4}{9}$$
$$D(M_2) = \frac{1}{N}\left(D_1(M_2) + D_2(M_2)\right) = \frac{1}{4}\left(\frac{2}{9}+\frac{4}{9}\right) = \frac{1}{6}$$

Our purpose is not only detecting if a matrix is an orthogonal design – $D(M) = 0$ – or not – $D(M) > 0$ – but also quantifying the departure from orthogonality using an orthogonality index (OI). To achieve this, I suggest the following procedure:

1. Construct a matrix that has the same number of rows and columns than the matrix in study. Each column (each factor) also has the same number of levels than the original matrix, but the whole column is filled with only one of the levels. This matrix is the least orthogonal matrix that can be created with factors and levels from the original matrix, and is called M^w
2. Calculate $D(M^w)$ using Equations (7.1) to (7.5). This value $D(M^w)$ is the biggest value possible for factors and levels from the original matrix.
3. In order to have an orthogonality index in percentatge, with 100% meaning an orthogonal design, and 0% meaning the least orthogonal matrix you can have, the following expression can be used:

$$OI = \left(1 - \frac{D(M)}{D(M^w)}\right) \cdot 100$$

Table 7.7 (on the right) shows the "worst" version of matrix M_2, matrix M_2^w. Calculation of $D(M_2^w)$ follows:

$$f_1(\boldsymbol{c}_1) = \frac{1}{3}\left(\left|4-\frac{4}{3}\right| + \left|0-\frac{4}{3}\right| + \left|0-\frac{4}{3}\right|\right) = \frac{16}{9}$$

$$f_1(\boldsymbol{c}_2) = \frac{1}{2}\left(\left|4-\frac{4}{2}\right| + \left|0-\frac{4}{2}\right|\right) = 2$$

$$f_2(\boldsymbol{c}_1, \boldsymbol{c}_2) = \frac{1}{3\cdot 2}\left(\left|4-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right| + \left|0-\frac{4}{3\cdot 2}\right|\right) = \frac{20}{3}$$

$$D_1(O_2^w) = \frac{1}{2}\left(\frac{16}{9}+2\right) = \frac{17}{9} \qquad D_2(O_2^w) = \frac{1}{6}\left(\frac{20}{3}\right) = \frac{10}{9}$$

$$D(O_2^w) = \frac{1}{N}\left(D_1(O_2^w) + D_2(O_2^w)\right) = \frac{1}{4}\left(\frac{17}{9}+\frac{10}{9}\right) = \frac{3}{4}$$

So, matrix M_2 has an orthogonality index of $\left(1-\frac{1/6}{3/4}\right)\cdot 100 = \frac{7}{9}\cdot 100 = 77.78\%$

Matrix M_1 has, obviously, an orthogonality index of 100%.

7.2.2. Confusion index (CI)

If two columns in a design matrix have the same sequence of levels (that is, they are equal when coded), the main effects of those two factors are totally confounded, and there is no way to remove that confusion without conducting more experiments. Of course, that is an extreme situation, where both factors are fully correlated. However, if some degree of correlation exists between factors, this phenomenon of aliasing also occurs, but in an even more complex manner, as partial aliasing appears.

To avoid problems when interpreting results, we would like to have uncorrelated factors (or at least, with very small correlations). Our aim is having a design matrix that will work well for estimating all main effects. Notice that my worries are focused only on main effects, and not on interactions. Although this might seem a bit unambitious, guaranteeing a good design matrix for main effects is a lot in the realm of KE studies[101].

As factors in KE studies are – almost always – nominal variables, the common Pearson correlation coefficient used for continuous magnitudes cannot be employed. Instead, my proposal is using Cramer's V. Cramer's V is a statistic measuring the strength of association between two nominal variables in a contingency table. It is named after the Swedish mathematician Harald Cramér.

---

101 In conjoint analysis, interactions are also almost always neglected, and main effects orthogonal designs are usually employed. The menu option "Generate orthogonal design" from SPSS, for instance, does exactly this.

Suppose X and Y are two factors. X has M different levels, labeled $X_1, \dots, X_M$. Y has N different levels, labeled $Y_1, \dots, Y_N$. X and Y can be arranged as a contingency table, as shown in Figure 7.1[102].

$$
\begin{array}{c|cccc}
X\backslash Y & Y_1 & Y_2 & \cdots & Y_N \\
\hline
X_1 & n_{11} & n_{12} & \cdots & n_{1N} \\
X_2 & n_{21} & n_{22} & \cdots & n_{2N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
X_M & n_{M1} & n_{M2} & \cdots & n_{MN}
\end{array}
$$

Figure 7.1. A contingency table with factors *X* and *Y*

In this contingency table, cell (i, j) contains the count n_{ij} of occurrences of level X_i in X and level Y_j in Y. n is the total number of pairs that can be done, and $n = \sum n_{ij}$.
The chi-squared statistic χ^2 can be computed from this contingency table. Then, Cramer's V is defined as:

$$V = \sqrt{\frac{\chi^2}{n \, min(M-1, N-1)}}$$

In a design matrix with r factors, there are $k = \binom{r}{2}$ pairs of factors. For each of this pairs of factors pf_i, with $i = 1, \dots, k$, its correlation V_{pf_i}(by means of its Cramer's V statistic) can be computed. I define the confusion index (CI) in the following manner:

$$CI = \left(1 - \max\left(V_{pf_i}\right)\right) \cdot 100$$

Only the worst pair of factors (those having the highest correlation index) is considered. However, the confusion index is not enough to summarize the correlation structure of main effects, as relevant information can be gathered from looking at the correlations.

Taking again matrix M_1 from Table 7.7, correlation between all 3 factors (AB, AC and BC) are all 0, so the confusion index is $CI = (1 - 0) \cdot 100 = 100\%$.

On the contrary, matrix M_2 has a correlation between factors A and B of 0.71. Its confusion index is, thus, 29%.

7.2.3. Examples of application of the suitability indexes

The use of both suitability indexes introduced in the previous subsection helps when assessing a design matrix for a KE study, before the data collection starts. I propose a

[102] This introduction to Cramer's V statistic is inspired in the Cramer's V entry from the online collaborative encyclopedia PlanetMath.org (http://planetmath.org/encyclopedia/CramersV.html, accessed November 2010).

visual representation of both the OI and the CI on a color scale, with a red, yellow and green area:

- Red area goes from 0% to 75%.
- Yellow areas goes from 75% to 90%
- Green area goes from 90% to 100%.

Taking again our first simulation sets from Section 7.1, Figure 7.2 shows the suitability indexes for simulation sets 1 and 2 (remember: a full factorial design). Of course, there are no problems: factors are totally uncorrelated and the matrix is orthogonal. Results were correct when using an ordinal logistic regression.

	Color	Sleeves	Printing	F1	F2	F3
1	Red	Long	Picture	A	A	A
2	White	Long	Picture	B	A	A
3	Red	Short	Picture	A	B	A
4	White	Short	Picture	B	B	A
5	Red	Long	Plain	A	A	B
6	White	Long	Plain	B	A	B
7	Red	Short	Plain	A	B	B
8	White	Short	Plain	B	B	B
9	Red	Long	Text	A	A	C
10	White	Long	Text	B	A	C
11	Red	Short	Text	A	B	C
12	White	Short	Text	B	B	C

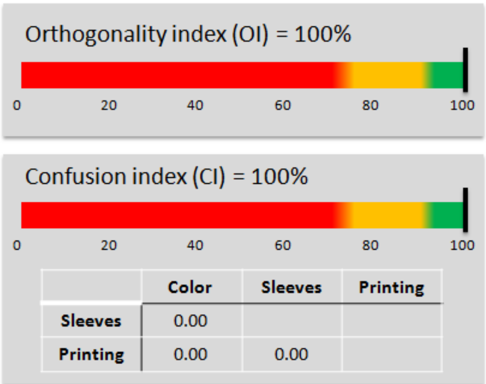


Figure 7.2. Orthogonality and confusion indexes for the design matrix of sets 1 and 2.

Simulation set 3 came from simulation set 1, but erasing 4 runs. The selection of these 8 runs from the total of 12 available was done trying to guarantee the estimation of main effects free of aliasing among them. This purpose has been fulfilled, as Figure 7.3 shows (CI=100%), although the matrix is no longer completely orthogonal. Results were also correct either using OLR or mOLR.

	Color	Sleeves	Printing	F1	F2	F3
1	Red	Long	Picture	A	A	A
2	White	Long	Picture	B	A	A
3	Red	Short	Picture	A	B	A
4	White	Short	Picture	B	B	A
5	Red	Long	Plain	A	A	B
8	White	Short	Plain	B	B	B
10	White	Long	Text	B	A	C
11	Red	Short	Text	A	B	C

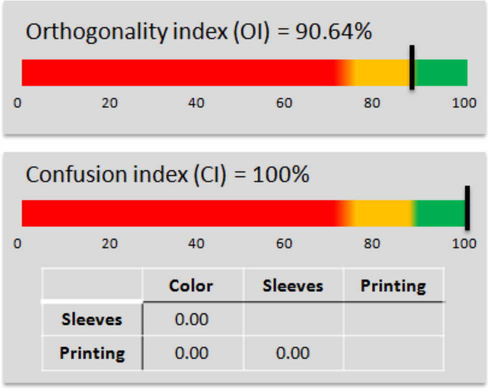


Figure 7.3. Orthogonality and confusion indexes for the design matrix of simulation set 3.

Things become worse with simulation sets 4 (Figure 7.4) and 5 (Figure 7.5). In simulation set 4, factors are correlated. However, the results obtained were quite good: mOLR correctly detected Color and Printing as significant, whereas OLR only detected Color as significant.

	Color	Sleeves	Printing	F1	F2	F3
1	Red	Long	Picture	A	A	A
3	Red	Short	Picture	A	B	A
4	White	Short	Picture	B	B	A
5	Red	Long	Plain	A	A	B
8	White	Short	Plain	B	B	B
10	White	Long	Text	B	A	C
11	Red	Short	Text	A	B	C

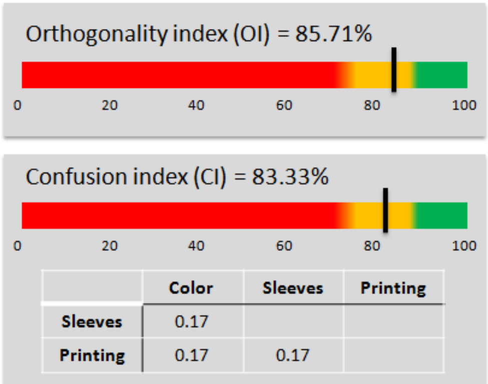


Figure 7.4. Orthogonality and confusion indexes for the design matrix of simulation set 4.

On the contrary, simulation set 5 has a CI of 50%. Two factors, Sleeves and Printing have a correlation of 0.50, too high for being able to estimate its effects free of confusion. No tool was able to correctly detect Color and Printing as the significant effects. Sleeves was wrongly declared significant because of the confusion with Printing.

	Color	Sleeves	Printing	F1	F2	F3
3	Red	Short	Picture	A	A	A
4	White	Short	Picture	B	A	A
5	Red	Long	Plain	A	B	B
8	White	Short	Plain	B	A	B
10	White	Long	Text	B	B	C
11	Red	Short	Text	A	A	C

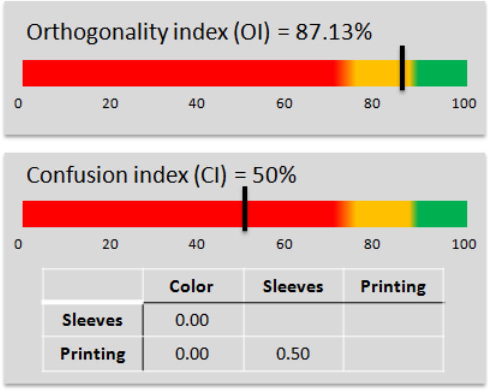


Figure 7.5. Orthogonality and confusion indexes for the design matrix of simulation set 5.

Suitability indexes: way to proceed

Based on the previous discussion, I propose the following method to evaluate the suitability of a design matrix for a KE study:

1. Calculate the orthogonality index and the confusion index, together with the correlation index for all pairs of factors[103].
2. Look at the color signal the orthogonality index and the confusion index give. Decide then what to do according to the criteria explain in Table 7.8[104]:

Table 7.8. Action plan for defining the design matrix depending on the color signals.

	Two green signals	The design will work well for calculating main effects. Go ahead!
	One or two yellow signals	The design might have problems. Try to improve it.
	One red signal (regardless of the other signal's color)	The design will almost sure have problems. Improve it.

How can we improve the design if we fall in the yellow or red areas? There are three different possibilities:

1. Increase the number of rows in the matrix (in a way that both OI and CI increase). This is the most obvious option and probably the best, but unfortunately, money or time restrictions make it impossible in many occasions.

[103] I have written a script in R to automatically calculate all these indexes.
[104] I have borrowed the idea of using a semaphore-like color code from pre-control charts.

2. Collapse levels in some factors. For example, imagine Color is a factor with 5 levels: blue, cyan, orange, red and yellow. These 5 levels can be collapsed into only 2: cold colors (blue and cyan) and warm colors (orange, red and yellow).
3. Join two different factors in one. This is a good option if the correlation between these two factors is high: the most straightforward way to solve the problem is making just one. For example, in the fruit juice experiment, Container and Decoration could have been, if necessary, collapsed in a new factor called Presentation.

If none of the previous three proposals work, there is always the possibility to drop a factor from the design (the one more prone to be non significant). Of course, this is a painful decision, but better than proceeding with the KE study and getting wrong conclusions, never discovering that they were wrong.

I have skipped the issue of interactions until now. This will be the topic of next section.

7.3. Introducing Interactions in the Synthesis Phase

I briefly talked about interactions in Section 5.6.2 of the literature review: interactions are hardly ever considered in Kansei Engineering. Sometimes, they are included under the more general term of non-linearities (which is, in fact, not incorrect, as an interaction is a second-order term in a model, although a bit special one). The use of neural networks in KE studies is in occasions justified by the need to consider these non-linearities in the model.

Table 7.9. Results from the T-shirts experiment (simulation set 6)

	Color	Sleeves	Printing	Carla	Joan	Marc	Maria	MEAN	
1	Red	Long	Picture	5	6	7	5	5.75	1
2	White	Long	Picture	3	4	5	3	3.75	2
3	Red	Short	Picture	7	4	5	6	5.50	3
4	White	Short	Picture	6	4	4	7	5.25	4
5	Red	Long	Plain	3	4	7	1	2.75	5
6	White	Long	Plain	1	3	2	1	1.75	6
7	Red	Short	Plain	3	5	5	2	3.75	7
8	White	Short	Plain	4	6	5	2	4.25	8
9	Red	Long	Text	5	6	6	4	5.25	9
10	White	Long	Text	1	2	5	1	2.25	10
11	Red	Short	Text	5	6	5	3	4.75	11
12	White	Short	Text	5	5	4	2	4.00	12

In this dissertation, I deal with interactions in the way done in factorial designs: I think this is the most understandable approach, and the most useful for obtaining practical conclusions. Table 7.9 shows a new simulation set (number 6). The design matrix is the same as in simulation set 1, but the model has been changed to include now an interaction.

Figure 7.6, in the left, shows the main effects plot for factor Printing. The conclusions for this factor are the same as the ones in Chapter 6. However, factor Color cannot be interpreted alone anymore, as it interacts with factor Sleeves (Sleeves was an inert factor in Chapter 6). What does this mean? The effect of factor Sleeves in the response depends on the level of factor Color: when Color is white, T-shirts with long sleeves are perceived as non colorful, whereas T-shirts with short sleeves are perceived as colorful. On the contrary, when Color is red, T-shirts are perceived as colorful, with no differences between long and short sleeves.

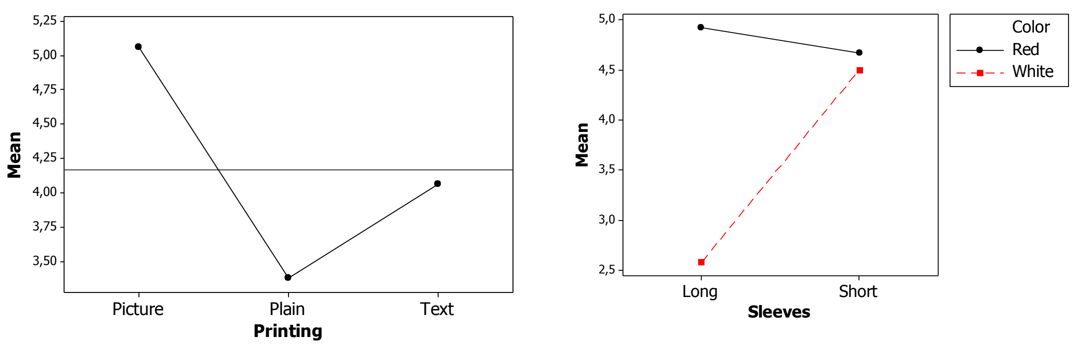


Figure 7.6. Main effects plot for factor Printing and interaction plot for factors Color and Sleeves, with data from Table 7.9.

Interactions are very important in experiments done in an industrial context and neglecting them is a mistake. In Kansei Engineering, interactions can be as important as in an industrial atmosphere, or even more (because the kind of factors used in KE studies are probably more prone to produce interactions[105]).

I think there are two main reasons that explain why interactions are usually not considered in Kansei Engineering:

- The first reason is that QT1 is designed to only include main effects in the model, and no interactions (we could say that interactions are "not supported" in QT1). As QT1 is very popular as an analysis tool in the

[105] Although this statement is only my perception and it is not based on data. In fact, reviewing the existing literature does not allow analyzing if interactions are common in KE studies, because they are almost never considered.

synthesis phase, interactions are often not considered. When another regression technique is used that allows the presence of interactions (a linear multiple regression, an ordinal logistic regression), interactions are not included because they tend to mimic the behavior of QT1.

For sure, some technicians using QT1 are aware of this fact, but maybe consider it a price to pay for the simplicity of interpretation of QT1. Having an interaction in the model always implies more complexity, and the nice representation of category scores as bars cannot be directly done with interactions. Probably, other technicians do not know about the importance of interactions, and they are happy just looking at the main effects.

- The second reason for not considering interactions is that, very often, they cannot be estimated with the collected data. As we have just seen in the previous section, the design matrix (the combination of factors and levels) determines the effects that can be estimated. Sometimes, the set of stimuli is so short that only main effects can be estimated (perhaps interactions can be calculated, but they are aliased with other interactions or even with main effects).

  If possible, I would recommend having a suitable design matrix for estimating interactions (a fractional factorial design with resolution V, or even IV, would be perfect). However, this is not feasible many times, because the number of stimuli needed would be too large. A common strategy in this situation is only bothering about main effects (this is usually the approach in conjoint analysis). At least, one must be aware of this. What I think cannot be accepted is, once the data is collected, not considering interactions when they can be estimated free of aliasing, and before collecting data, not trying – at least – to have a design matrix that allows the estimation of interactions.

7.3.1. Interactions in linear and logistic regression

Using data from Table 7.9 (and working with the mean for each row), assume we already know there is a significant interaction between factors Color and Sleeves (not asking too much, as data was simulated). Row labelled as MODEL 1 in Table 7.10 shows the coefficients when performing a linear regression with all linear terms and that significant interaction (each factor has a reference level: red for Color, long for Sleeves and picture for Printing). Including an interaction between dummy variables is like fitting two different models: one when Sleeves are short (MODEL 2) and another one when Sleeves are long (MODEL 3). Predictions using either model are the same. In some occasions, having a single model (like MODEL 1) is preferred; but probably interpretation is easier splitting that single model in two different models, one for each level of factor Sleeves (MODEL 2 and MODEL 3).

Table 7.10. Coefficients in the linear regression models with data from the T-shirts experiment with interaction

	Constant	Color (white)	Sleeves (short)	Color (white) x Sleeves (short)	Printing (plain)	Printing (text)
MODEL 1	5.8125	-2.3333	-0.2500	2.1667	-1.6875	-1.0000
MODEL 2 Sleeves (short)	5.5625	-0.1666			-1.6875	-1.0000
MODEL 3 Sleeves (long)	5.8125	-2.3333			-1.6875	-1.0000

Section 6.2.2 showed how to use an ordinal logistic regression in KE studies, but with no interactions. Of course, a logistic regression model can also include interactions, but interactions in logistic regression can be tricky to interpret. Again, assume we already know that interaction between factors Color and Sleeves is significant in an ordinal logistic regression with data from Table 7.9. The linear predictor $g(x)$ of the regression is the following:

$$g(x) = \beta_{0g} - 3.0184\, Color(white) - 0.2829\, Sleeves(short) + \\ +2.7698\, Color(white) \cdot Sleeves(short) - 2.1523\, Printing(plain) \\ - 1.3169\, Printing(text)$$

β_{0g} is the threshold constant (different for each cutting point; as we have 7 different categories in the response, $g = 1 \ldots 6$, and of no interest to us). Interpretation of factor Printing is straightforward, as it does not interact with any other factor. Odds ratios of all levels with respect to the reference (picture) are the following:

$$OR_{Printing(plain)\ vs\ Printing(picture)} = e^{-2.1523} = 0.116$$

$$OR_{Printing(text)\ vs\ Printing(picture)} = e^{-1.3169} = 0.268$$

$$OR_{Printing(picture)\ vs\ Printing(picture)} = e^{0} = 1$$

This last OR is obviously 1, as picture is the reference level for factor Printing. However, explicitly showing it facilitates interpretation and visual representation, as we have seen in Section 6.2.2.

Interpretation of factors Color and Sleeves is more complicated, as they interact. The OR for Color(white) vs. Color(red) is $e^{-3.0184} = 0.049$, but only for T-shirts with long sleeves (the reference level of the other factor interacting with Color). Similarly, the OR for Sleeves(short) vs. Sleeves(long) is $e^{-0.2829} = 0.754$, but only for red T-shirts.

What does the coefficient of the interaction term represent? When Color is red, the OR (having long Sleeves as reference) is 0.754, as we have seen. Conditioned to the fact of having white Color, the OR (having long Sleeves as reference) is $e^{-0.2829+2.7698} = 12.024$. If there was no interaction between Color and Sleeves, OR with Color red should be the same as OR with Color white, so the ratio of both of them should be equal to 1. In this occasion of a significant interaction between Color and Sleeves, the ratio is $12.024/0.754 = 15.95$, far away from 1. The exponent of the interaction coefficient is precisely $e^{2.7698} = 15.95$.

The exponent of interactions coefficients are, thus, ratios of odds ratios (a ratio of a ratio of a ratio of probabilities!) (Jaccard 2001). The whole procedure is quite confusing and unfriendly for technicians not skilled in logistic regression (and even for those who do know how to correctly interpret the coefficients). Our interest in the synthesis phase of a KE study is discovering how each item affects each Kansei word, and which categories of each item are preferred for conveying the desired emotions. Two factors interacting must be always interpreted together. So, in my opinion, the best thing to do is calculating OR for all possible combinations of levels for both factors interacting, one combination having both factors in their reference levels.

In our T-shirts' example, the results are the following:

$$Color: red \begin{cases} OR_{Sleeves(long) vs\ Sleeves(long)} = e^{0} = 1 \\ OR_{Sleeves(short) vs\ Sleeves(long)} = e^{-0.2829} = 0.75 \end{cases}$$

$$Color: white \begin{cases} OR_{Sleeves(long) vs\ Sleeves(long)} = e^{-3.0184} = 0.05 \\ OR_{Sleeves(short) vs\ Sleeves(long)} = e^{-3.0184-0.2829+2.7698} = 0.59 \end{cases}$$

7.3.2. Visual representation adapted to interactions and detection of significant interactions.

The visual representation using bars introduced in Section 6.2.2 greatly facilitates interpretation of results: one must choose those categories with longest bars to convey the emotion described by the Kansei word studied (and the opposite, select the shortest bars if the purpose is not conveying that emotion).

This visual representation can be adapted to allocate a significant interaction. To empathise the fact that both factors interacting cannot be interpreted separately, a different color is used to fill the bars representing those factors. Figure 7.7 shows the final result, and conclusions are easy to depict:

- On one hand, T-shirts with a picture are perceived as more colorful than T-shirts with text. Plain T-shirts are the ones assessed as less colorful.

- On the other hand, red T-shirts are perceived as similarly colorful regardless of being short or long-sleeved. Short-sleeved white T-shirts get a similar evaluation on colorful than red T-shirts, whereas long-sleeved white T-shirts are clearly perceived as the least colorful option.

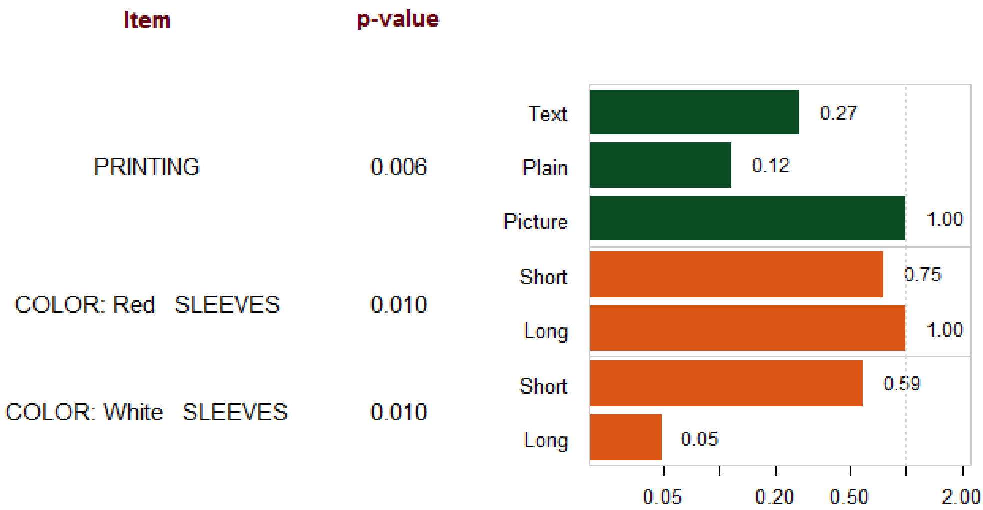


Figure 7.7. A visual representation of odds ratios from the ordinal logistic regression with data from Table 7.9.

The procedure showed in this subsection with an interaction in an ordinal logistic regression can be directly translated to a mixed effects ordinal logistic regression. Everything is the same, but subjects enter the model as a random effect. The final graph with odds ratios represented as bars is the same.

In the discussion until now, I have assumed we knew that the interaction between Color and Sleeves was significant. How do we know this? A common method for selecting the subset of variables that produces the best model is using a stepwise procedure. A forward stepwise could be used to select the best subset of main effects and interactions for the logistic regression model.

My proposal is limiting the number of interactions in the model to one. The main reason for this is that the usually not very large set of stimuli are probably enough for estimating all main effects, but quite often not enough degrees of freedom are left for estimating many interactions. If one interaction can be calculated, that is already something remarkable in this realm of KE studies. There are other practical reasons to limit the number of interactions:

- The KE studies proposed in this dissertation work with a relatively small set of factors (say, something between 1 and 8). Thanks to the effect sparsity principle (Wu, Hamada 2000) – which basically states that only a few effects in a factorial experiment will be statistically significant – assuming that at maximum one interaction is significant is not that risky.

- If more than one interaction is included in the logistic regression model, interpretation of results becomes very complex, and the visual representation using bars is no longer possible.

So my proposed procedure for detecting interactions in KE studies is checking, if possible, the significance of interactions in the model one by one. Probably none of them will be significant, or at maximum one of them. If more than one interaction should be included in the model, one has to consider if only the most clearly significant is used with the visual representation, or a more complex model is preferred.

7.3.3. Using rough sets to confirm interactions

As stated in Section 7.1, rough sets can be used as a check of results once conclusions have been reached using an ordinal logistic regression. This is especially true when an interaction is present. The nature of rough sets is giving rules where, usually, more than one factor is present: many times, rules comprise two factors at the same time. This is exactly the situation in an interaction, where both factors must be studied together.

As Figure 7.7 illustrates, T-shirts are perceived as colorful (1) either when Color is red, or when Color is white and Sleeves is short. Only when Color is white and Sleeves is long are T-shirts perceived as non-colorful (0). Table 7.11 show all rules involving two factors extracted using original rough sets with data from simulation set 6. Almost all rules involving Color red give 1 as result (this is not followed by the second listed rule, as Printing: plain “wins” and causes a result of 0). The first listed rule shows how Color: white and Sleeves: long give 0.

Table 7.11. Extracted rules using OrigRS with simulation set 6

COLOR: White	SLEEVES:◎Long	→	**0**
COLOR: Red	PRINTING: Plain	→	**0**
COLOR: White	PRINTING: Text	→	**0**
SLEEVES:◎Long	PRINTING: Plain	→	**0**
COLOR: Red	PRINTING: Picture	→	**1**
COLOR: Red	PRINTING: Text	→	**1**
SLEEVES:◎Short	PRINTING: Picture	→	**1**

The interaction is even more clearly using variable precision bayesian rough sets. Table 7.12 shows all rules involving two factors, ordered by its index of strength. Not only all rules having Color red give 1, but the first rule is explicitly Color: red → 1. The same happens with Sleeves: short → 1. Similarly to OrigRS, the interaction is correctly captured with the rule Color: white & Sleeves: long → 0.

Table 7.12. Extracted rules using VPBRS with simulation set 6

				str
	COLOR: Red	→	**1**	0.29
	PRINTING: Plain	→	**0**	0.27
COLOR: White	SLEEVES:◎Long	→	**0**	0.25
	SLEEVES:◎Short	→	**1**	0.20
SLEEVES:◎Long	PRINTING: Plain	→	**0**	0.15
COLOR: White	PRINTING: Plain	→	**0**	0.15
COLOR: Red	PRINTING: Picture	→	**1**	0.13
COLOR: Red	PRINTING: Text	→	**1**	0.13
COLOR: Red	SLEEVES:◎Short	→	**1**	0.10

So rough sets can always be used as a check of results. However, if something has not clearly been discovered with a logistic regression, it is difficult to derive it from the rules in rough sets. Rough sets is also influenced by the quality of the design matrix: if a design matrix have low suitability indexes, neither a logistic regression nor a rough set analysis will be able to reach correct conclusions.

7.4. The Synthesis Phase in the Fruit Juice Experiment, and a final simulation set.

This last section of this chapter resumes the fruit juice experiment first introduced in Chapter 4 as an example of a complete Kansei Engineering study. A final simulation set (simulation set 7) is later used to introduce one of the topics of next chapter.

7.4.1. The Synthesis Phase in the Fruit Juice Experiment

The design matrix in the fruit juice experiment is a 2^{5-1}, so we stayed in the safe area of a resolution V fractional factorial design. In Section 4.6 I analyzed the data collected averaging data over all subjects and using normal probability plots to identify the significant effects. Many "pages" have passed and I am now armed with better tools to reanalyze the data...

One of the aims of this chapter was comparing the behavior of QT1, OLR and mOLR. This has already been done with simulated data: it is now time for these tools to compete in the field of real data. Data has been modeled in a backward stepwise manner: all factors were first included in the model, and those non significant were progressively removed. For OLR and mOLR, all main effects and one interaction at a time has been introduced in the model. If an interaction was detected as significant, it was included in the model, and the backward stepwise process started with the interaction and all main effects (never more than one interaction appeared as significant) [106]. Table 7.13 summarizes all significant effects (with a level of significance of 5%) found using QT1 (first columns, red sign), OLR (second columns, green sign) and mOLR (third columns, blue signs). A rounded sign (•) indicates a significant effect. When two factors interact, they are both marked with a squared signal (▪).

Table 7.13. Significant effects with QT1 (red), OLR (green) and mOLR (blue) in the fruit juices experiment.

	Refreshing	Healthy	Exotic	Seductive	Natural	Relaxing	Tasty
Straw		■			■ ■	● ● ●	■ ■
Decoration			● ● ■	● ● ●			
Ice	● ● ■	■	● ● ●	● ● ■	■ ■		■ ■
Container		● ●	● ● ●	● ● ■	● ●		
◎Color	■		● ● ■	● ● ●			● ● ●

As stated before in Section 7.1, more significant effects appear when using OLR and mOLR than when using QT1. Besides, OLR and mOLR allow the use of interactions. OLR and mOLR give very similar results. Sometimes, factors that are significant only in their main effects in OLR appear interacting in mOLR. This happens with Kansei words Exotic and Seductive. Differences are, however, very small, as a detailed inspection of Figure 7.8 (with results for Seductive) shows.

The combined use of mOLR and interactions give interesting outputs, as the one obtained with the Kansei word Healthy (Figure 7.9). To convey a perception of a healthy fruit juice, a good choice is presenting it in a glass, without straw and without ice. If there is a straw, the presence or absence of ice does not seem to affect (exactly the same when there is ice, but no straw). The graphical representation is really useful to quickly understand the situation.

---

[106] I have written several scripts in R to automate all this modelling process.

mOLR results for Seductive

OLR results for Seductive

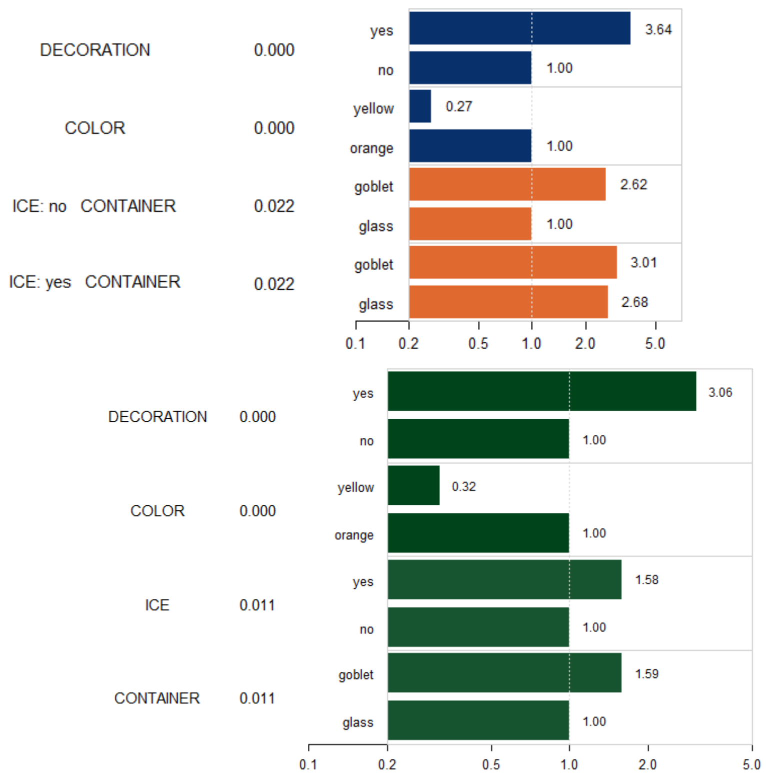


Figure 7.8. Results for Kansei word Seductive using mOLR (top) and OLR (bottom)

The fruit juice experiment will be more deeply analyzed in the next chapter.

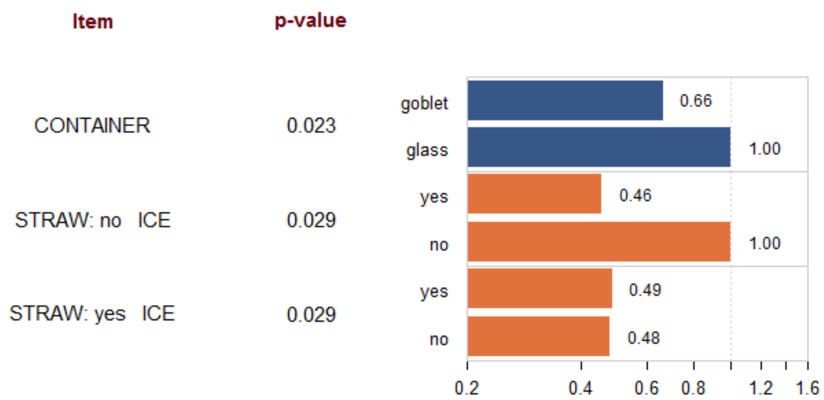


Figure 7.9. Results for Kansei word Healthy using mOLR.

7.4.2. A final simulation set

I finish this chapter with a small "trick"... Consider simulation set 7, shown in Table 7.14. When this dataset is analized either using QT1, OLR or mOLR, no significant effects appear. In particular, p-values for the three main effects using mOLR are shown in Table 7.15.

What happens? Are we missing something here?

Table 7.14. Results for the T-shirts experiment (simulation set 7)

	Color	Sleeves	Printing	Carla	Joan	Marc	Maria	MEAN	
1	Red	Long	Picture	6	3	1	6	4.00	1
2	White	Long	Picture	5	7	4	4	5.00	2
3	Red	Short	Picture	5	4	3	6	4.50	3
4	White	Short	Picture	5	4	5	4	4.50	4
5	Red	Long	Plain	5	4	5	5	4.75	5
6	White	Long	Plain	4	6	7	3	5.00	6
7	Red	Short	Plain	4	4	4	4	4.00	7
8	White	Short	Plain	3	6	6	3	4.50	8
9	Red	Long	Text	5	4	4	4	4.25	9
10	White	Long	Text	2	5	4	4	3.75	10
11	Red	Short	Text	5	2	3	6	4.00	11
12	White	Short	Text	5	5	4	5	4.75	12

This dataset has been generated in the following manner: the model for subjects Carla and Maria is the same as in simulation set 1; the model for subjects Joan and Marc is also the same, but with values of the main effects having opposite signs.

Table 7.15. p-values for each factor when using mOLR (simulation set 7)

Color	Sleeves	Printing
0.4454	0.7048	0.7331

Figure 7.10 represents the averages for each level, separating by the two groups.

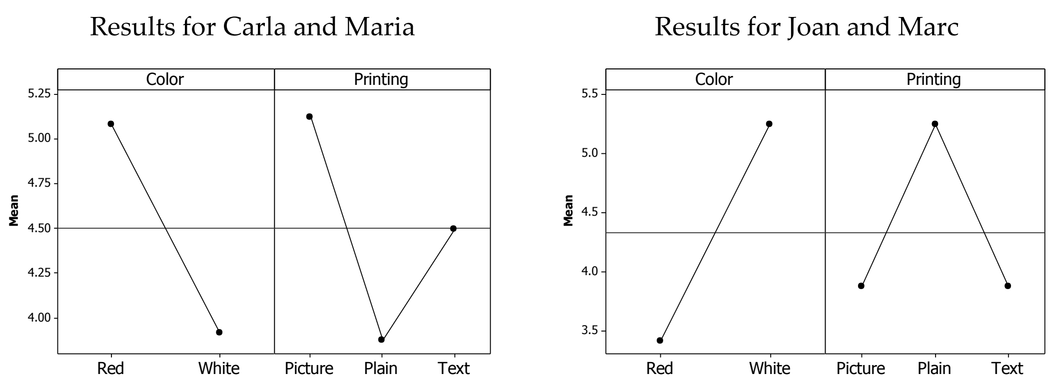


Figure 7.10. Main effects for simulation set 7, separated in two groups: Carla and Maria (left) and Joan and Marc (right)

When data from all subjects is analyzed together, we do not detect any significant factor. However, it is clear from Figure 7.10 that Color and Printing are significant, but subjects fall into two different groups. This topic will be treated in detail in the next chapter.

8 Addressing Heterogeneity Among and Within Participants

This chapter introduces some innovations in the analysis of data from Kansei Engineering studies. After detecting outliers, a methodology for performing an emotional segmentation of participants will be explained. Some proposals for computing a coherence index for participants will be suggested. Finally, two graphs will be proposed to summarize all the extracted information.

8.1. Two Important, but Undertreated Topics

When the fruit juice experiment was introduced in Chapter 4, I devoted its last section, Section 4.8, to share some thoughts that rose when carefully looking at the data. Basically, I addressed the following two topics:

- Heterogeneity among participants: Although the selection of subjects for a KE study is done choosing people from a clearly defined target group, once data is collected there are usually great differences among subjects. How can we treat this?
- Heterogeneity within participants: The perceived emotion when faced with a specific stimulus is only recorded once for each participant. What happens

when, a minute later, the same stimulus is shown to the same person? Does he or she give the same rating – or at least, a similar one – than the first time?

In this chapter I make proposals on these two topics.

A part of Section 5.5.1, in the literature review, was devoted to how several KE studies make comparisons on the results from several predefined groups. van Lottum, Pearce & Coleman (2006), for example, took data on the emotions elicited by shoes with people from two different countries: UK and Spain. The study is done separately for these two countries, and the results are then compared. This approach is followed in other papers (a list of these papers from the literature review can be found in Section 5.5.1). In some KE studies, demographic and socioeconomic data (age, gender, level of studies, etc.) is recorded for each subject before collecting Kansei data. This data could be used to make groups before or after the Kansei data collection, but this is not commonly done. But even if we do not collect additional data about the KE study participants, we surely have their ratings to all stimuli for every Kansei word. Why not using this data to check if different groups of people can be detected (people in the same group having a similar profile of ratings for all stimuli on a Kansei word)? Section 8.3 will offer a procedure to perform this so-called emotional segmentation.

The emotional segmentation is sensible to outliers so before conducting it the Kansei dataset must be free from outliers. Of course, the meaning of outlier in this marshy topic of emotions is not clear. And as always happens when analyzing data, outliers are not something to simply throw away: they can give loads of information. However, some people give ratings in a way so different from the others that keeping their responses could bias the global results. Section 8.2 offers a method to do this outlier detection.

As KE studies are based on the statistical analysis of collected data, data quality is something that must be taken into account. If we cannot rely on our measurement instrument, reached conclusions will be probably wrong no matter the sophistication of the posterior analysis. The idea of using reliable data is common in industrial environments, where gage R&R studies are often conducted in the framework of improvement projects (for example, in the Measure phase in Six Sigma projects). Validity and reliability are concepts also studied in statistical methods for social sciences. I am not aware of KE studies directly considering the topic of reliability. I will tackle this topic in Section 8.4. In particular, I will focus on how to deal with repeatability in Kansei, the degree of coherence among different ratings of the same stimuli and the same Kansei word within a participant.

8.2. Outliers in Kansei Engineering

Simply looking at the data, it is sometimes possible to detect participants in a KE study who give ratings in a weird way. Consider for example the ratings shown in Table 8.1 (with data from the fruit juice experiment). Some people, like Héctor or Núria, give ratings that cover the whole range of possibilities (from 1 to 7), and their profile of ratings seem perfectly "normal". On the contrary, some people, like Imma or Lydia, show an odder distribution of ratings. The results from Lydia for the Kansei word Exotic are an especially clear example. Why did Lydia give a rating of 1 to all stimuli but the first one? There are two possibilities:

1. She was not in the mood of doing the study, and she decided to write down something silly.
2. She really perceived all the juices as not exotic, so she heartily wrote down low ratings for all of them.

Table 8.1. Four participants results for Kansei words Exotic and Seductive in the fruit juice experiment

Exotic				Seductive			
Imma	Lydia	Héctor	Núria	Imma	Lydia	Héctor	Núria
1	3	1	3	5	4	3	4
1	1	1	3	1	1	1	3
1	1	2	5	1	1	1	4
1	1	5	7	1	1	7	5
1	1	1	4	1	1	1	1
1	1	5	1	1	2	4	3
3	1	6	6	3	1	5	5
2	1	3	7	3	1	2	6
1	1	1	5	3	1	1	3
2	1	3	5	1	2	6	4
3	1	3	6	3	2	5	5
2	1	2	6	1	1	1	7
1	1	6	4	2	1	4	5
2	1	1	3	2	1	4	4
1	1	7	7	3	1	3	7
2	1	5	5	2	3	3	6

It would have been very interesting discovering which one of the two possibilities is the right one; the only way to do it is keeping a conversation with Lydia. If the first possibility is right, one can soundly remove that data[107]. If the second one is right,

[107] In fact, what is really desired is not having people who give ratings in a silly way. This can be accomplished motivating subjects to participate in the study in a serious way, generating

thought-provoking information can be probably extracted talking to Lydia. In any case, the presence of ratings from one person very different from others' ratings might alter the results from the synthesis phase.

How can we detect unusual ratings once the data collection has finished? A thorough analysis of all the ratings for the word Exotic would lead to the consideration of Lydia's ratings for that word as an outlier. But the procedure is not always that simple (especially when all dimensions – participants, stimuli and Kansei words – must be taken into account at the same time). The multidimensional nature of KE data (several participants rate different stimuli on a number of Kansei words) makes the manual detection of outliers an almost impossible task.

A PhD thesis submitted in 2009 was devoted to the topic of detecting outliers in KE studies (Álvarez 2009). This work presents a methodology based on robust principal component analysis and discusses the consequences of outliers in KE. Two conclusions from this dissertation are worth mentioning here:

- As it might be expected, the higher the number of participants in the study, the less sensible the results are to the presence of outliers. Based on simulations done, 50 is an approximate number above which the results are quite robust to the presence of outliers. However, as indicated in Chapter 5, many KE studies are done with few participants (in particular, 65% of the reviewed papers used less than 50 participants).
- When the design matrix is orthogonal, the impact of outliers in the results is less than when the matrix departs from orthogonality [108]. However, orthogonality of design matrices is usually a topic either ignored or dealt in a superficial way in Kansei Engineering.

Outliers often change the results of the synthesis phase. Therefore, it is better to avoid them.

8.2.1. A methodology for detecting outliers

As usual detecting and taking into account outliers while analyzing Kansei data gives protection from reaching wrong conclusions. Furthermore, keeping outliers in our database will make the segmentation based on the emotional response that I will

---

the appropriate atmosphere and having a questionnaire short enough to avoid fatigue when answering.

[108] This is another reason that justifies the convenience of having orthogonal matrices in KE studies, and it adds up to all the reasons already presented in Section 7.2.

propose in Section 8.3 unfeasible. So the first step in Kansei data analysis has to be detecting and erasing outliers.

My selected methodology for detecting outliers is the one proposed in the above-mentioned PhD thesis (Álvarez 2009). Two situations are described in this work:

- Detecting outliers in the so-called simple Kansei tables (only considering one response, one Kansei word).
- Detecting outliers in the so-called multiple Kansei tables (considering all responses, all Kansei words, at the same time).

Outliers in simple Kansei tables

Consider n subjects rate m stimuli on a Kansei word. A simple Kansei table $\boldsymbol{X} = \{x_{ij}\}$, with $i = 1, \dots, n$ and $j = 1, \dots, m$ can be created (the rows are subjects and the columns are stimuli). x_{ij} is the rating given by participant i to stimuli j.

Detecting outliers in these simple Kansei tables can be done using the ROBPCA algorithm proposed by Hubert and Rousseeuw (2005)[109]. ROBPCA stands for Robust Principal Component Analysis. Its purpose is twofold: it allows the calculation of principal components resistant to outliers in the data and it gives a diagnostic plot that allows the detection of those outliers. We are especially interested in the diagnostic plot to identify participants giving ratings in a very different manner from the others.

The details of the ROBPCA algorithm are complicated. It uses projection-pursuit (PP) techniques (Li, Chen 1985) and the minimum covariance determinant (MCD) method (Rousseeuw 1984). The procedure can be summarized in the following steps (Hubert, Rousseeuw 2005):

1. Data is preprocessed by reducing their space to the subspace spanned by the n observations. This is done by singular value decomposition of the original matrix $\boldsymbol{X}$. The transformed data lies then in a subspace with a dimension that is, at most, $n - 1$.

[109] Methods as ROBPCA require working with continuous variables. We know that data from KE studies are usually ordinal, and not continuous. However, I think the procedure can be safely used when using the common 7-point scale response. In fact, the use of methods that require continuous variables when having ordinal variables is quite common: classical ANOVA is used all the time in sensorial analysis, where 5-point scales are usually employed. However, particular attention must be paid here: while all main statistical techniques suggested in this dissertation work well with ordinal data having few categories (even with binary data), ROBPCA would probably be inadequate for 3-point scales or binary data, for example.

2. A measure of outlyingness is calculated for each data point: the data points are projected on many univariate directions, each time the univariate MCD estimator of location and scale is computed and the standardized distance to the center is measured. The largest of these distances is the outlyingness measure of that data point. The data points with smallest outlyingness measures are used to compute a preliminary covariance matrix $\boldsymbol{S}_0$. This covariance matrix $\boldsymbol{S}_0$ is used for selecting the number of components p that will be retained.
3. The data points are finally projected on the p-dimensional subspace where their location and scatter matrix are robustly estimated.

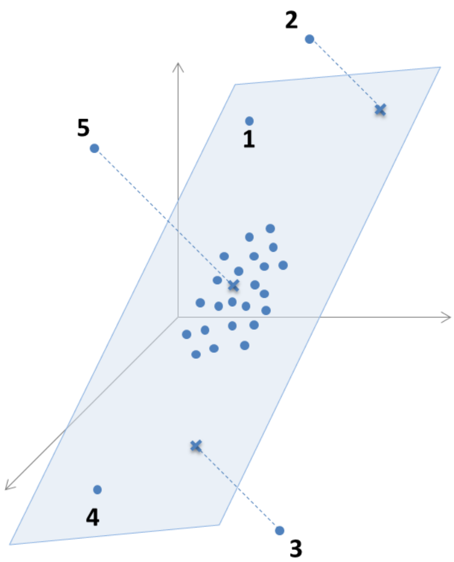


Figure 8.1. Different types of outliers when a three-dimensional dataset is projected on a robust two-dimensional PCA subspace, reproduced from Hubert and Rousseeuw (2005)

An extremely useful output from a ROBPCA is the diagnostic plot. The diagnostic plot allows the detection of outliers and the determination of its type. Consider the graph in Figure 8.1, where $m = 3$ and $p = 2$. Four types of observations can be described:

- Regular observations: they all form a homogenous group that is close to the PCA subspace.
- Good leverage points: they fall close to the PCA subspace, but far from the regular observations (points 1 and 4 in Figure 8.1).
- Orthogonal outliers: they have a large orthogonal distance to the PCA subspace, but cannot be detected as outliers if we just look at their projection on the PCA subspace (point 5 in Figure 8.1).
- Bad leverage points: they have a large orthogonal distance and its projection on the PCA subspace is far from the typical projections (points 2 and 3 in Figure 8.1).

The diagnostic plot places each observation in a scatterplot where the horizontal axis shows a robust score distance and the vertical axis shows an orthogonal distance. To classify the observations, two lines are drawn that divide the plot in four quadrants. These lines are drawn on the 97,5% quantiles for both the score distance and the orthogonal distance (which implies knowing the underlying distribution for these magnitudes, or at least an approximation). Details on how to calculate the distances for each observation and the dividing lines can be found in Hubert and Rousseeuw (2005).

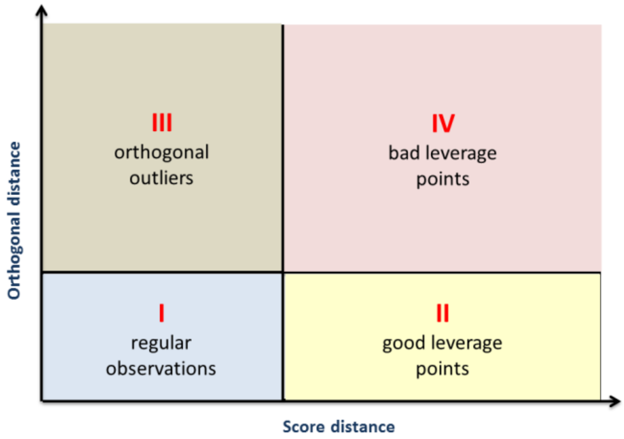


Figure 8.2. The diagnostic plot for detecting outliers according to the ROBPCA method

The diagnostic plot can be used to detect participants in a Kansei Engineering study that give ratings on a Kansei word in a manner very different from the others.

Outliers in multiple Kansei tables

Although it could be possible to detect outlier participants for each Kansei word (preparing a single Kansei table for each Kansei word), it would be interesting having a procedure to detect which participants can be flagged as outliers considering all Kansei words at the same time. This is done using a multiple Kansei table. The multiple Kansei table is created juxtaposing m tables (one for each stimuli in the study). Each table has n rows (participants) and r columns (Kansei words). Its structure is shown in Figure 8.3.

The method for detecting global outliers combines a multiple factor analysis (Escofier, Pagès 1994) and the ROBPCA method. Data in the multiple Kansei table is centered (columns' means are subtracted to each element) before starting. These are the two steps of the procedure:

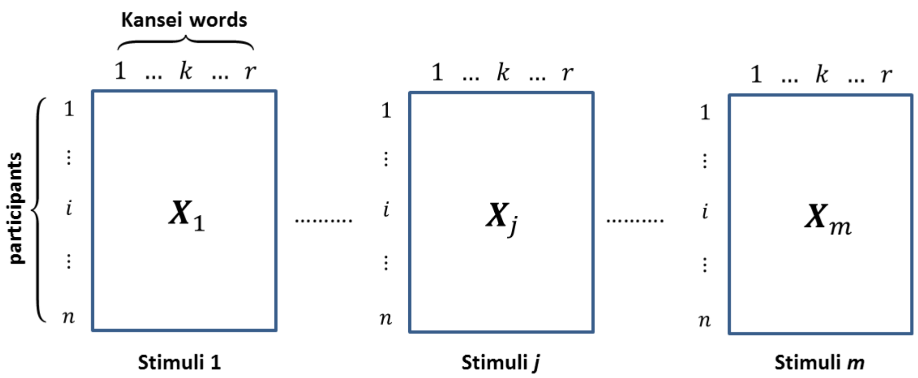


Figure 8.3. Structure of a multiple Kansei table.

1. The multiple Kansei table is treated as in a multiple factor analysis. A subtable could have a contribution in the first factorial axis much higher than the others. If this happens, the posterior analysis would be basically based only on this subtable. To bypass this undesired phenomenon, each subtable should be weighted. This is done through a principal component analysis (with all observations, including outliers) performed in each subtable. Each subtable is then "normalized" by dividing all its elements by the square root of the first eigenvalue obtained from its PCA.
   A balanced multiple Kansei table is then obtained:
$$\left({}^{1}\!\Big/\!{\sqrt{\lambda_1^1}}\, X_1 \;\mid \ldots \mid {}^{1}\!\Big/\!{\sqrt{\lambda_1^k}}\, X_k \;\mid \ldots \mid\; {}^{1}\!\Big/\!{\sqrt{\lambda_1^r}}\, X_r \right)$$
2. A ROBPCA is applied to the weighted juxtaposed table. This new table is considered as a simple table with n rows and $r + r + \cdots + r = m \cdot r$ columns. The diagnostic plot from ROBPCA allows then the detection of global outliers.

My proposal is detecting global outliers using the multiple Kansei table, and removing those participants flagged as outliers for the posterior analysis. As always when dealing with outliers in datasets, erasing those outliers with no reasoning is not recommended. In this case, it would be interesting discovering why some people gave so different ratings from their colleges in the experiment. This requires, almost surely, more particular information or even a personal interview with each of the "anomalous" participants. As I will explain in Section 8.5, I recommend complementing the results from the synthesis phase of a KE study with other qualitative information gathering technique (as interviews or focus groups). That moment can also be used to find the reasons for those outliers.

The methodology for detecting outliers has been applied to data from the fruit juices experiment[110]. Figure 8.4 shows the diagnostic plot for detecting outliers. Participants 14 (Lydia), 15 (Manolo) and 12 (Imma) appear as good leverage points. These participants could have been detected as outliers simply inspecting the dataset because they systematically give either low or high ratings:

- Lydia and Imma give too low ratings (74% of Imma's ratings are 1 and 2, 46% of Lydia's ratings are 1 and 2).
- Manolo give too high ratings (59% of Manolo's ratings are 6 and 7).

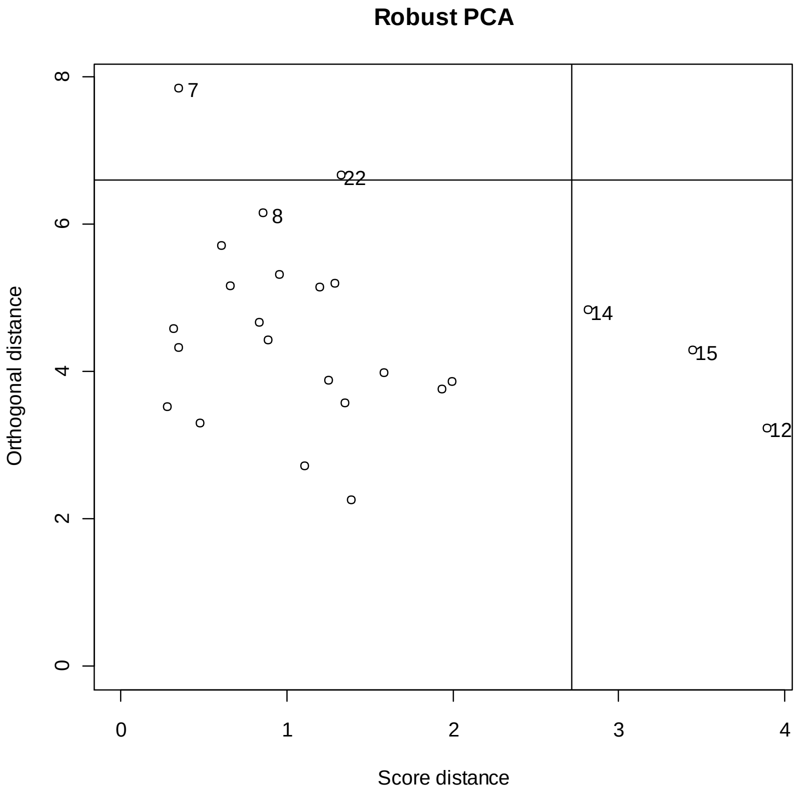


Figure 8.4. A robust PCA for detecting outliers in the fruit juice experiment.

Notice that the more extreme the ratings are, the higher score distance the participant has in the diagnostic plot.

---

110 The ROBPCA method is available in `R` using the function `PcaHubert` from the package `rrcov` (Filzmoser, Todorov 2009). A function has been written to perform the two-step process for detecting outliers in multiple Kansei tables: balancing the simple Kansei tables and conducting a ROBPCA on the juxtaposed multiple Kansei table.

Participant 7 (Erli) and 22 (Sergi) appear as orthogonal outliers. These participants cannot be detected as outliers simply inspecting the dataset. They are outliers because the profile of ratings from these two participants is far from the others.

Outliers identified as good leverage points (Lydia, Manolo and Imma) should be removed in order not to pollute posterior analysis. As stated before, it would be interesting finding – if possible – the reason for those extreme ratings.

It can be argued if participants identified as orthogonal outliers (Erli and Sergi) should be removed (their only sin is having a profile of ratings far away from the others). Probably these participants deserve a deeper analysis, but that needs personal interviews with them. However, if they are not removed, they will systematically appear as different groups when performing the emotional clustering that I will introduce in Section 8.3. My suggestion is thus also removing Erli and Sergi.

From now on, I will use the rest of participants (19 in total) for the fruit juice experiment.

8.2.2. Effects of outliers in KE studies

According to Álvarez (2009), the presence of outliers can affect conclusions from the synthesis phase in several ways:

- Some non-significant properties when having outliers can appear as significant without outliers.
- Some significant properties when having outliers can appear as non-significant without outliers.
- Some properties can be significant both with and without outliers, but its category scores can change its value and even its sign.
- The semantic space (Kansei words projected in the first two principal components) can be different depending on the fact of having outliers or not.

I assume the synthesis phase is done using QT1 or ordinal logistic regression. Similar consequences apply when the synthesis phase is done using rough sets.

Following the procedure explained in this section to detect anomalous participants is recommended, especially if the number of participants in the study is low (less than 50). A very reasonable approach is performing the synthesis phase with and without outliers to assess the magnitude of the changes.

Section 7.4.1 analyzed the data from the fruit juice experiment using QT1, ordinal logistic regression (OLR) and mixed effects ordinal logistic regression (mOLR). The results were summarized in Table 7.13 (this table is reproduced in Table 8.2 for convenience). Results were basically the same using the three methods of analysis for some words (Refreshing, Exotic, Seductive or Relaxing). For some other words, both flavours of ordinal logistic regression gave more factors as significant (this happens in Kansei words Healthy, Natural and Tasty).

Table 8.2. Significant effects with QT1 (red), OLR (green) and mOLR (blue) in the fruit juices experiment (reproduced from Table 7.13).

	Refreshing	Healthy	Exotic	Seductive	Natural	Relaxing	Tasty
Straw		🟦			🟩 🟦	🔴 🟢 🔵	🟩 🟦
Decoration			🔴 🟢 🟦	🔴 🟢 🔵			
Ice	🔴 🟢 🟦	🟦	🔴 🟢 🔵	🔴 🟢 🟦	🟩 🟦		🟩 🟦
Container		🟢 🔵	🔴 🟢 🔵	🔴 🟢 🟦	🟢 🔵		
◎Color	🟦		🔴 🟢 🟦	🔴 🟢 🔵			🔴 🟢 🔵

Table 8.3 shows the results of the analysis phase using the three proposed methods when all outliers from the dataset have been removed. Two interesting issues come up when comparing Table 8.2 (with outliers) and Table 8.3 (without outliers):

- The results from the three methods of analysis (QT1, OLR and mOLR) are more similar now. The only difference is in the Kansei word Relaxing (in this word, OLR and mOLR are able to detect some factors as significant – as usual – whereas QT1 is not).
- The significant effects without outliers with OLR and mOLR are similar to the significant effects with outliers with QT1. For example, no factors were detected as significant for Kansei words Healthy and Natural when having outliers and using QT1. OLR and mOLR flagged some factors as significant for these two words when having outliers, but give no signal of significant factors when working without outliers.

As QT1 collapses data from all participants using the mean, it gives more robust results in the presence of outliers. The higher the number of participants, the more robust QT1's results will be. On the contrary, OLR and mOLR take into account all individual data, without summarizing them. This causes an increased sensibility to outliers in OLR and mOLR.

Does this mean that QT1 is better than OLR or mOLR? If no process of outlier detection is executed prior to the synthesis phase, probably yes. At least, the results will be more robust to the presence of outliers. But if outliers are identified and removed, OLR and mOLR must be preferred, as they are able to give more accurate results (as I showed in Section 7.1). Furthermore, OLR and mOLR must be used after the emotional segmentation proposed in the next section, where they will show its maximum utility.

Table 8.3. Significant effects with QT1 (red), OLR (green) and mOLR (blue) in the fruit juices experiment, once the 5 people detected as outliers have been removed

	Refreshing	Healthy	Exotic	Seductive	Natural	Relaxing	Tasty
Straw						■ ■	
Decoration			● ● ●	● ● ●			
Ice	● ● ●		● ● ●	● ● ●			
Container			● ● ●	● ● ●			
◎Color			● ● ●	● ● ●		■ ■	● ● ●

8.3. Finding Segments Based on Emotional Responses

The common sequence in a KE study states that, after the data collection session, the synthesis phase starts. In the synthesis phase, regression analysis (I recommend mixed effects ordinal logistic regression) can be used to link the semantic space with the space of properties. As stated, a previous step before performing the regression analysis has been added: outlier detection. Outlier detection consists in identifying people giving anomalous ratings, so that the dataset can be free of outliers prior to the synthesis phase.

I will add yet another step prior to the regression analysis of the synthesis phase and after the outlier detection: emotional segmentation. Although some KE studies reported in the literature compare the results among different groups of people, these groups are always defined outside the data collected for the KE study. Segmentation of participants is done *a priori*. For example, Bahn et al. (2009) compare results from designers and users, van Lottum, Pearce & Coleman (2006) participants from two different countries, Huang et al. (2009) participants belonging to four different age groups and Mondragon, Company & Vergara (2005) participants who are production managers, university lecturers and machine tool operators.

My proposal is doing *a posteriori* segmentation, based on the ratings given to all stimuli on every Kansei word. In the next section I justify that the proposal makes sense both from a conceptual and a practical perspective.

8.3.1. Justifying the need of an emotional segmentation

Quite often in Kansei Engineering studies, demographic and socioeconomic data (age, gender, level of studies, etc.) is recorded for each subject before starting the proper Kansei data collection. This data is valued because it can be used, in one way or another, to make groups before or after the Kansei data collection. However, nothing is usually done with this data. If nothing is done, it could reasonably be argued not losing time collecting it.

In fact, as the target group for a KE study is clearly defined in the choice of domain phase, chances are that all subjects are very similar in demographic and socioeconomic terms. However, can we really infer from this social homogeneity that the emotions conveyed by products will be the same for all of them?

As in market research, requirements for collecting subjects follow guidelines such as: "women living in cities with more than one hundred thousand inhabitants, aged 25-30, from middle-upper class, that practice a non-competitive sport at least 3 hours per week". Imagine all girls in Figure 8.5 are in accordance with these guidelines. Just looking at her appearance, can we really believe that they all like the same kind of, say, watch? I think that is something rather unrealistic. Does not it make more sense to find groups based on their own ratings to Kansei words?

Figure 8.5. Five different girls, all in the same group according to its socio-economical status.

So asking questions such as age or education level is reasonable and should be done, but probably not for making *a priori* groups, but to try to characterize *a posteriori* groups. To recap:

1. Socioeconomic and demographic data should be collected together with the Kansei ratings.
2. Socioeconomic and demographic data should be used to try to typify the *a posteriori* groups, not to make them.

In addition to this conceptual reason for performing what I have called an emotional segmentation prior to the synthesis phase, a practical "numerical" reason is also available. If the emotional segmentation is not performed and data from all participants are analyzed together, chances are that active factors are not detected as significant. This is precisely what happened in simulation set 7, shown in Section 7.4.2. Three factors were used in that dataset: Color, Sleeves and Printing. None of them were detected as significant when data from all four participants were analyzed together. But in fact, Color and Sleeves were active factors, but with effects having different signs depending on the participants.

Although this phenomenon is very clear with simulated data as that shown in Section 7.4.2, it can also be observed in real data. Table 8.4 shows ratings for the Kansei word Refreshing in the fruit juice experiment. A careful inspection of this data illustrates that the profiles of ratings for Mònica and David are similar, and so are those for Antonio, Helena and Marta. However, profiles are quite different between the two groups of people.

Stimuli	Mònica	David	Antonio	Helena	Marta
1	5	4	2	4	2
2	3	1	4	6	4
3	3	1	5	5	5
4	3	2	3	4	3
5	4	5	5	6	5
6	4	5	3	3	3
7	4	7	2	3	4
8	5	6	5	6	5
9	3	1	5	6	6
10	3	4	3	3	2
11	5	6	2	4	4
12	3	1	5	6	4
13	3	4	3	3	4
14	4	5	4	6	5
15	5	6	3	5	5
16	4	4	2	5	4

Table 8.4. Ratings for the Kansei word Refreshing by two groups of subjects: Mònica - David and Antonio - Helena - Marta.

This can be more easily perceived in Figure 8.6. Ratings have been standardized (by substracting its mean and dividing by the standard deviation) to make comparison among participants clearer. If the synthesis phase is done only with ratings from Mònica and David, or only with ratings from Antonio, Helena and Marta, results will surely differ.

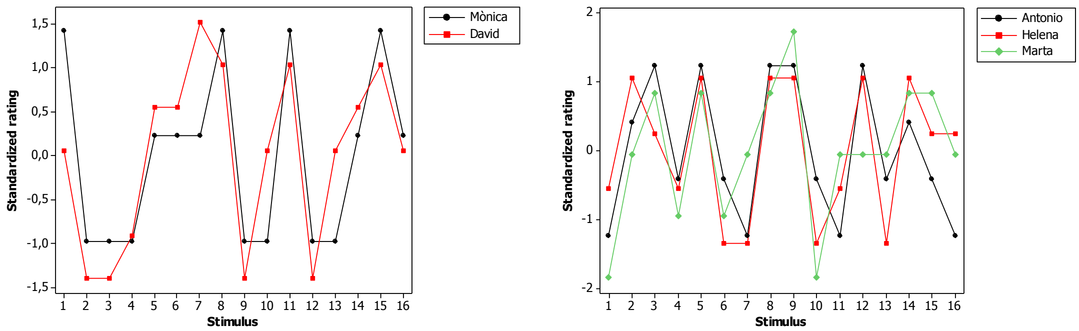


Figure 8.6. Profiles of ratings for the Kansei word Refreshing by two groups of subjects: Mònica - David and Antonio - Helena – Marta (standardized ratings)

How can we perform this emotional segmentation prior to the regression analysis of the synthesis phase? I think the most natural procedure is a cluster analysis, as the next section will explain.

8.3.2. Cluster analysis as the basis for the emotional segmentation

Manly (2005) gives a concise explanation of the purpose of cluster analysis:

> *Suppose that there is a sample of n objects, each of which has a score on p variables. Then the idea with a cluster analysis is to use the values of the variables to devise a scheme for grouping the objects into classes so that similar objects are in the same class. The method used must be completely numerical, and the number of classes is not usually known.*

With our datasets coming from Kansei Engineering studies, for each Kansei word, the objects are all the stimuli (there are *n* stimuli) and the variables are the participants (there are *p* participants). Many algorithms have been proposed for cluster analysis, but there are basically two approaches:

- Hierarchic methods. In hierarchic methods, a distance of each object to all other objects is calculated. Groups are then formed by agglomeration or division. With agglomeration, all objects start alone in groups of one. The closest groups are gradually merged until finally all objects are in a single group. With division, all objects start in a single group, and those more distant are progressively split until each object is in a group of its own. Agglomeration is more common than division.

In agglomerative hierarchic cluster analysis, once an object is placed in a group, it cannot leave that group in posterior iterations. Both the process and the results can be summarized with a dendogram (a very useful graphical representation of the agglomerative process).

- Non-hierarchical methods. Non-hierarchic methods involve partitioning, with objects being allowed to move from one group to another at different stages of the procedure. K-means is a common algorithm to perform this analysis. The first step is assigning each object to a group. This is done either introducing the number of desired groups and assigning each object to a group randomly, or defining the group for each object (based on previous information). Group centers are calculated as the average of objects in the groups. An object is moved to a new group if it is closer to that group's center than to the center of its current group. The process continues until no objects change from one group to another.

  One advantage of non-hierarchic methods over hierarchic methods is that objects can change groups during the process. However, the result usually depends on the initial placing of objects in groups, so it is more appropriate when there is some reasonable initial classification available. Furthermore, a dendogram cannot be used to illustrate the procedure and the results.

I have already suggested the use of cluster analysis in this dissertation as a help to reduce the number of Kansei words when spanning the semantic space in a KE study (refer to Section 4.3.2). In that occasion, I proposed a non-hierarchical cluster analysis, as initial groups were available as a result of an affinity diagram, and the purpose of the cluster analysis was simply refining those initial groups.

For the emotional segmentation here, my proposal is using a hierarchical cluster analysis for several reasons:

- There are no initial groups available and the number of groups is not previously known.
- It is easy to define a convenient distance and to incorporate it to the available functions for performing hierarchic cluster analysis (for example, using the function `hclust` in `R`).
- The use of a dendogram is very convenient to decide on the number of groups, and the results after testing the procedure with datasets from KE studies are good.

Two decisions must be taken before applying the hierarchic cluster analysis: the distance and the linkage method used.

Defining a distance for the hierarchic cluster analysis

In the final stage of a cluster analysis, observations are classified into clusters: objects in the same cluster are very similar among them, but are very dissimilar to objects in other clusters. Of course, "similar" means similar with respect to some criterion. This criterion can be specified by a distance between objects. In that way, very dissimilar objects must have a big distance, and very similar objects a small distance.

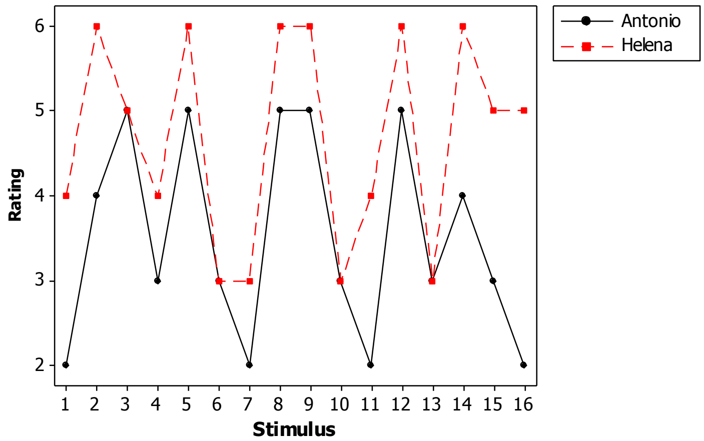


Figure 8.7. Profiles of ratings for the Kansei word Refreshing for participants Antonio and Helena (real ratings, without standardization)

The most commonly used distance is the Euclidean distance. This is the standard measure of distance (the square root of the sum of squared differences), the one given by a rule on a piece of paper. Other distances can be used, such as the Manhattan distance (the sum of absolute distances)[111]. However, we are not interested in distances that measure the proximity of observations, but in distances that take into account the correlation between ratings for all the stimuli by two different participants. This can be exemplified with Figure 8.7. Figure 8.7 shows the ratings by Antonio and Helena to all stimuli for the Kansei word Refreshing (with data from the fruit juice experiment). Helena systematically gives an average of 2 points – sometimes 1, sometimes 3 – more than Antonio to each stimulus. If the Euclidean distance is used, Helena and Antonio would appear as being far one from the other. But what really interests us is the fact that the profile of ratings by Antonio and Helena are very similar. Stimulus rated high by Helena are also rated high by Antonio; stimulus rated low by Helena are also rated low by Antonio (each one on a

[111] SPSS is the menu-driven software with more implemented distances for cluster analysis that I know (more than 30 in SPSS version 17!).

different reference level, we could say). Taking data from Helena or from Antonio would lead to the same conclusions on which effects are significant and how factors affect the response.

So what we need is a distance based on a correlational measure. In particular, I suggest the following – very simple – distance:

$$distance(i,j) = 1 - correlation(i,j)$$

The correlation between participants i and j goes from –1 (one profile completely opposite to the other) to +1 (both profiles identical, although not necessarily at the same level). So the distance goes from 0 (when both profiles are identical, both participants must be in the same cluster, as they surely produce the same results in the synthesis phase) to 2 (when participants must surely be in different clusters, as they produce opposite results in the synthesis phase).

How can we compute correlation between participants i and j? When 7-point scales are used, probably one can safely use the common Pearson product-moment correlation coefficient. However, as both vectors have ordinal data, a more appropriate correlation coefficient would be the polychoric correlation. The polychoric correlation assumes there are two continuous normally distributed latent variables, but they become ordinal variables when observed. I think this is a realistic assumption with ratings from KE studies (in fact, ratings are sometimes done on visual analogue scales, as I explained in Section 3.4.2, in an attempt to capture this continuous nature of the latent variables). Furthermore, Coenders, Satorra & Saris (1997) found that "*polychoric and polyserial correlations are somewhat robust to nonnormality of the underlying continuous variables*". Polychoric correlations have been found to perform better than Pearson correlations in factor analysis studies (Brown 1989), so I think its use in cluster analysis is suitable.

Holgado–Tello et al. (2010) give a brief description of the polychoric correlation. Suppose Z_1 and Z_2 are two ordinal variables with m_1 and m_2 levels each. Their sample distribution can be shown in a contingency table like the one shown in Table 8.5, where n_{ij} is the number of cases in level i of variable 1 and in level j of variable 2. If we assume that variables Z_1 and Z_2 are in fact generated by latent variables Z_1^* and Z_2^*, which are normally distributed, their combined distribution is a normal bivariate distribution with correlation ρ. The polychoric correlation is the correlation ρ in the following bivariate normal distribution:

$$P[X = i, Y = j] = p_{ij} = \int_{a_{i-1}}^{a_i} \int_{b_{j-1}}^{b_j} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} dxdy$$

It can be estimated by maximizing the likelihood of the multinomial distribution:

$$\ln L = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} n_{ij} \log p_{ij}$$

Details on the maximum likelihood estimation of the polychoric correlation coefficient can be found in Olsson (1979). I have computed polychoric correlations using the function `polychor` from the `polycor` package in `R`.

$$\begin{matrix} n_{11} & n_{12} & \cdots & n_{1m_2} \\ n_{21} & n_{22} & \cdots & n_{2m_2} \\ \vdots & \vdots & n_{ij} & \vdots \\ n_{m_1 1} & n_{m_1 2} & \cdots & n_{m_1 m_2} \end{matrix}$$

Table 8.5. A contingency table for showing the calculation of polychoric correlation.

Choosing a linkage method for the hierarchic cluster analysis

At the first step of an agglomerative hierarchic cluster analysis, when each observation constitutes a group, the distance between two groups is simply the distance between two observations. But after observations are joined together, a linkage method is needed to know how to calculate the distance between clusters. There are several linkage methods available (Hair Jr et al. 1995):

- Single linkage, or nearest neighbor. The distance between two clusters is the distance between the observation in one cluster and the observation in the other cluster that have the minimum distance.
- Complete linkage, or furthest neighbor. The distance between two clusters is the distance between the observation in one cluster and the observation in the other cluster that have the maximum distance.
- Average linkage. Similar to single linkage and complete linkage, but the distance is not the minimum or the maximum, but the average of all distances from all observations in one cluster to all observations in the other cluster.
- Centroid linkage. The distance between two clusters is the distance between the cluster centroids. Cluster centroids are the mean values of the observations in the cluster. Every time a cluster adds an observation, a new centroid is computed.
- Ward's linkage. In Ward's linkage, clusters are joined so that the information loss associated with each grouping is minimized. This loss is defined as an error sum of squares (the squares of the differences between observations and centroids). At each stage of the clustering procedure, the union of every cluster pair is considered. The clusters finally joined are those that result in

minimum increase in information loss, so the within-cluster sum of squares is minimized (Ward 1963).

Single linkage and complete linkage only use one observation in each cluster, and disregards all the others. Average and centroid linkage methods consider all observations (participants in this case). Ward's linkage is different from the others as it uses an ANOVA approach. Several studies – such as Milligan (1980) – show that no single linkage method can be consider superior for all types of data. Depending on the structure of the data, very similar results could be obtained no matter the selected linkage method. Trying some of them and checking if results are similar is a common recommendation. Personal preferences also play a role in deciding the linkage method for a hierarchic cluster analysis.

When the data contains a true cluster structure masked by the presence of noise, Ward's linkage is reported as having a very good performance (Morey, Blashfield & Skinner 1983, Dimitriadou et al. 2004, Blashfield 1976). Therefore, my recommendation is using the Ward's linkage in Kansei Engineering studies. Our main interest is detecting these true cluster structures. Nothing happens if some artificial clusters are created: when we analyze results inside each cluster, we can always join clusters if we detect the same results in several of them (these are situations that fit in the first possibility described in the next section).

8.3.3. The process of emotional segmentation

The emotional segmentation is performed Kansei word by Kansei word. The procedure is simple: for each Kansei word, stimuli are placed in columns and participants in the KE study in rows. A cluster analysis using the distance defined in Section 8.3.2 is then completed. The number of groups is decided "cutting the dendogram" at a desired level.

Both the process and the possible outcomes are more easily understood when explained with an example, so I will use the fruit juice experiment as an illustration. Take the Kansei word Refreshing to begin with. Figure 8.8 shows the resulting dendogram. A reasonable criterion to decide the number of groups is cutting the dendogram where the vertical lines are the longest – or at least quite long. As the vertical axis represent distances, long vertical lines imply that the distances between the clusters being split are high. Each cluster of participants will enter the synthesis phase separately, so a practical consideration is not having more than 3 or 4 clusters (the number of realistic clusters probably depends on the number of participants: the more participants, the more clusters). For Kansei word Refreshing, participants have been split into two clusters. Recall data from Table 8.4 and Figure 8.6 and notice how Mònica and David are in the same cluster, whereas Antonio, Helena and Marta are in the other cluster.

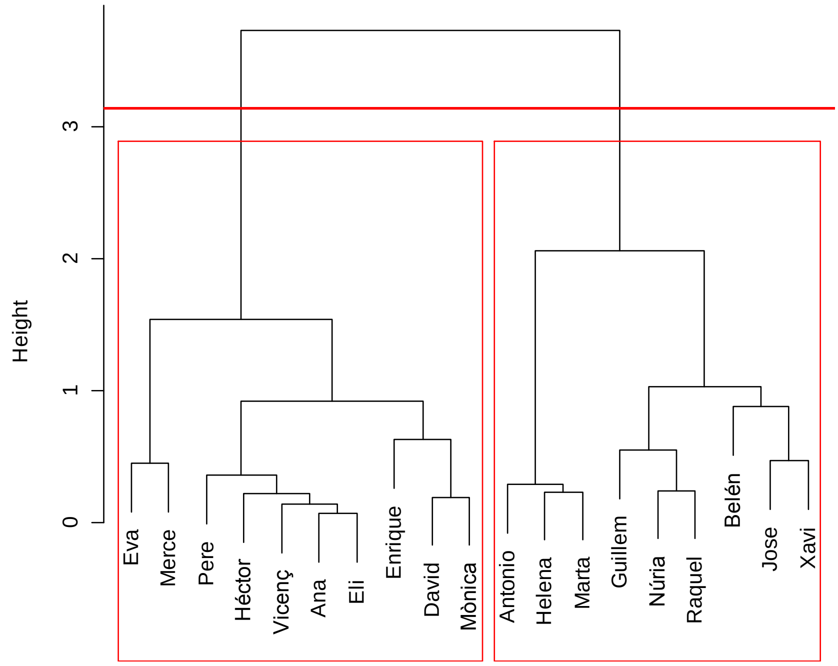


Figure 8.8. Dendogram for the Kansei word Refreshing

The synthesis phase (linking the semantic space with the space of properties) is then done separately for each cluster. Any of the methods explained in Chapter 6 are valid (although my personal preference is regression analysis, and mixed effects ordinal logistic regression in particular).

Table 8.6 shows the results for each group of participants (A and B) and for all participants analyzed together. Looking at the results, Straw and Decoration are always inert factors. Ice and Color interact when data from all participants are analyzed together; the same result that we have when only data from group A is used. In group A, looking at the interaction, ICE: yes gives much higher ratings than ICE: no. When having ICE: yes, Color is not important. When having ICE: no, COLOR: orange gives higher ratings than COLOR: yellow. In group B, COLOR: yellow gives higher ratings than COLOR: orange (something somehow opposite to what happened in group A). Ice and Container interact, the highest ratings are obtained for ICE: yes and CONTAINER: glass.

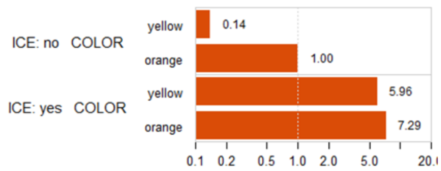
Table 8.6. Significant effects for Kansei word REFRESHING using mOLR

Ana	Antonio	Belén	David	Eli	Enrique	Eva	Guillem	Héctor	Helena	Jose	Marta	Merce	Mònica	Núria	Pere	Raquel	Vicenç	Xavi
A	B	B	A	A	A	A	B	A	B	B	B	A	A	B	A	B	A	B

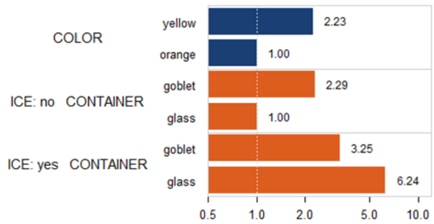
	Straw	Decoration	Ice	Container	Color
A + B			■		■
A			■		■
B			■	■	●

Ice affects in groups A and B in a similar way. In group A, color orange give higher ratings. In group B, color yellow give higher ratings. Container interacts with Ice in group B.

A



B



The procedure continues analyzing results for each Kansei word, and this understandably produces loads of information[112]. There is no shortcut to this process. However, it is possible to characterize two different kinds of situation:

1. Results from all the groups are more or less similar, and probably similar to what we obtained when analyzing data from participants together. In this situation, for the sake of simplicity, it is better not to keep the groups.

   Notice that it will be always possible to cut the dendogram at some point, so it will be always possible to create different groups. Our check to detect if keeping these groups makes sense or not is simply looking at the results from the synthesis phase in each of the groups: join the groups if results are very similar.

2. Results are different for each group. In this situation, it is better to keep the groups. It can happen that active factors are different in each group, or that a factor is active in several groups, but with different sign. This last possibility is especially relevant as we are probably detecting a factor as active when it was – wrongly – declared inert when data from all participants were analyzed at the same time (this is what happened in simulation set 7 in Section 7.4.2).

[112] It would be probably desirable to have a way to summarize all extracted information. I will make my proposal in Section 8.5.1

Of course, we can also have something in the middle of situations 1 and 2: several groups have the same results (so these ones are joined into one unique group), while some other groups give different results (so they are kept as separate groups).

Results from all the groups are more or less similar

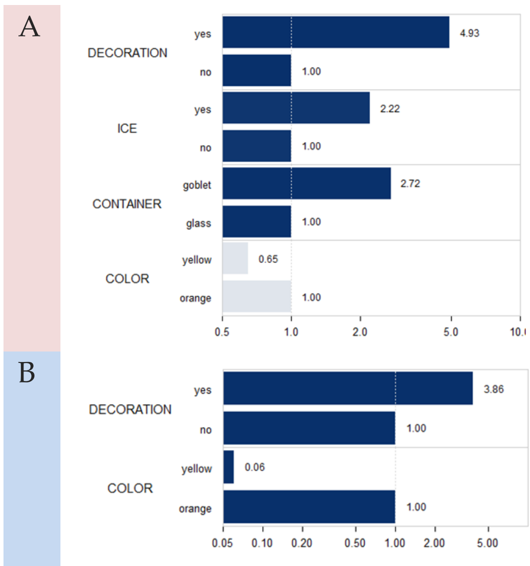
Going back again to the fruit juice experiment, Kansei words Exotic and Seductive (which we have already seen are quite similar) fit to the first described situation: “results from all the groups are more or less similar”. Consider first the results from Kansei word Seductive, shown in Table 8.7. More factors are active in group A than in group B, but the sign of the effects is the same. This is the kind of situation where joining the groups is recommended: interpretation of results is intricate enough, it is probably worthless to make it more complicated.

Table 8.7. Significant effects for Kansei word SEDUCTIVE using mOLR

Ana	Antonio	Belén	David	Eli	Enrique	Eva	Guillem	Héctor	Helena	Jose	Marta	Merce	Mònica	Núria	Pere	Raquel	Vicenç	Xavi
A	B	B	B	B	A	A	A	B	A	A	A	B	A	A	A	A	B	B

	Straw	Decoration	Ice	Container	Color
A + B		●	■	■	●
A		●	●	●	
B		●			●

Decoration and Color affect both groups A and B in a similar way. Ice and Container affect only in group A.



Something similar happens with Kansei word Exotic. This Kansei word, however, shows a peculiarity worth noting here. Figure 8.9 shows the dendogram for this response.

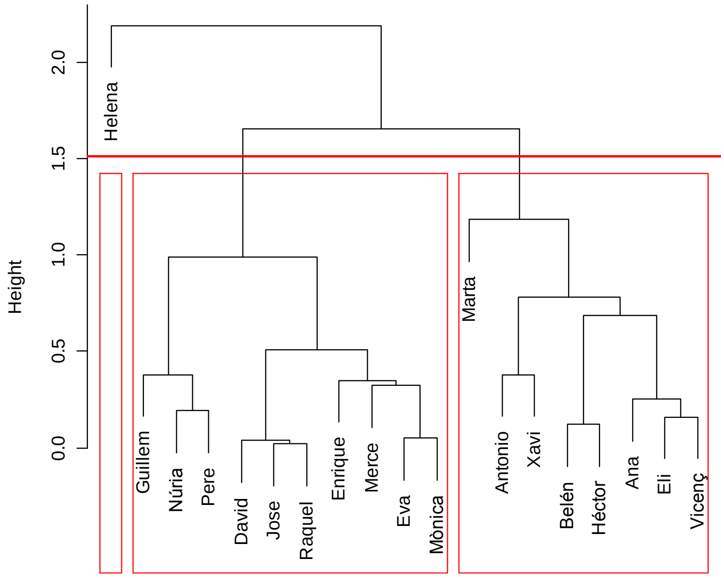


Figure 8.9. Dendogram for the Kansei word Exotic

Helena appears as a one-person group, very different from the others. Why does this happen? Simply because its profile of ratings for this Kansei word on all stimuli falls far from other's participants profile of ratings. Participants with ratings very dissimilar from the "normal" ones on all or many Kansei words have been detected as outliers and removed from the analysis. Helena was kept in the analysis, but she is an outlier just for this Kansei word. When a robust principal component analysis (ROBPCA) is performed on data only on the simple Kansei table from response Exotic, Helena clearly appears as an orthogonal outlier (Figure 8.10).

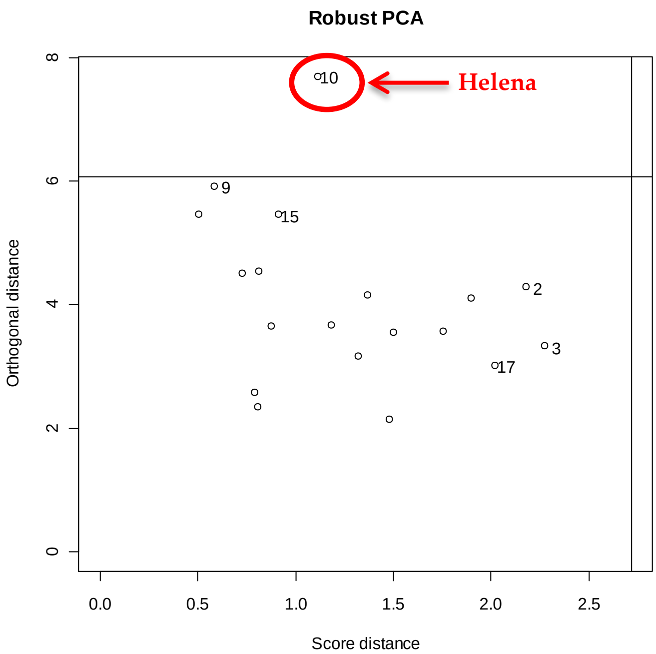


Figure 8.10. A robust PCA for the Kansei word Exotic. Subject 10 (Helena) appears as an outlier for this Kansei word.

The best option in this situation is not considering that participant only for this Kansei word. Doing the outlier detection explained in Section 8.2.1 is especially important for the emotional segmentation, as people giving anomalous ratings would systematically appear as one-person groups in the cluster analysis if not done. The outlier detection and the emotional segmentation should be always done – in this order – before performing the regression analysis for linking the semantic space with the space of properties.

Table 8.8 shows the results for Kansei word Exotic. As happened with Seductive, results are not the same for groups A and B, but they are similar enough to justify joining both groups into just one, and analyzing data from all participants together (setting Helena aside).

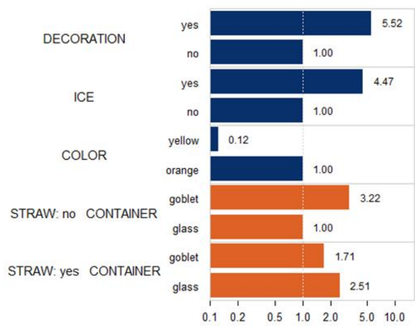
Table 8.8. Significant effects for Kansei word EXOTIC using mOLR

Ana	Antonio	Belén	David	Eli	Enrique	Eva	Guillem	Héctor	Helena	Jose	Marta	Merce	Mònica	Núria	Pere	Raquel	Vicenç	Xavi
A	A	A	B	A	B	B	B	A	C	B	A	B	B	B	B	B	A	A

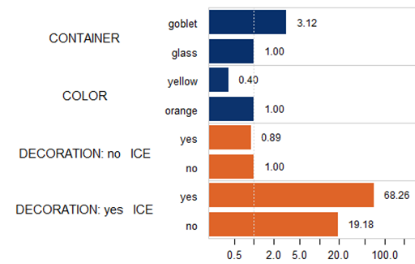
	Straw	Decoration	Ice	Container	Color
A + B		■	●	●	■
A	■	●	●	■	●
B		■	■	●	●

Decoration and Ice affects both groups A and B in a similar way (although it interacts with Ice in group B). Straw and container interact in group A, but not in group B. Straw is inert in group B. Color affects both groups A and B in a similar way.

A



B



Results are different for each group

When results are different for each group, we can be faced to three situations: active factors are different; active factors are the same but with different signs; or a combination of the first two.

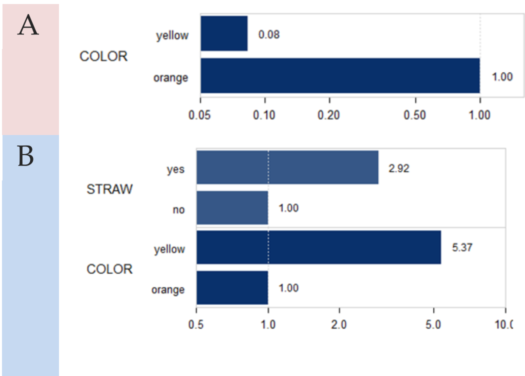
Consider the results for Kansei word Tasty, shown in Table 8.9. Straw appears as a significant factor in group B and as an inert factor in group A. But the most relevant issue here is the behavior of Color. Color appears as significant in both groups, but with different sign. COLOR: orange is perceived as more tasty in group A, and COLOR: yellow is perceived as more tasty in group B. Color also gave signal as significant when both groups A+B were analyzed together. In that situation, the results from group A were dominant, and COLOR: orange was perceived as more tasty. This happens because 15 participants are in group A whereas only 4 are in group B.

Table 8.9. Significant effects for Kansei word TASTY using mOLR

Ana	Antonio	Belén	David	Eli	Enrique	Eva	Guillem	Héctor	Helena	Jose	Marta	Merce	Mònica	Núria	Pere	Raquel	Vicenç	Xavi
A	B	A	A	A	A	A	B	A	A	B	B	A	A	A	A	A	A	A

	Straw	Decoration	Ice	Container	Color
A + B	■		■		●
A					●
B	●				●

Color affects both to groups A and B in opposite ways.
Straw also affects in group B.



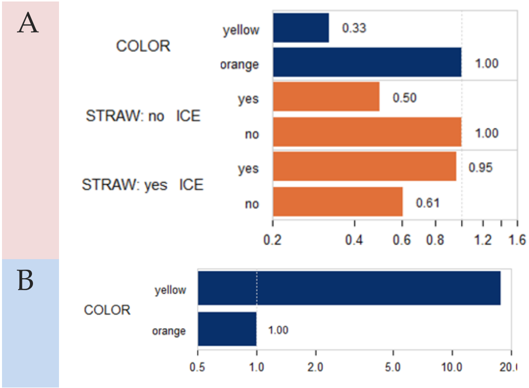
The results for Kansei word Natural are shown in Table 8.10. The same effect as in response Tasty happens with factor Color: it is active in both groups, but affecting with different sign. However, in this occasion Color did not give signal as active when data from both groups A+B were analyzed. The reason is that the number of people in each group is here more balanced (12 in group A and 7 in group B), so the effect of Color is compensated. We were missing an important conclusion here: we though Color was inert; but in fact Color has a significant effect (just that it is opposite from one group of people to the other).

Table 8.10. Significant effects for Kansei word NATURAL using mOLR

Ana	Antonio	Belén	David	Eli	Enrique	Eva	Guillem	Héctor	Helena	Jose	Marta	Merce	Mònica	Núria	Pere	Raquel	Vicenç	Xavi
A	B	A	A	A	A	B	A	A	B	A	B	A	B	B	A	A	B	A

	Straw	Decoration	Ice	Container	Color
A + B	■		■	●	
A	■		■		●
B					●

Color affects both to groups A and B in opposite ways.
Straw interacts with Ice in group A.



The rest of Kansei words (Healthy and Relaxing) are analysed in the same manner.

8.3.4. Looping the loop: clustering clusters of participants

The emotional segmentation has been done Kansei word by Kansei word. My aim now is doing a global grouping of participants, taking into account information from all Kansei words together. However, a note of attention must be done here. In general, participants fall in different groups for each Kansei word, so suggesting a global grouping of participants will necessarily oversimplify reality. In my opinion, if one tries to make simple something in fact complex by excessively summarizing and collapsing, chances are that the result falls into one of these two undesired categories:

- Everything seems simple, but nothing is really understood.
- Everything seems simple and seems understood, but it is false as it does not reflect reality.

I dare not say which option is worse. So it is better not to avoid complexity (Gummesson 2007).

Nevertheless, I will propose a method for doing this "clustering of clusters of participants" for two reasons:

- It might happen that every participant falls – basically – in the same group for all Kansei words. In this exceptional situation, making a global grouping of participants would absolutely make sense. The next step should be trying to characterize the groups. For example, imagine two global groups of participants are found, and one group is basically constituted by males and the other by females. That would be a remarkable conclusion, full of practical possibilities: women perceive a given product in a way different from men (and we know in which way each group perceives the product).

- Even if each participant falls in different groups depending on the Kansei word, some participants are globally more similar than others. Having a way to detect these similarities and dissimilarities is interesting if a qualitative study wants to be done later. I will propose in Section 8.5 complementing the information from the Kansei Engineering study with other qualitative techniques for gathering emotional perceptions (interviews, focus groups, etc.). Having a global clustering of participants can help in choosing a range of participants that cover the whole spectrum of opinions.

Cluster analysis for the global clustering of participants

The global clustering of participants will be done, again, using a hierarchic cluster analysis. I recommend Ward's linkage method, as I did before. A distance must be created for the cluster analysis: it must be low if participants are close (0 if both participants fall in the same groups for all Kansei words) and high if participants are far away.

Data from the fruit juice experiment will be used again to exemplify the procedure. Table 8.11 shows, for each Kansei word, the group where each participant was placed. Ana, Eli, Enrique, Héctor and Pere are always in the same group for all Kansei words, so they should constitute a cluster. Surely other participants are not exactly in the same groups for all Kansei words, but they do not differ in many.

Table 8.11. Groups for each participant and Kansei word in the fruit juice experiment.

	Ana	Antonio	Belén	David	Eli	Enrique	Eva	Guillem	Héctor	Helena	Jose	Marta	Merce	Mònica	Núria	Pere	Raquel	Vicenç	Xavi
Refreshing	A	B	B	A	A	A	A	B	A	B	B	B	A	A	B	A	B	A	B
Healthy	A	B	B	A	A	A	B	A	A	B	A	B	A	A	B	A	A	B	A
Exotic	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A
Seductive	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Relaxing	A	B	A	B	A	A	A	A	A	B	A	B	B	A	B	A	B	B	A
Natural	A	B	A	A	A	A	B	A	A	B	A	B	A	B	B	A	A	B	A
Tasty	A	B	A	A	A	A	A	B	A	A	B	B	A	A	A	A	A	A	A

I propose the following distance to perform the cluster analysis:

$$distance(i,j) = Number\ of\ KW\ where\ participants\ i\ and\ j\ fall\ in\ different\ groups$$

For example, $distance(Eli, Enrique) = 0$, as columns Ana and Eli in Table 8.11 are exactly the same (all A). $distance(Pere, Raquel) = 2$, as they fall in different groups in Kansei words Refreshing and Relaxing.

Figure 8.11 shows the result of the hierarchic cluster analysis using the above proposed distance.

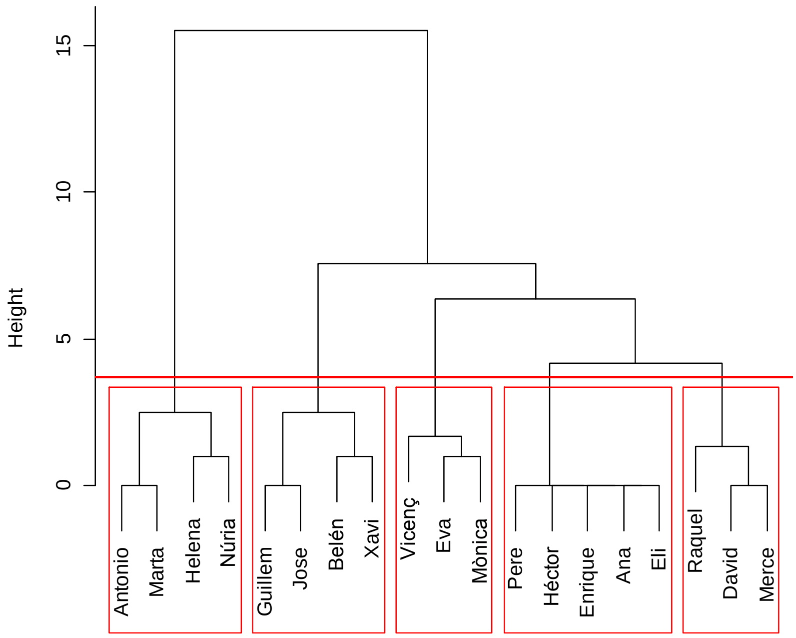


Figure 8.11. Dendogram grouping subjects

In this occasion, the dendogram can be cut depending on the number of groups desired for the posterior qualitative analysis. For example, if we have resources to later interview 5 people, or prepare a focus groups with 5 people, it makes sense to cut the dendogram leaving 5 groups: one person from each group can be later selected. Section 8.5 will give additional clues on which person to select from each group.

Figure 8.12 shows the groups for each participant, with columns ordered according to the global grouping. Notice how the five created groups are quite reasonable, as participants inside each group usually fall in the same group for all Kansei words, or differ in just one.

	Antonio	Marta	Helena	Núria	Guillem	Jose	Belén	Xavi	Vicenç	Eva	Mònica	Pere	Héctor	Enrique	Ana	Eli	Raquel	David	Merce
Refreshing	B	B	B	B	B	B	B	B	A	A	A	A	A	A	A	A	B	A	A
Healthy	B	B	B	B	A	A	B	A	B	B	A	A	A	A	A	A	A	A	A
Exotic	A	A	B	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Seductive	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Relaxing	B	B	B	B	A	A	A	A	B	A	A	A	A	A	A	A	B	B	B
Natural	B	B	B	B	A	A	A	A	B	B	B	A	A	A	A	A	A	A	A
Tasty	B	B	A	A	B	B	A	A	A	A	A	A	A	A	A	A	A	A	A

Figure 8.12. Groups for each participant and Kansei word in the fruit juice experiment (ordered by the final cluster analysis)

8.4. Introducing Repeatability in Kansei Engineering Studies

When collecting data for a KE study, each participant in the study rates each stimulus on every Kansei word. This process is done only once. If one of the participants in the study repeats the study one year after, it could happen that he or she gives different ratings to the same stimulus and Kansei words: one year is a long time, and we could assume that his or her tastes have changed.

However, if the study is repeated just five minutes after finishing the first one, we do not expect big changes in ratings. If the same person rates a product with 7 points on the Kansei word modern, and five minutes later rates the same product with 1 point on the same Kansei word, it is obvious that his or her ratings are worthless. In general, if ratings differ too much[113], we can safely assume that there is a lack of coherence in that person.

In this section, I will propose a methodology for evaluating participants' coherence. The idea is having each participant flagged with a coherence index, so that they can be ordered from more coherent to less coherent. Section 8.4.1 suggests several possibilities for this coherence index. Section 8.4.2 deals with changes in the Kansei Engineering data collection procedure to allow the calculation of these indexes.

8.4.1. Using an index to measure coherence

In order to be able to evaluate the coherence of a participant in the study, one must ask for the same ratings more than once. This idea is common both in industrial statistics (where gage R&R studies are done to assess if different operators measuring a part several times give always the same quantities) and social sciences research

[113] "Too much" is not a precise measure of anything. I will tackle the meaning of "ratings differing too much" later in this section.

(where reliability is interpreted as the correlation among multiple measurements of the same construct). However, I am not aware of KE studies addressing the coherence issue.

The concern with participants' coherence comes not only from a theoretical perspective, but also looking at real data. The data collection in the fruit juice experiment was done twice: that gave the opportunity to repeat the analysis and compare the results from both rounds. I already shared some thoughts that arose from this experiment when it was first presented in Chapter 4. In particular, Figure 4.15 (reproduced in Figure 8.13 for convenience), illustrates the fact that some people (like Ana) are more coherent than others (like Guillem). Ana gives exactly the same ratings in the first and second round in more than 60% of occasions. On the contrary, half of the juices get ratings differing in 2 or more units in Guillem.

	Ana 1st round	Ana 2nd round	Difference
1	6	6	0
2	4	4	0
3	2	3	-1
4	6	6	0
5	5	5	0
6	7	6	1
7	7	6	1
8	6	4	2
9	4	4	0
10	6	5	1
11	6	6	0
12	4	4	0
13	6	6	0
14	5	5	0
15	5	6	-1
16	6	6	0

	Guillem 1st round	Guillem 2nd round	Difference
1	3	5	-2
2	7	5	2
3	3	5	-2
4	5	5	0
5	4	5	-1
6	5	6	-1
7	3	5	-2
8	6	6	0
9	5	6	-1
10	7	4	3
11	5	5	0
12	5	3	2
13	5	4	1
14	6	3	3
15	6	5	1
16	3	6	-3

Figure 8.13. First and second round ratings on the Kansei word Tasty for two participants (Ana and Guillem). (Replicated from Figure 4.15)

Figure 8.14 shows the data structure for computing the coherence index. Each participant l (with $l = 1, \dots, p$) has repeated ratings on every Kansei word k (with $k = 1, \dots, r$). For each Kansei word, some or all of the stimuli get repeated ratings. All the stimuli with repeated ratings are stacked in one matrix, with a total of n rows. The number of ratings for each stimuli with repeated ratings is m. Usually, $m = 2$. However, m could theoretically be any value. We finally have a $n \cdot m$ matrix $\boldsymbol{X}^l$ for each participant l. This matrix is used to compute the coherence index for each participant.

An intraclass correlation coefficient (ICC) can be used as coherence index. The ICC is an ANOVA-type model that considers ratings as the response. A number of ICC estimators have been proposed depending on the underlying model (McGraw, Wong 1996). The simplest one resembles that of a one-way ANOVA, where stimuli are considered as a random effect. This model fits to our data in matrix $\boldsymbol{X}^l$, as the ordering on j (the replica) is irrelevant. This is the ICC(1) according to the nomenclature proposed in McGraw and Wong (1996). Basically, the ICC is a ratio of object variation relative to the total observed variation: $ICC = \sigma_p^2/\sigma_p^2 + \sigma_e^2$. Although the ICC is defined in terms of proportions of variances, it is possible to get negative estimates when the samples are very badly correlated (Nichols 1999).

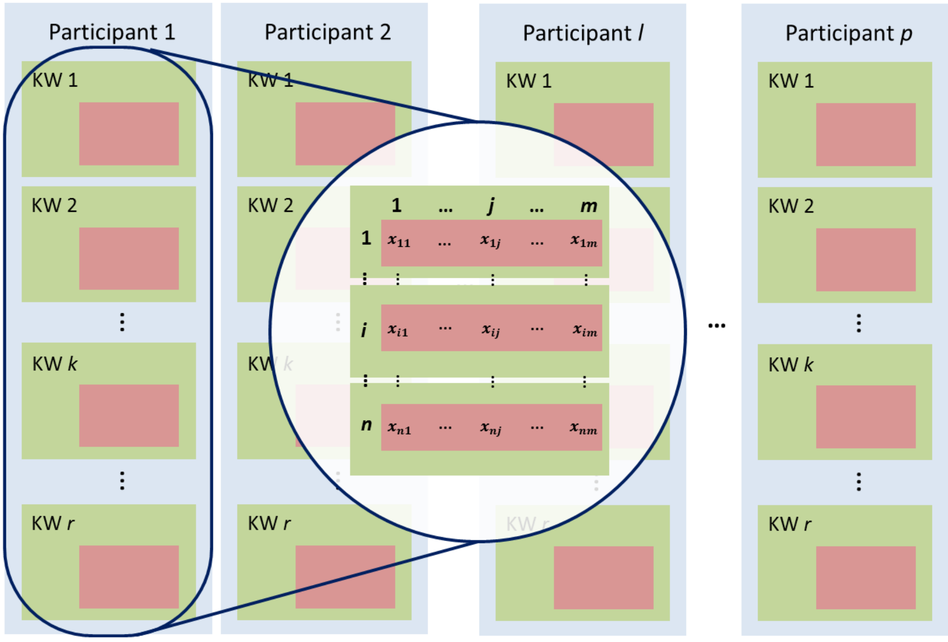


Figure 8.14. Matrix of data for calculating the coherence index (n stimuli, rows; m participants, columns)

The intraclass correlation coefficient assumes that data are continuous. However, it can be safely used in KE studies with 5 or more points scales (like the common 7-point scale). In KE scales, integer anchors are usually used, and the format implies that participants regard the rating levels as evenly-spaced. This is a safeguard to consider data as continuous.[114]

If a scale with very few points (a 3-point scale, for example, or even a binary scale) is used, alternatives exist to the proposed intraclass correlation coefficient. One solution is using a modified version of ICC for ordinal data, like the one proposed in de Mast and van Wieringen (2004). Another option is jumping into the world of non-parametric methods. For example, the Kendall's coefficient of concordance can be used.

---

[114] A good discussion on when ordinal scales can be considered continuous for calculating reliability indexes can be found in the web page http://www.john-uebersax.com/stat/cont.htm (accessed March 2011).

Considering matrix $\boldsymbol{X}$ with n rows and m columns, Kendall's coefficient of concordance (also known as Kendall's W) is computed using the following formula: $W = \frac{12\,S}{m^2(n^3-n)-m\,T}$ with $S = \sum_{i=1}^{n}(R_i - \bar{R})^2$ and $T = \sum_{k=1}^{g}\left(t_k^3 - t_k\right)$. Figure 8.15 shows an example of computation of Kendall's W using Ana's responses to Kansei word Tasty in the fruit juice experiment. The idea is ranking all stimuli in each replica. When there are ties, the average of ranks for those cells is used. R_i is the sum of ranks for each stimulus. T is a correction factor for tied ranks. t_k is the number of tied ranks in each group of ties (there are g groups of ties).

As stated in Legendre (2005), *"Kendall's W is an estimate of the variance of the row sums of ranks* R_i *divided by the maximum possible value the variance can take; this occurs when all variables are in total agreement. Hence 0≤W≤1, 1 representing perfect concordance."*

	C1	C2	Rangs		R_i
1	6	6	11	12,5	23,5
2	4	4	3	3,5	6,5
3	2	3	1	1	2
4	6	6	11	12,5	23,5
5	5	5	6		13
6	7	6	15,5	12,5	28
7	7	6	15,5	12,5	28
8	6	4	11	,5	14,5
9	4	4	3	3,5	6,5
10	6	5	11	7	18
11	6	6	11	125	23,5
12	4	4	3	3,5	6,5
13	6	6	11	12,5	23,5
14	5	5	6	7	13
15	5	6	6	12,5	18,5
16	6	6	11	12,5	23,5

$$\bar{R} = 17 \Rightarrow S = 1050.5$$

$$\left.\begin{aligned} t_1 &= 3 \\ t_2 &= 3 \\ t_3 &= 7 \\ t_4 &= 2 \\ t_5 &= 4 \\ t_6 &= 3 \\ t_7 &= 8 \end{aligned}\right\} \Rightarrow T = 978$$

$$\left.\begin{aligned} S &= 1050.5 \\ T &= 978 \\ m &= 2 \\ n &= 16 \end{aligned}\right\} \Rightarrow$$

$$W = 0.8776$$

Figure 8.15. An example of calculation of Kendall's W using Ana's data from Figure 8.13

The intraclass correlation coefficient (ICC) and Kendall's coefficient of concordance (Kendall's W) have been calculated for each participant in the fruit juice experiment[115]. Data from the two replicas of all 16 juices have been used. The results are shown in Table 8.12. Both ICC and Kendall's W give the same information, as the bars next to the figures help to notice: when ICC is high Kendall's W is also high; when ICC is low, Kendall's W is also low.

[115] The function `icc` and `kendall` of the `R` package `irr` have been used to compute intraclass correlations coefficients and Kendall's coefficient of concordance.

Participants can be ordered from more to less coherent using the figures in Table 8.12. The two participants that motivated the discussion on reliability in KE studies, Ana and Guillem, appear as the most and the least coherent, respectively.

Table 8.12. ICC and Kendall W for participants of the fruit juice experiment.

Participant	ICC	Kendall W
Ana	0,7339	0,8620
Antonio	0,4869	0,7671
Belén	0,5849	0,8616
David	0,6317	0,8323
Eli	0,6727	0,8421
Enrique	0,4218	0,7212
Eva	0,5043	0,7715
Guillem	0,1170	0,5517
Héctor	0,5936	0,8002
Helena	0,2232	0,6986
Jose	0,4398	0,7049
Marta	0,4744	0,7620
Merce	0,5344	0,7757
Mònica	0,2088	0,6789
Núria	0,3570	0,7251
Pere	0,5405	0,7628
Raquel	0,6201	0,8349
Vicenç	0,6337	0,8268
Xavi	0,3300	0,6542

8.4.2. Data collection for measuring coherence

As it has been stated in the previous section, evaluating coherence necessarily requires asking the same more than once. In the fruit juice experiment, this has been strictly done: the same data collection procedure was repeated twice[116]. However, one might soundly think that Kansei Engineering questionnaires are long enough to make them longer asking again for something already been asked. Repeating the whole data collection procedure when it has already been done seems difficult to achieve in practice. My proposal is thus not repeating the whole data collection twice (or even more times), but presenting only some of the stimuli more than once in the same data collection procedure in order to have some replicates.

When not all the stimuli are presented again for being rated, the coherence index (either ICC or Kendall's W) can vary quite a lot, depending on the subset of stimuli

[116] All the analysis done in this dissertation up to this section used data only from the first round of the fruit juice experiment. The whole data collection process was repeated again just 20 minutes after finishing the first round with the purpose of considering reliability in Kansei Engineering studies.

repeated. Using data from the fruit juice experiment, a simulation has been performed creating all combinations of stimuli taken one at a time (16 in total), in pairs (120 in total), in threes (560 in total), in groups of four (1820 in total), five (4368 in total), six (8008 in total), seven (11440 in total), eight (12870 in total), nine (11440 in total), ten (8008 in total), eleven (4368 in total), twelve (1820 in total), thirteen (560 in total), fourteen (120 in total), fifteen (16 in total) and sixteen (only 1). The ICC and Kendall's W have been computed for each of these combinations of stimuli and participant.

Table 8.13 shows some summary statistics (minimum, 5% quantile, median, 95% quantile and maximum) of the ICC and Kendall's W for each participant for all combinations of stimuli (juices) taken in groups of three. The first column shows the "real" figure of the index (calculated using all 16 juices). Variability is rather high. Looking at Ana's figures, for example, the computed ICC ranges from 0.18 (minimum) to 0.97 (maximum) depending on the three juices selected for being rated again.

Table 8.13. Minimum, 5% quantile, median, 95% quantile and maximum ICC and Kendall W for participants of the fruit juice experiment, when only 3 stimuli from the 16 available are repeated.

	ICC						Kendall W					
	Real	Min	5%	Median	95%	Max	Real	Min	5%	Median	95%	Max
Ana	0,73	0,18	0,41	0,72	0,92	0,97	0,86	0,53	0,69	0,86	0,96	0,98
Antonio	0,49	-0,12	0,15	0,52	0,70	0,76	0,77	0,48	0,62	0,78	0,87	0,91
Belén	0,58	-0,31	0,01	0,52	0,85	0,90	0,86	0,45	0,63	0,84	0,94	0,98
David	0,63	0,08	0,31	0,64	0,84	0,90	0,83	0,56	0,67	0,83	0,92	0,96
Eli	0,67	-0,32	0,15	0,66	0,89	0,95	0,84	0,41	0,65	0,84	0,95	0,97
Enrique	0,42	-0,28	0,09	0,42	0,72	0,81	0,72	0,38	0,52	0,72	0,87	0,93
Eva	0,50	-0,43	-0,15	0,52	0,89	0,95	0,77	0,30	0,50	0,78	0,96	0,98
Guillem	0,12	-0,38	-0,22	0,11	0,46	0,67	0,55	0,27	0,38	0,55	0,73	0,83
Héctor	0,59	0,05	0,24	0,57	0,77	0,83	0,80	0,61	0,66	0,79	0,89	0,93
Helena	0,22	-0,43	-0,23	0,17	0,61	0,83	0,70	0,31	0,53	0,68	0,82	0,93
Jose	0,44	-0,10	0,12	0,40	0,62	0,71	0,70	0,45	0,55	0,69	0,82	0,89
Marta	0,47	0,06	0,18	0,47	0,63	0,70	0,76	0,47	0,59	0,75	0,86	0,93
Merce	0,53	-0,49	-0,12	0,50	0,77	0,85	0,78	0,35	0,48	0,78	0,91	0,95
Mònica	0,21	-0,51	-0,15	0,18	0,57	0,74	0,68	0,35	0,50	0,68	0,83	0,90
Núria	0,36	-0,36	-0,11	0,33	0,65	0,73	0,73	0,37	0,52	0,72	0,87	0,93
Pere	0,54	-0,13	0,26	0,55	0,70	0,77	0,76	0,43	0,59	0,76	0,88	0,91
Raquel	0,62	-0,07	0,23	0,65	0,83	0,90	0,83	0,45	0,64	0,86	0,93	0,97
Vicenç	0,63	0,03	0,27	0,64	0,88	0,92	0,83	0,56	0,66	0,85	0,95	0,97
Xavi	0,33	-0,51	-0,19	0,29	0,67	0,82	0,65	0,28	0,42	0,65	0,86	0,95

Obviously, these deviations from the "real" index of coherence decrease when the number of repeated stimuli is increased. Figure 8.16 shows the 5% and 95% quantile of the ICC for Ana and Guillem (the most and less coherent participants) depending on the number of repeated stimuli.

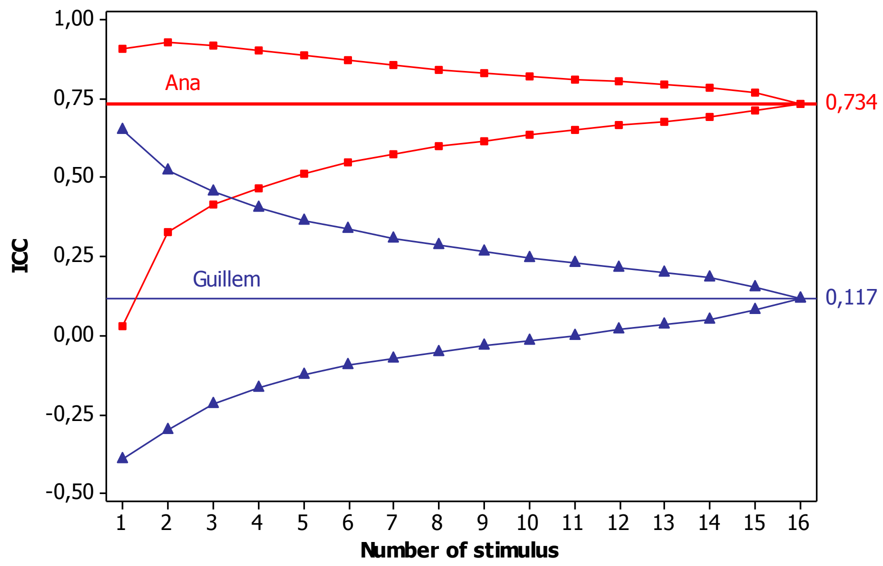


Figure 8.16. 5% and 95% ICC quantiles for Ana and Guillem depending on the number of taken stimuli. The horizontal lines show the "real" ICC value for each participant.

So two questions must be now answered:

- How many stimuli should be repeated to get a decent estimation of the index of coherence (either ICC or Kendall's W)?

  On one hand, it is better to have as many stimuli repeated as possible to obtain a good estimation of the coherence index. But obviously, the more stimuli shown again, the longer the data collection is. My recommendation is: repeat as many stimuli as possible, but at least repeated around 20% of them. For example, if 8 stimuli are used, a minimum of 2 should be repeated (8·0.20 = 1.6 ≅ 2); if 16 stimuli are used, a minimum of 3 should be repeated (16·0.20 = 3.2 ≅ 3), if 20 stimuli are used, a minimum of 4 should be repeated (20·0.20 = 4)

  I think the figure of 20% represents a good compromise between the desire to have a decent estimation and the need to keep the data collection short.

- Which stimuli should be selected?

  One possibility is selecting randomly the stimuli to be repeated. However, there are other choices that produce better results. Based on my experience with KE studies and after deeply analyzing data from the fruit juice experiment, I think the best option is choosing stimuli – say, juices – that are quite different among them, and that are prone to be rated with extreme values.

For example, if one has to repeat 3 juices from the 16 available, a good selection would be the 3 juices shown in the left part of Figure 8.17. The computed ICCs for each participant using only those 3 juices are quite similar to the "real" ICCs using all 16 juices.

Participant	ICC (3 juices)	ICC (all juices)
Ana	0,7257	0,7339
Antonio	0,7417	0,4869
Belén	0,5813	0,5849
David	0,7455	0,6317
Eli	0,7415	0,6727
Enrique	0,5489	0,4218
Eva	0,6092	0,5043
Guillem	0,1482	0,1170
Héctor	0,6751	0,5936
Helena	0,3420	0,2232
Jose	0,5363	0,4398
Marta	0,6217	0,4744
Merce	0,5020	0,5344
Mònica	0,1555	0,2088
Núria	0,4444	0,3570
Pere	0,5487	0,5405
Raquel	0,6383	0,6201
Vicenç	0,7308	0,6337
Xavi	0,3959	0,3300

Figure 8.17. Recommended selection of three juices (on the left) and obtained ICCs for each participant (on the right)

To recap, this is the proposed method to collect data allowing the calculation of a coherence index (summarized in Figure 8.18 for the case of having 8 available stimuli):

- All stimuli are evaluated as usual, and all Kansei words are rated for each one of the stimuli. The ideal procedure is randomizing the order of presentation for each participant. This is easy when stimuli are presented on a computer for each participant and ratings are directly recorded with the computer, but can be more complicated in other situations.
- Some stimuli (at least 20% of the original ones) are presented again for evaluation, as if they were new stimuli (there is no signal that the stimuli to be shown are repeated). The order of presentation of these repeated stimuli can also be randomized for each participant.

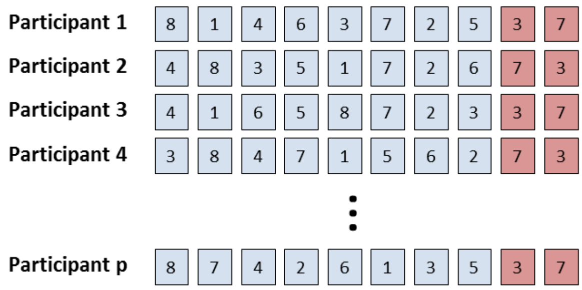


Figure 8.18. An example of data collection repeating two stimuli.

Although, in theory, each repeated stimuli could be shown more than twice, this is very rare. If more stimuli can be shown, it is probably a better option expanding the subset of repeated stimuli, showing each one only twice.

The fact of presenting the repeated stimuli at the end, once all the original stimuli have been presented, tries to prevent participants from detecting that repeated stimuli are shown again. If the order of presentation of stimuli were kept totally randomized (putting together the original stimuli and the repetitions), chances are that the same stimuli is shown for evaluation twice in a row (or with very few other stimuli between). This fact would go against the attempt to keep participants without noticing that some stimuli are repeated. Anyway, according to my experience, I think that participants very often detect that an already rated stimulus is being presented again to evaluation, and they feel upset about this. So it is better upsetting them at the end, when the data collection is almost finished.

8.5. Presenting results from a Kansei Engineering study.

Although a bunch of statistical methods have been presented in this dissertation to analyze data collected for KE studies, some of the "customers" of Kansei Engineering (designers, engineers, technicians, marketing experts, etc.) might have little knowledge of statistics. Therefore, there is a need to present results in a visual way that allows an error-free interpretation. In my opinion, making presentation of results appealing is something appreciated by everybody, also by statisticians who are accustomed to interpret complex outputs.

Section 6.2 made proposals in the direction of having visual tools for correctly interpreting KE results. A relevant one was introducing a new visual graph for easily showing the results from an ordinal logistic regression, including the representation of interactions.

Conclusions can be extracted studying this graph for each Kansei word. It would also be interesting to give visual tools to smooth the interpretation of innovations presented in this chapter: the emotional segmentation of participants and the coherence index. In this section, I propose two graphs to facilitate the incorporation of these two innovations in Kansei Engineering results. First, the emotional overview will be presented in Section 8.5.1. Secondly, the participants' compass will be introduced in Section 8.5.2. Section 8.5.3 explains how part of the extracted information can be used as the input for a qualitative analysis.

8.5.1. The Emotional Overview

Figure 2.5 (in page 19) showed the resemblance between emotional design and the structure of the house of quality commonly used in QFD. Emotional needs (Kansei words) were on the left part, whereas technical properties (factors) were on the top. The central part is a matrix that links (filling the appropriate cells with a sign) the emotional needs with the technical properties. This idea will be now used to create a graph that summarizes the results from the synthesis phase, taking into account the emotional segmentation. This graph is called the emotional overview.

The emotional overview is necessarily simplistic (especially when considering interactions) and does not substitute deeply looking at the results for each group and Kansei word. However, it does give a nice "overview" of the situation.

The emotional overview is built in the following manner:

1. All Kansei words are shown in a column on the left. Similar words (according to the representation of Kansei words in the semantic space, see Section 4.7 for an example) are grouped together. For example, in the fruit juice experiment, Seductive and Exotic, or Healthy and Natural, are shown together.
2. All properties (factors) are shown in columns. Below each factor, all levels for that factor are written.
3. Each Kansei word has several rows, one for each of the groups resulting from the emotional segmentation. Participants are shown in columns on the right of the table, and those belonging to a group have a colored cell in that row.
4. For each group inside each Kansei word, the level of a significant factor that gives a high response is marked with a green dot (•). When two factors interact, the combination of levels that give the highest response is marked with a red square (▪).

As in the house of quality, those factors affecting many Kansei words will have columns with many signs (dots or squares) in it. Those Kansei words with many significant factors will have rows with many signs in it.

Table 8.14. Emotional overview for the fruit juice experiment.

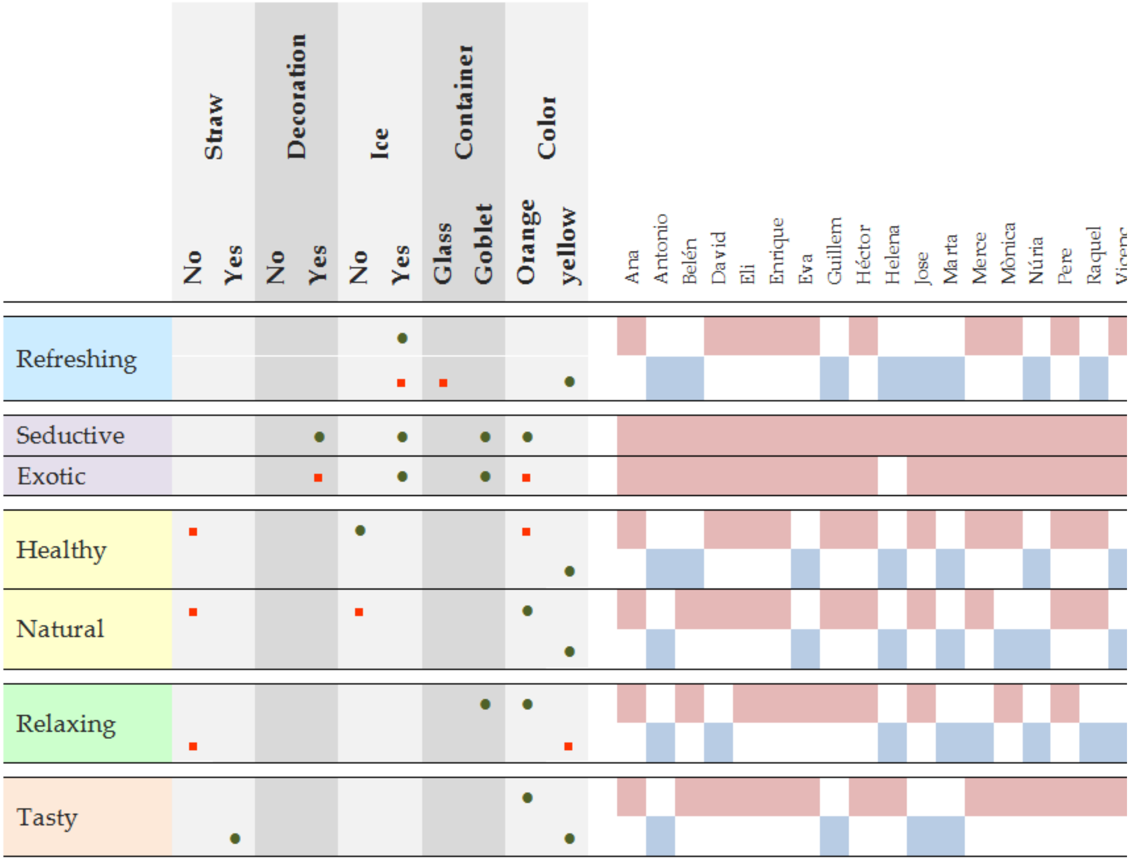


Table 8.14 shows the emotional overview for the fruit juice experiment. From the table it is very easy to see that Color appears as a factor affecting many Kansei words, and in opposite directions (remember that this was not the case before performing the emotional segmentation). Perhaps the best solution to cover all possibilities is offering juices with both colors. On the contrary, Decoration affects only a few Kansei words (Seductive and Exotic; and this Kansei words are affected by many other factors). So Decoration is not an important factor in the study.

8.5.2. The Participants' Compass

The participants' compass makes relevant the individual information that can be extracted from each participant in the study. It is a scatterplot with two axes:

- Horizontal axis: the average of all ratings given by each participant to each Kansei word is computed. Later, the mean of all the averages is calculated for each participant. This axis is labeled emotionality (the higher the number for a participant, the higher ratings he or she tends to give).

- Vertical axis: the variance of all ratings given by each participant to each Kansei word is computed. Later, the mean of all variances (a pooled variance) is calculated for each participant, and the squared root is computed to convert it into a standard deviation. This axis is labeled diversity (the higher the number for a participant, the more diverse ratings he or she tends to give for each Kansei word).

Each participant is represented in the scatterplot using a bubble. The size of the

bubble is proportional to the coherence index for that participant. Finally, each bubble is filled with a color. These colors stratify participants according to the emotional segmentation[117]. Figure 8.19 shows the participants' compass for the fruit juice experiment.

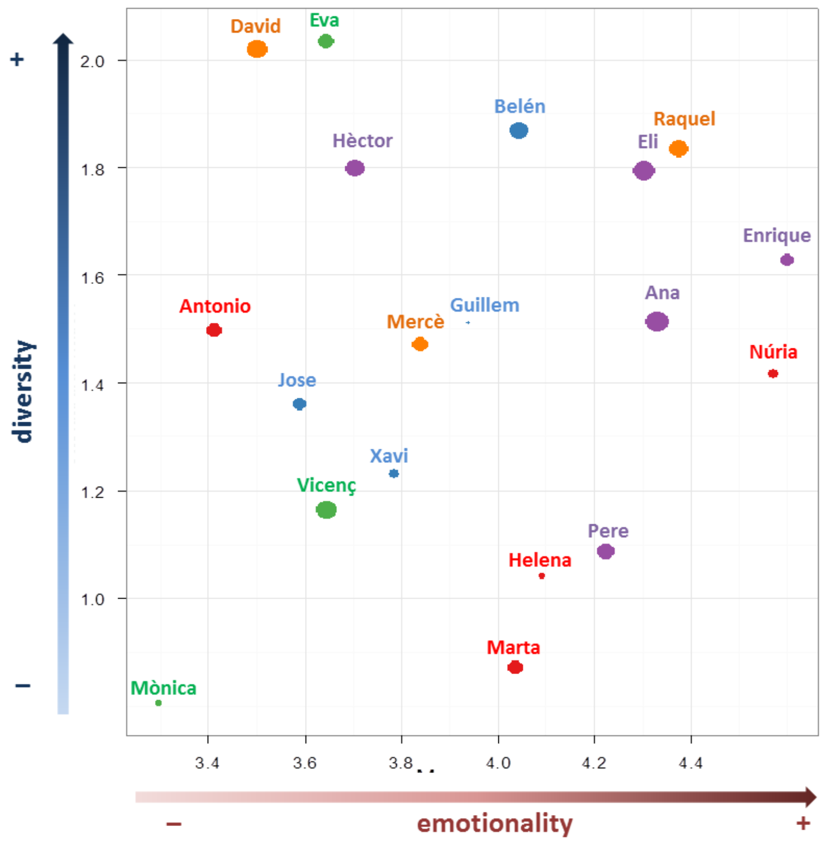


Figure 8.19. Participants' compass for the fruit juice experiment.

[117] The idea of representing four different variables in a scatterplot came to my mind from the work by statistician Hans Rosling. Prof. Rosling adds a fifth variable to the graph, time, making it dynamic. Many examples can be found in http://www.gapminder.org (accessed April 2011). Some software packages – such as JMP from SAS – have incorporated these kind of graphs to their set of visualization tools.

The use of the participants' compass helps in combining quantitative results from a KE study with posterior qualitative approaches. This will be deeper explained in the next section.

8.5.3. Qualitative follow up

I presented a list of research methods used in product design in Section 3.1 of this dissertation. Figure 3.1 placed all these research methods in a scatterplot, where they were roughly divided into belonging to a qualitative or a quantitative approach. Most of them (focus groups, interviews, experience diaries) fell in the qualitative approach. I made clear in Section 3.2 that what I understand by Kansei Engineering – the line used in the whole dissertation – is a quantitative approach based in the Kansei data collection.

A well-done Kansei Engineering study brings lots of information useful for designing – or redesigning – a new product. Although most of the sections in this dissertation have been devoted to statistical methods used in the synthesis phase, one has to keep in mind that a KE study involves going through all the steps presented in the Kansei Engineering model (Chapter 4). Some of these steps (especially spanning the semantic space and the space of properties) involve using lots of qualitative information that should be brought by designers, engineers, users, etc. (for example, when doing affinity diagrams for reducing the initial semantic space). After the Kansei data collection, in the synthesis phase, the spotlight goes on quantitative methods.

My last proposal in this dissertation is using, if necessary, a qualitative technique, but after the synthesis phase of the Kansei Engineering study, and selecting participants based on the information shown by the participants' compass. Perhaps all the information extracted from the KE study seems enough to go on with the design process: great. But perhaps there is a desire to delve deeper. If this is the case, a qualitative method can then be used. Some people could be interviewed (a repertory grid technique could provide rich information[118]), or a group of people participate in

---

[118] The repertory grid is an interviewing technique based on constructivism epistemology. It was devised by psychologist George Kelly in the 1950s. The outcome from applying the repertory grid technique is a matrix whose rows contain constructs (in our case, that would be Kansei words) and whose columns represent elements (in our case, that would be stimuli). Each participant produces his or her own constructs, so the most important ones will appear. The production of constructs is usually done using triads (the participant is exposed to sets of three elements at a time and is asked to explain which property separates one of the elements from the other two). In a second part, the participant is asked to rate each element on his or her own constructed bipolar scales. Details on how to use repertory grids when designing

a focus group. The participants could be a subset of the participants in the KE study: those more relevant for our purposes. And for this purpose the participants' compass can be very helpful.

It is very different selecting people for a focus group almost blindly, that doing it based on their emotional reaction to the product of interest. Imagine we want to conduct a focus group with five participants for the fruit juice experiment. How can we select these participants? Looking at the participants' compass in Figure 8.19, a reasonable possibility would be choosing one participant from each of the five emotional groups. In this way, we have a representation of all range of reactions. The representative of each group could be the most coherent person in that group (biggest bubble), and with most diverse opinions, if possible (at the upper part of the graph). We could combine people with higher ratings (right part of the graph) and lower ratings (left part of the graph). For example, a selection could be David, Belén, Eli, Antonio and Vicenç.

If a computer is used for the data collection step and a skillful person is in charge of analyzing data in the synthesis phase, I think it is possible to have the participants' compass fast. That could allow conducting the qualitative second part of the study with some of the participants the same day the data collection takes place.

---

products can be found in Tomico, Pifarré & Lloveras (2006) and Fallman and Waterworth (2005)

9 Conclusions

This last chapter starts with a brief summary of the dissertation (with references to previous sections in the thesis). The complete model for performing Kansei Engineering studies, incorporating all procedures suggested in this work, is then presented. Finally, suggestions for further research and a discussion on the future of Kansei Engineering close this thesis.

9.1. A Brief Summary

Many "chapters" have passed since Chapter 1 of this thesis explained the motivations for studying Kansei Engineering, and especially the application of statistical methods in Kansei Engineering. It is time to make a brief summary.

Chapter 2, Emotions in Product Design, started with a general description of the role emotions have in our lives and progressively focused in the topic of emotional design. The order in which topics were presented and the extensive use of examples to illustrate ideas constitutes an original way to introduce the subject of this dissertation. Some relevant contributions from this chapter are:

- The explanation on why attractive products sell better (page 11) and work better (page 15).

- The link between emotional design to the four fitnesses described by Shoji Shiba and the hierarchy of consumer needs by Patrick Jordan (page 16). The example of the evolution of Microsoft Windows (page 21) is original.

The chapter was also used to clarify some nomenclature – for example, the use of words emotion, feeling, sensation and perception in this dissertation (page 8). Last section of this chapter described a set of important topics, such as emotion marketing (page 26), brands (page 29), co-creation (page 31) or the increasing role of emotional products (page 32). All examples in this section are original and have never been written before. Last section of the chapter, Section 2.4, also delimitated the scope of the dissertation, as the topics described were later not directly treated in this thesis.

Chapter 3, an Introduction to Kansei Engineering, started with a review of research methods used in product design (page 35). Kansei Engineering was then introduced as a quantitative approach to emotional design (page 38): Kansei Engineering was defined and a list of distinctive properties of KE was shown (page 39), together with a short historical review (page 42). A model for conducting KE studies, taken from Schütte et al. (2004), was then presented (page 44). Two modifications to this model were suggested (the final model, incorporating all proposals from this dissertation, is presented in Section 9.2).

The second part of Chapter 3 was devoted to methods for measuring the Kansei. Several alternatives were progressively presented. Section 3.3 compared capturing the Kansei through physiological body reactions (page 50) and self-reporting the Kansei (page 52). Section 3.4 (page 55) compared different methods of obtaining ratings for products when self-reporting the Kansei. From the vast area of emotional design, the scope of the dissertation was closely bounded by the end of this chapter.

Chapter 4, the Model for Kansei Engineering Studies, described the model proposed in the previous chapter step by step. The procedure was exemplified with the fruit juice experiment. This KE academic experiment was specifically conducted for this dissertation. Some available papers discuss theoretical aspects of KE studies and many others give case studies; this chapter combines discussions on the purpose, methodology and difficulties of each phase with real data analysis. It is specially apropiate as a reading for someone who wants to quickly understand how a Kansei Engineering study is performed (this is the chapter I would have wanted to read when I started studying this topic). As the semantic space is not treated anymore in depth in the dissertation, it was studied in great detail in this chapter (Section 4.3). Section 4.4 was devoted to the space of properties (a topic treated with more detail and taking into account statistical considerations in Section 7.2). The data collection phase (Section 4.5) included an explanation of the procedure and relevant pieces of advice I have been not able to find in other sources. The synthesis phase, although one of the most important ones, was briefly schematized in Section 4.6, as Chapters 6, 7 and 8 were basically devoted to it. The chapter finished with descriptive graphs that can be used to present results from the study. Section 4.8 advanced some of the problems found in data from KE studies that were later addressed in Chapter 8.

Chapter 5 contains an exhaustive literature review on Kansei Engineering. The chapter began framing Kansei Engineering from different perspectives: location of research centers where the topic is studied (page 106), journals that usually have papers on the subject (page 109) and common products where KE is applied (page 110). A list of reasons for doing research in this area was given in Section 5.3. One of the main purposes of this chapter was answering, based on data collected from the literature review, several questions regarding the KE model:

- How Kansei words are selected, and how many are used (page 115).
- How many factors and stimuli are used – this necessarily implies looking at how design matrices are created (page 119).
- How many subjects participate in the data collection, and how they are selected (page 124).
- How stimuli are presented and the data collection session conducted (page 126).
- The kind of scales used to rate the stimuli (page 129).
- Which are the most common tools in the synthesis phase (page 131), and how interactions and variability is treated – these topics were emphasized because of their importance, and were later studied in posterior chapters.
- How results are presented (page 142).

After having discovered what are the most common tools in the synthesis phase from the literature review in the previous chapter, the purpose of Chapter 6 was studying those tools in detail. Two kinds of approaches were considered: a statistical approach and an automatic learning approach. A simple invented example with T-shirts was used through the entire chapter to illustrate each one of the tools. The example was presented in page 149. Two interesting tools that fall beyond the scope of this dissertation – neural networks (page 152) and categorical regression (page 156) – were briefly introduced for the sake of completeness.

Three tools were described under the statistical approach:

- Quantification theory type I (QT1), page 161. The rarely revealed inner calculations of QT1 were explained in page 163. Although partial correlation coefficients are commonly used in QT1, I proposed a modified version using p-values to assess the significance of each factor.

- Ordinal logistic regression (OLR), page 173. Ordinal logistic regression is not commonly used in KE studies, although I find it very appropriate. Besides an explanation of the technique, the most important contribution from this section was a new visual representation of results from OLR that resembles that from QT1 (page 180).
- Mixed effects ordinal logistic regression (mOLR), page 183. The idea of incorporating subjects as random effects in the ordinal logistic regression is new. The visual representation from OLR can also be used in mOLR.

Rough sets were described under the automatic learning approach. The classical rough sets (OrigRS) were introduced with an example devoted to watches (page 189). Besides the description of the technique, an original procedure to add tests of significance to rules from rough sets was proposed (page 195). Variable precision Bayesian rough sets (VPBRS) was introduced in page 197 as a way to apply rough sets taking into account the individual ratings from all subjects.

The thread of Chapter 7 was the use of simulated data. The first part used several simulation sets to compare the performance of the regression tools presented in Chapter 6 (page 208). The main conclusion was that a mixed effects ordinal logistic regression is the best method for analyzing data from KE studies (although a simple ordinal logistic regression may give similar results in many cases).

Regardless of the used tool in the synthesis phase, good conclusions depend on the adequacy of the design matrix – something decided before the data collection. So the second part of this chapter suggested two indicators to assess the suitability of design matrices for KE studies: an orthogonality index (page 214) and a confusion index (page 217). The design matrices used in the simulation sets were evaluated according to these two indexes (page 218). A simple and visual procedure to decide if the design matrix is good or bad was proposed in page 221, as were directions on how to proceed when the design matrix is not suitable.

The third part of Chapter 7 contained a discussion on how to incorporate interactions in KE studies (page 222). Although interpreting interactions in logistic regression is a bit tricky, the visual output for presenting results was adapted to accommodate the presence of an interaction (page 226), facilitating its understanding. The use of rough sets for checking the presence of interactions was also explained (page 228). Data from the fruit juice experiment was revisited in the last part of Chapter 7.

The two proposals for treating data from KE studies made in Chapter 8, Addressing heterogeneity among and within participants, are new:

- The emotional segmentation. The need of segmenting participants from a KE study was justified in Section 8.3.1. A cluster analysis with the raw data was then proposed (page 247) for each Kansei word (a specific distance, based in the polychoric correlation was defined). The process of emotional segmentation was explained in Section 8.3.3, using data from the fruit juice experiment. A global clustering of participants, using results from the emotional segmentation for each Kansei word, was suggested in Section 8.3.4.

  Before conducting the emotional segmentation, the dataset must be free of outliers. A procedure for detecting outliers in KE datasets was explained in Section 8.2. This procedure was proposed in Álvarez (2009), but it is included in this work as it is appropriate for the posterior emotional segmentation.

- The coherence index for participants. Intuition and real data show that some participants in a KE study are more coherent when rating stimuli than others. The coherence index introduced in Section 8.4 serves the purpose of ranking participants according to its coherence. An intraclass correlation coefficient (or alternatively the non-parametric Kendall's W) were suggested in page 262 as coherence index. Of course, computing a coherence index is only possible when the same stimuli are rated more than once by the same people. A proposal was made in Section 8.4.2 to collect data repeating some stimuli. A simulation using data from the fruit juice experiment was conducted and presented in page 266 to discover the effect of making subsets of the stimuli to be repeated. Recommendations were made based in this simulation.

Chapter 8 finished with the proposal of two new graphs designed to summarize the results from a Kansei Engineering study: the emotional overview (page 271) and the participants' compass (page 272). The participants' compass can help in the selection of subjects for a qualitative study (page 274).

The last chapter of this work, Chapter 9, starts with the above summary. The next section presents a complete model for conducting Kansei Engineering studies, incorporating proposals made in the thesis. The chapter finishes with suggestions for further research on the topic (page 283) and some ideas on the future of Kansei Engineering (page 287).

9.2. The Complete Model for Kansei Engineering Studies

A model for conducting Kansei Engineering studies, based in the model suggested by Schütte et al. (2004), was first presented in Section 3.2.4 and followed in detail with the fruit juice experiment, in Chapter 4. The model can be now updated with all the proposals made in this thesis, and summarized with higher levels of detail. Figure 9.1 shows the complete model for conducting KE studies.

After the choice of domain – where the product to be studied and the target group are defined – the semantic space and the space of properties must be spanned.

Spanning the semantic space in a proper way is very important to end up with a list of Kansei words that cover the whole range of possible emotions elicited by the product. Nothing really innovative has been suggested for spanning the semantic space in this dissertation. The common procedure was followed, for example, in Section 4.3: from a long initial collection of Kansei words, several procedures can be used to progressively reduce the list – affinity diagrams, cluster analysis – until the final reduced list of Kansei words is obtained.

The procedure for spanning the space of properties has been enhanced with the idea of checking the adecuacy of design matrices. This is a vital issue, since the appropriateness of the matrix will determine the validity of the final conclusions. When the design matrix is not appropriate, a new design matrix must be choosen (perhaps collapsing levels or even factors). So the whole procedure evolves iteratively, as explained in Section 7.2, until a good set of stimuli is achieved.

The data collection was not in the original model as a separate phase, but I introduced it as a way to emphatise its importance: all conclusions depend on that collected data, so its quality is vital.

The synthesis phase has been one of the main areas of research in this thesis and the object of many improvements. A completely new approach is the emotional segmentation developed in Chapter 8. After detecting and removing outliers, several clustering techniques can be employed to discover groups of people reacting differentely on the same set of stimuli. The statistical analysis conducted in the synthesis phase is then done for each group. A mixed effects ordinal logistic regression is the recommended tool for this analysis, as explained in Section 7.1.

Although a topic not deeply treated in this dissertation (some coments can be found in Section 4.7), and almost never done in KE studies, conducting some confirmatory experiments as a test of validity is undobtely a good idea.

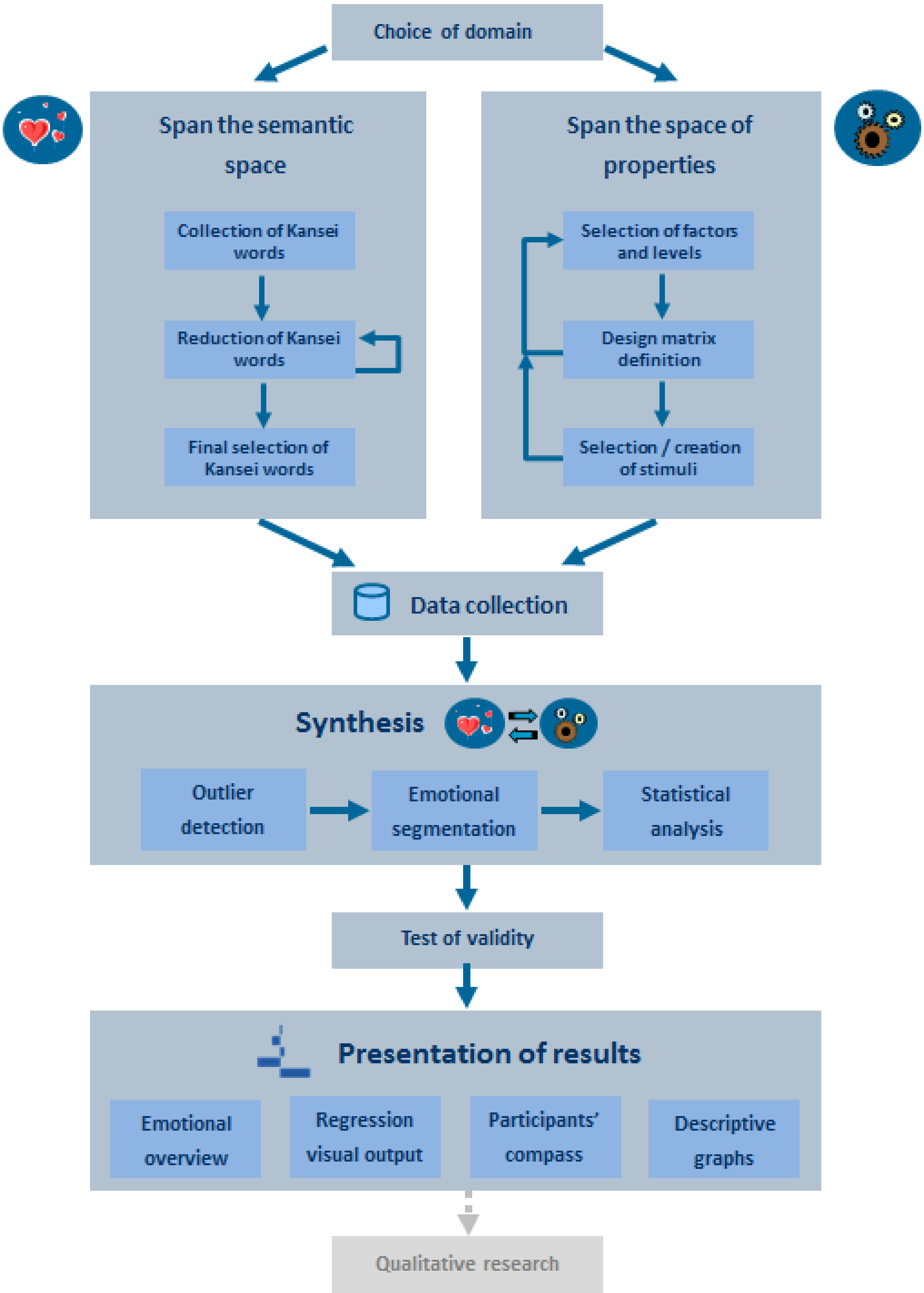


Figure 9.1. The complete model for performing Kansei Engineering studies

The last part of the model, presentation of results, is basically new from ideas developed in this thesis. Designers are one of the most important "consumers" of results from Kansei Engineering studies. As they are not necessarily experts in reading the common numerical output from statistical methods, a visual representation of results that facilitates understanding and interpretation is very convenient. Several graphs – such as the emotional overview and the regression visual output – are devised to accomplish this purpose. The participants' compass can help in the design of an optional qualitative research follow-up.

9.3. Further developments

Several further developments can be now suggested on the issues about Kansei Engineering presented in this thesis. I will divide them in two areas: further research and aids for using Kansei Engineering.

9.3.1. Suggestions for further research

Several methodological improvements for Kansei Engineering studies have been suggested in this dissertation. Some of them allow sophistications that could further expand its usefulness. Other ideas appeared while working in this thesis, but were not treated in the text. The following is a list of them:

- The number of participants needed to achieve reliable results in a Kansei Engineering study stays as an unstudied topic. Of course, the more people the better. But how many are enough? I think a good idea would be dynamically changing the number of required participants depending on the variability of ratings to each Kansei word and stimulus. The list of participants stops when results are "sufficiently" clear. If variability among participants is high, more raters are needed: quite reasonable! It could even be that not all stimuli and Kansei words are rated by the same amount of people. A computer-based method for collecting data could allow the real-time decision of presenting or not specific stimulus or responses to be rated depending on the previous consensus of raters. For example, if 5 people rate fruit juices with ice on the Kansei word Refreshing with ratings from 6 to 7 in a 7-points scale, there is no need to ask more about that: clearly juices with ice are perceived as more refreshing. On the contrary, if opinions diverge a lot, more participants are required to see is that factor has no effect, or simply too few people has been asked to discover its real behaviour.
- Instead of using ratings as response for each Kansei word and stimuli, presenting stimuli in pairs (or in groups of three, four) and asking the participant to select the one he or she finds more modern, comfortable, or whatever Kansei word used. This is the approach followed in discrete choice conjoint analysis. Probably the effort needed in the data collection phase is alleviated using this procedure, as selecting among different stimuli seems

quite natural, and it is probably easier than rating. Using this kind of data collection in KE needs research both from methodology and practice.

- During the whole thesis, the focus has been in discovering how several technical properties (independent variables) elicited some emotions (dependent variables). Why not considering the analysis the other way around (emotions as independent variables and technical properties as dependent variables). For example, if I want to feel happy, which drink should I order? [119]
- The regression analysis in the synthesis phase could be done using a Bayesian approach. The possibility of incorporating *a priori* distributions for the coefficients in the logistic regression could have a twofold benefit. First, the philosophy of incorporating all the previous "emotional" information we have about our product and later updating that information based on the collected data seems very adequate (for example, we are almost sure pink watches will be perceived as femenine, why not having this into account before starting the data collection?). Second, the need to accumulate and organize this previous information would be a good opportunity to involve designers and technicians – those having to provide that information – from the very beginning of the KE study. Furthermore, designers would be needed in every step of the procedure, perhaps decreasing the mistrust they sometimes have in quantitative techniques.
- The regression analysis in the synthesis phase could be done weighting the ratings of participants depending on their coherence index. In this way, results would capture the idea of "believing" more those ratings coming from the most coherent participants. A starting point to this area of research could be the paper "A weighted logistic regression for conjoint analysis and Kansei engineering" (Barone, Lombardo & Tarantino 2007), which suggests weighting a logistic regression (although using a different weighting criterion).

  In a similar way, the ratings of participants could be weighted depending on where they are placed in the diagnostic plot for detecting outliers. Instead of directly removing anomalous values, they are simply given less importance. Of course, this procedure requires a measure of "outlierness", something not yet defined.
- A method for developing robust emotional products could be created. It would be like using Taguchi's ideas, but with emotional responses. It does not seem very hard to accommodate this methodology to Kansei Engineering

[119] This idea has been kinkly suggested by one of the external referees of this thesis.

studies. Perhaps the most difficult part would be deciding what noise factors mean in this context.

- Introducing some kind of sequentiality in Kansei Engineering studies seems a good idea. The recommendation when performing factorial designs in an industrial environment is conducting an initial round of experiments and, in the light of the results, deciding on new experiments: a much sounded strategy. In fact, nothing prevents to follow a similar approach in Kansei Engineering studies, but this is actually never done. Guidelines should be given on how to proceed after getting the results of a first round of experiment in a KE study.

  Perhaps the sequentiality is difficulted by the fact that data collection in a KE study is challenging, and it is not easy to continue asking participants after it has been done once. Computer software able to collect participants' ratings and decide in real-time which stimuli to be shown next to that person seems a promising path. In any case, that needs both methodological and software developments.

9.3.2. Aids for Kansei Engineering Use

Kansei Engineering studies require the use of computers – probably for collecting data, surely for analyzing it. During this dissertation, the open-source statistical software R was used as a tool to implement all of the methodological improvements suggested. It would be nice to organize all written code in several functions, and consolidate these functions in an R package. R packages are a usefull way to keep related functions together, making its mantenaince and – if desired – distribution easier. It would not be very hard doing this, only some functions should be rewritten to improve efficiency and others be slightly changed to adapt them to the usual look and feel of R functions.

The data collection phase in a Kansei Engineering study is time-consuming and demanding (in fact, these characteristics are common in many data collection procedures). In the fruit juice experiment, data was collected on pieces of paper, and later transferred into an electronic format (a tedious task, and prone to errors). I think it would be interesting developing a platform to allow the collection of Kansei Engineering data with a computer. The platform should allow doing an on-line data collection using the internet, but it should also work off-line if preferred. Data from this platform could be formatted in a suitable way so that it could be directly processed by the proper R functions.

The joint use of the platform for collecting data and the R package for analyzing it would speed up the time needed to conduct a Kansei Engineering study. In particular, it would be possible to obtain all results shortly after finishing the data collection.

9.4. A Final Word

Visualizing the future is always difficult and dangerous. But I think it is not too risky stating that the ideas and philosophy behind Kansei Engineering are here to stay. Not only will the emotions elicited by products become more and more important: emotions will play a leading role in the design of user experiences. More and more, companies will incorporate emotional design techniques, so they will achieve an emotional know-how that will progressively reduce the effort for conducting Kansei Engineering studies.

Presentation of stimuli and data collection will be easier thanks to advances in technology. During the first trimester of 2011, Nintendo released Nintendo 3DS, a portable game console able to produce 3D effects without the need for any special glasses[120]. This kind of "virtual reality" devices will allow the presentation of virtual prototypes in a realistic way.

In the last trimester of 2010, Microsoft released Kinect (originally known by the code name Project Natal). Kinect is an extension for the Xbox video game console that allows a controller-free gaming experience. Kinect includes full-body 3D motion capture, facial recognition and voice recognition. The capture of body reactions will be much easier in the near future with devices like this, making the data collection in Kansei Engineering studies more attractive and opening new possibilities.

Whatever the future brings to the field of emotional design, there will be a need to collect data – probably more and more data – and analyze it to extract useful information. Therefore statistics, the discipline which helps in this process of collecting and analyzing data, will have an increasing important role.

Kansei Engineering has a multidisciplinary character; the collaboration of many experts is necessary: designers, psycologists, marketing experts, engineers, statisticians, etc. So more in this discipline than in any other is the sentence from the book "Statistics for Experimenters" – my favorite in the topic of experimental design – true:

---

[120] The ability to produce 3D glasses-free images is accomplished by a method called autostereoscopy (Dodgson 2005). The device also includes eye-tracking.

Close collaboration between the subject matter expert and the statistical practitioner almost always catalyzes the generation of new ideas. (Box, Hunter & Hunter 2005, p.549)

References

Agresti, A. 2002, *Categorical data analysis,* 2nd edn, John Wiley & Sons, Hoboken N.J.

Agresti, A. 2007, *An Introduction to categorical data analysis,* 2nd edn, Wiley, New York.

Akioka, S., Fukumori, H. & Muraoka, Y. 2009, "Search in the mood: The information filter based on ambiguous queries", *International Journal of Computer Applications in Technology,* vol. 34, no. 4, pp. 322-329.

Aktar Demirtas, E., Anagun, A.S. & Koksal, G. 2009, "Determination of optimal product styles by ordinal logistic regression versus conjoint analysis for kitchen faucets", *International Journal of Industrial Ergonomics,* vol. 39, no. 5, pp. 866-875.

Álvarez, E., Hernández, A.C., Talavera, N., Guerrero, R. & Bernard, C. 2001, "Rough analysis applied to ergonomics: a first approximation", *Proceedings of the International Conference on Computer-Aided Ergonomics and Safety*, July 29 - August 1, 2001.

Álvarez, H.R. 2009, *Análisis de valoraciones atípicas en los estudios de ingeniería Kansei: consideraciones estadísticas y prácticas*. Doctoral thesis.

Ariely, D. 2008, *Predictably irrational: the hidden forces that shape our decisions,* Harper, New York, NY.

Axelrod, R. 1997, "Advancing the art of simulation in the social sciences", *Complexity,* vol. 3, no. 2, pp. 16-22.

Ayas, E. & Eklund, J. 2008, "Affective Engineering Approach to Understand Servicescape Effects", *Second Conference on Affective Design and Kansei Engineering.*, eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 20-22 August, 2008.

Ayas, E., Eklund, J. & Ishihara, S. 2008, "Affective design of waiting areas in primary healthcare", *TQM Journal,* vol. 20, no. 4, pp. 389-408.

Baba, K., Shibata, R. & Sibuya, M. 2004, "Partial correlation and conditional correlation as measures of conditional independence", *Australian & New Zealand Journal of Statistics,* vol. 46, no. 4, pp. 657-664.

Bahn, S., Lee, C., Nam, C.S. & Yun, M.H. 2009, "Incorporating Affective Customer Needs for Luxuriousness into Product Design Attributes", *Human Factors and Ergonomics in Manufacturing,* vol. 19, no. 2, pp. 105-127.

Barnes, C. & Lillford, S.P. 2009, "Decision support for the design of affective products", *Journal of Engineering Design,* vol. 20, no. 5, pp. 477.

Barnes, C., Childs, T., Henson, B. & Lillford, S. 2008, "Kansei engineering toolkit for the packaging industry", *TQM Journal,* vol. 20, no. 4, pp. 372-388.

Barnes, C.J. & Lillford, S.P. 2007, "Affective design decision-making - issues and opportunities", *CoDesign,* , pp. 135-46.

Barnes, C.J., Childs, T.H.C., Henson, B. & Southee, C.H. 2004, "Surface finish and touch - a case study in a new human factors tribology", *Wear,* vol. 257, no. 7-8, pp. 740-750.

Barone, S., Lombardo, A. & Tarantino, P. 2007, "A weighted logistic regression for conjoint analysis and Kansei engineering", *Quality and Reliability Engineering International,* vol. 23, no. 6, pp. 689-706.

Ben Ahmed, W. 2009, "A Bayesian learning of probabilistic relations between perceptual attributes and technical characteristics of car dashboards to construct a perceptual evaluation model", *International Journal of Product Development,* vol. 7, no. 1, pp. 47.

Blashfield, R.K. 1976, "Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods.", *Psychological bulletin,* vol. 83, no. 3, pp. 377.

Bouchard, C. 2009, "A European emotional investigation in the field of shoe design", *International Journal of Product Development,* vol. 7, no. 1, pp. 3.

Box, G.E.P., Hunter, J.S. & Hunter, W.G. 2005, *Statistics for experimenters: design, innovation, and discovery,* Wiley-Interscience New York.

Brown, R.L. 1989, "Using Covariance Modeling for Estimating Reliability on Scales with Ordered Polytomous Variables", *Educational and Psychological Measurement,* vol. 49, no. 2, pp. 385-398.

Burchell, B. & Marsh, C. 1992, "The effect of questionnaire length on survey response", *Quality and Quantity,* vol. 26, no. 3, pp. 233-244.

Chang, H., Lai, H. & Chang, Y. 2006, "Expression modes used by consumers in conveying desire for product form: A case study of a car", *International Journal of Industrial Ergonomics,* vol. 36, no. 1, pp. 3-10.

Chang, Y. & Chen, H. 2007, "A neural network-based computer aided design tool for automotive form design", *International Journal of Vehicle Design,* vol. 43, no. 1-4, pp. 136-150.

Chen, C. & Chuang, M. 2008, "Integrating the Kano model into a robust design approach to enhance customer satisfaction with product design", *International Journal of Production Economics,* vol. 114, no. 2, pp. 667-681.

Chen, H. & Chang, Y. 2009, "Extraction of product form features critical to determining consumers' perceptions of product image using a numerical definition-based systematic approach", *International Journal of Industrial Ergonomics,* vol. 39, no. 1, pp. 133-145.

Chen, X., Barnes, C.J., Childs, T.H.C., Henson, B. & Shao, F. 2009, "Materials' tactile testing and characterisation for consumer products' affective packaging design", *Materials & Design*, vol. 30, no. 10, pp. 4299-4310.

Chen-Jimenez, I.E., Kornecki, A. & Zalewski, J. 1998, "Software safety analysis using rough sets", *IEEE Southeastcon'98.Proceedings,* , pp. 15-19.

Childs, T., De Pennington, A., Rait, J., Robbins, T., Jones, K., Workman, C., Warren, S. & Colwill, J. 2003, "Affective Design (Kansei Engineering) in Japan", *A Report from a DTI International Technology Service Mission,* .

Choi, K. & Jun, C. 2007, "A systematic approach to the Kansei factors of tactile sense regarding the surface roughness", *Applied Ergonomics,* vol. 38, no. 1, pp. 53-63.

Chuang, M.C., Chang, C.C. & Hsu, S.H. 2001, "Perceptual factors underlying user preferences toward product form of mobile phones", *International Journal of Industrial Ergonomics,* vol. 27, no. 4, pp. 247-258.

Coenders, G., Satorra, A. & Saris, W.E. 1997, "Alternative approaches to structural modeling of ordinal data: A Monte Carlo study", *Structural Equation Modeling: A Multidisciplinary Journal,* vol. 4, no. 4, pp. 261.

Coleman, S. 2011, *Use of an overall liking question in Kensys project*. [Review of Doctoral Thesis *Statistical Methods in Kansei Engineering Studies*] (Personal communication, 25 July 2011).

Czerniak, J. & Zarzycki, H. 2003, "Application of rough sets in the presumptive diagnosis of urinary system diseases", *Artificial intelligence and security in computing systems: 9th international conference, ACS'2002, Międzyzdroje, Poland, October 23-25, 2002: proceedings*Springer, , pp. 41.

Dahlgaard, J.J., Schutte, S., Ayas, E. & Dahlgaard-Park, S. 2008, "Kansei/affective engineering design: A methodology for profound affection and attractive quality creation", *TQM Journal,* vol. 20, no. 4, pp. 299-311.

Damasio, A.R. 1994, *Descartes' error: emotion, reason and the human brain,* Picador, London.

Damasio, A.R., Everitt, B.J. & Bishop, D. 1996, "The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]", *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences,* vol. 351, no. 1346, pp. 1413-1420.

Daniel, C. 1959, "Use of half-normal plots in interpreting factorial two-level experiments", *Technometrics,* vol. 1, no. 4, pp. 311-341.

Davis, S.M. 1987, "Future perfect", Basic Books, New York.

Dawes, J. 2008, "Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales", *International Journal of Market Research,* vol. 50, no. 1, pp. 61.

de Mast, J. & van Wieringen, W. 2004, "Measurement System Analysis for Bounded Ordinal Data", *Quality and Reliability Engineering International,* vol. 20, no. 5, pp. 383-395.

Delin, J., Sharoff, S., Lillford, S. & Barnes, C. 2007, "Linguistic support for concept selection decisions", *Ai Edam-Artificial Intelligence for Engineering Design Analysis and Manufacturing,* vol. 21, no. 2, pp. 123-135.

Desmet, P. 2002, *Designing Emotions*. Doctoral Thesis.

Dimitriadou, E., Barth, M., Windischberger, C., Hornik, K. & Moser, E. 2004, "A quantitative comparison of functional MRI cluster analysis", *Artificial Intelligence in Medicine,* vol. 31, no. 1, pp. 57-71.

Dodgson, N.A. 2005, "Autostereoscopic 3D displays", *Computer,* vol. 38, no. 8, pp. 31-36.

Draper, N.R. & Smith, H. 1998, *Applied regression analysis,* Wiley-Interscience, New York.

Eklund, J. & Kiviloog, L. 2003, "Kansei ratings and time dependencies", *Proceedings of the XVth Triennial Congress of the International Ergonomics Associatio*n.

Escofier, B. & Pagès, J. 1994, "Multiple factor analysis (AFMULT package)", *Computational Statistics & Data Analysis,* vol. 18, no. 1, pp. 121-140.

Fallman, D. & Waterworth, J. 2005, "Dealing with user experience and affective evaluation in hci design: A repertory grid approach", *Conference on Human Factors in Computing Systems, CHI2005*, April 2-7, pp. 2.

Filzmoser, P. & Todorov, V. 2009, "An Object-Oriented Framework for Robust Multivariate Analysis", *Journal of Statistical Software,* vol. 32, no. i03.

Forlizzi, J. 2007, "How robotic products become social products: an ethnographic study of cleaning in the home", *Proceedings of the ACM/IEEE international conference on Human-robot interaction*ACM New York, NY, USA, , pp. 129.

Fox, J. 1997, *Applied Regression Analysis, Linear Models, and Related Methods,* Sage Publications, Inc.

Fox, J. 2008, *Applied regression analysis and generalized linear models,* Sage, California.

Friborg, O., Martinussen, M. & Rosenvinge, J.H. 2006, "Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience", *Personality and Individual Differences,* vol. 40, no. 5, pp. 873-884.

Fukushima, K., Kawata, H., Fujiwara, Y. & Genno, H. 1995, "Human sensory perception oriented image processing in a color copy system", *International Journal of Industrial Ergonomics,* vol. 15, no. 1, pp. 63-74.

Genjo, K., Tanabe, S., Matsumoto, S., Hasegawa, K. & Yoshino, H. 2005, "Relationship between possession of electric appliances and electricity for lighting and others in Japanese households", *Energy and Buildings,* vol. 37, no. 3, pp. 259-272.

Giles, J. 2005, "Internet encyclopaedias go head to head", *Nature,* vol. 438, no. 7070, pp. 900-901.

Gonzalez, M.E., Mueller, R.D. & Mack, R.W. 2007, "An alternative approach in service quality: an e-banking case study", *Quality Management Journal,* vol. 15, no. 1, pp. 41-58.

Grant, S., Aitchison, T., Henderson, E., Christie, J., Zare, S., McMurray, J. & Dargie, H. 1999, "A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg Scales, and Likert Scales in Normal Subjects During Submaximal Exercise", *Chest,* vol. 116, no. 5, pp. 1208.

Grzymala-Busse, J.W. & Ziarko, W. 2000, "Data mining and rough set theory", *Communications of the ACM,* vol. 43, no. 4, pp. 109.

Gummesson, E. 2007, "Quality in the Value-Creating Network Society", *10th QMOD Conference. Quality Management and Organiqatinal Development. Our Dreams of Excellence.* Linköping University Electronic Press, , 18-20 June 2007.

Hair Jr, J.F., Anderson, R.E., Tatham, R.L. & Black, W.C. 1995, *Multivariate data analysis: with readings,* Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Hassenzahl, M. 2004, "The Interplay of Beauty, Goodness, and Usability in Interactive Products", *Human-Computer Interaction,* vol. 19, no. 4, pp. 319.

Hayashi, C. 1952, "On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view", *Annals of the Institute of Statistical Mathematics,* vol. 3, no. 2, pp. 69.

Hiltz, S.R. & Johnson, K. 1990, "User satisfaction with computer-mediated communication systems", *Management Science,* , pp. 739-764.

Holgado–Tello, F., Chacón–Moscoso, S., Barbero–García, I. & Vila–Abad, E. 2010, "Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables", *Quality & Quantity,* vol. 44, no. 1, pp. 153-166.

Hosmer, D.W. & Lemeshow, S. 2000, *Applied logistic regression,* 2nd edn, John Wiley & Sons, New York.

Hsu, S.H., Chuang, M.C. & Chang, C.C. 2000, "A semantic differential study of designers' and users' product form perception", *International Journal of Industrial Ergonomics,* vol. 25, no. 4, pp. 375-391.

Huang, X., Sobue, S., Kanda, T. & Chen, Y. 2009, "Mapping functions between image color features and KANSEI", *Information-an International Interdisciplinary Journal,* vol. 12, no. 1, pp. 137-146.

Hubert, M. & Rousseeuw, P.J. 2005, "ROBPCA: a new approach to robust principal component analysis", *Technometrics,* , pp. 64-79.

Isen, A.M., Ashby, F.G. & Turken, U. 1999, "A neuropsychological theory of positive affect and its influence on cognition", *Psychological Review,* vol. 106, no. 3, pp. 529-550.

Ishihara, K., Nakagawa, R., Ishihara, S. & Nagamachi, M. 2008, "An e-commerce site for gift flower arrangements that fit kansei and social manners", *TQM Journal,* vol. 20, no. 4, pp. 338-347.

Ishihara, S., Ishihara, K., Nagamachi, M. & Matsubara, Y. 1995, "An automatic builder for a Kansei Engineering expert system using self-organizing neural networks", *International Journal of Industrial Ergonomics,* vol. 15, no. 1, pp. 13-24.

Ishihara, S., Ishihara, K., Nagamachi, M. & Matsubara, Y. 1997, "An analysis of Kansei structure on shoes using self-organizing neural networks", *International Journal of Industrial Ergonomics,* vol. 19, no. 2, pp. 93-104.

Jaccard, J. 2001, *Interaction effects in logistic regression,* Sage, Thousand Oaks etc.

Jang, D. 2002, "Measures for Evaluating Non-orthogonality of Experimental Designs", *Communications in Statistics - Theory and Methods,* vol. 31, no. 2, pp. 249.

Jeong, K. & Lim, H. 2009, "Factor analysis on hazards for safety assessment in decommissioning workplace of nuclear facilities using a semantic differential method", *Annals of Nuclear Energy,* vol. 36, no. 10, pp. 1639-1647.

Jiao, J.X., Zhang, Y.Y. & Helander, M. 2006, "A Kansei mining system for affective design", *Expert Systems with Applications,* vol. 30, no. 4, pp. 658-673.

Jindo, T. & Hirasago, K. 1997, "Application studies to car interior of Kansei engineering", *International Journal of Industrial Ergonomics,* vol. 19, pp. 105-114(10).

Jordan, P.W. 2002, *Designing pleasurable products: An introduction to the new human factors,* CRC.

Nakada, K. 1997, "Kansei engineering research on the design of construction machinery", *International Journal of Industrial Ergonomics,* vol. 19, pp. 129-146(18).

Kano, N., Seraku, N., Takahashi, F. & Tsuji, S. 1984, "Attractive quality and must-be quality", *The Journal of the Japanese Society for Quality Control,* vol. 14, no. 2, pp. 39-48.

Kleinbaum, D.G. & Klein, M. 2002, *Logistic regression: a self-learning text,* 2nd edn, Springer, New York.

Komine, K., Sawahata, Y., Uratani, N., Yoshidayoshida, Y. & Inoue, T. 2007, "Evaluation of a Prototype Remote Control for Digital Broadcasting Receivers by using Semantic Differential Method", *IEEE Transactions on Consumer Electronics,* vol. 53, no. 2, pp. 561-568.

Kuang, J. & Jiang, P. 2009, "Product platform design for a product family based on Kansei engineering", *Journal of Engineering Design,* vol. 20, no. 6, pp. 589.

Kurosu, M. & Kashimura, K. 1995, "Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability", *Conference on Human Factors in Computing Systems*ACM New York, NY, USA, , pp. 292.

Lai, H., Lin, Y., Yeh, C. & Wei, C. 2006, "User-oriented design for the optimal combination on product design", *International Journal of Production Economics,* vol. 100, no. 2, pp. 253-267.

Lai, H.H., Chang, Y.M. & Chang, H.C. 2005, "A robust design approach for enhancing the feeling quality of a product: a car profile case study", *International Journal of Industrial Ergonomics,* vol. 35, no. 5, pp. 445-460.

Lai, H.H., Lin, Y.C. & Yeh, C.H. 2005, "Form design of product image using grey relational analysis and neural network models", *Computers & Operations Research,* vol. 32, no. 10, pp. 2689-2711.

Lanzotti, A. & Tarantino, P. 2008, "Kansei engineering approach for total quality design and continuous innovation", *TQM Journal,* vol. 20, no. 4, pp. 324-337.

Lanzotti, A., Tarantino, P. & Matrone, G. 2008, "An empirical approach to optimal experimental design selection and data analysis for the synthesis phase of Kansei Engineering", *Second Conference on Affective Design and Kansei Engineering.,* eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 20-22 August, 2008.

Lawless, H.T., Popper, R. & Kroll, B.J. 2010, "A comparison of the labeled magnitude (LAM) scale, an 11-point category scale and the traditional 9-point hedonic scale", *Food Quality and Preference*, vol. 21, no. 1, pp. 4-12.

Legendre, P. 2005, "Species associations: the Kendall coefficient of concordance revisited", *Journal of agricultural, biological, and environmental statistics,* vol. 10, no. 2, pp. 226-245.

Leon, N. 2009, "The future of computer-aided innovation", *Computers in Industry,* vol. 60, no. 8, pp. 539-550.

Li, G. & Chen, Z. 1985, "Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo", *Journal of the American Statistical Association,* vol. 80, no. 391, pp. pp. 759-766.

Lin, Y., Lai, H. & Yeh, C. 2007, "Consumer-oriented product form design based on fuzzy logic: A case study of mobile phones", *International Journal of Industrial Ergonomics,* vol. 37, no. 6, pp. 531-543.

Llinares, C. & Page, A. 2007, "Application of product differential semantics to quantify purchaser perceptions in housing assessment", *Building and Environment,* vol. 42, no. 7, pp. 2488-2497.

Llinares, C. & Page, A.F. 2008, "Differential semantics as a Kansei Engineering tool for analysing the emotional impressions which determine the choice of neighbourhood: The case of Valencia, Spain", *Landscape and Urban Planning,* vol. 87, no. 4, pp. 247-257.

Lokman, A.M., Noor, N.L.M. & Nagamachi, M. 2009, "ExpertKanseiWeb: A tool to design kansei website", *Lecture Notes in Business Information Processing,* vol. 24, pp. 894-905.

Ma, C., Fang, K. & Liski, E. 2000, "A new approach in constructing orthogonal and nearly orthogonal arrays", *Metrika,* vol. 50, no. 3, pp. 255-268.

Ma, T. & Nakamori, Y. 2007, "Agent-based simulation for Kansei Engineering: Testing a fuzzy linear quantification method in an artificial world", *Journal of Systems Science and Systems Engineering,* vol. 16, no. 3, pp. 308-322.

Manly, B.F.J. 2005, *Multivariate statistical methods: a primer,* Chapman & Hall/CRC.

Marco-Almagro, L., Tort-Martorell Llabrés, X. & García Subirats, I. 2007, "Methods for Linking the Semantic Space and the Space of Properties in Kansei Engineering: a Comparison", *First Conference on Affective Design and Kansei Engineering.*, eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 18-20 June, 2007.

Maslow, A.H. 1943, "A theory of human motivation", *Psychological review,* vol. 50, no. 4, pp. 370-396.

Matsubara, Y. & Nagamachi, M. 1997, "Hybrid Kansei Engineering System and design support", *International Journal of Industrial Ergonomics,* vol. 19, no. 2, pp. 81-92.

Matsumura, T. 2004, "Analysis of ovipositional environment using Quantification Theory Type I: the case of the butterfly, Luehdorfia puziloi inexpecta (Papilionidae)", *Journal of Insect Conservation,* vol. 8, no. 1, pp. 59-67.

McCullagh, P. & Nelder, J.A. 1989, *Generalized linear models,* 2nd edn, Chapman & Hall, London.

McCullagh, P. 1980, "Regression Models for Ordinal Data", *Journal of the Royal Statistical Society.Series B (Methodological),* vol. 42, no. 2, pp. 109-142.

McCulloch, W.S. & Pitts, W. 1943, "A logical calculus of the ideas immanent in nervous activity", *Bulletin of mathematical biology,* vol. 5, no. 4, pp. 115-133.

McGraw, K.O. & Wong, S.P. 1996, "Forming Inferences About Some Intraclass Correlation Coefficients", *Psychological methods,* vol. 1, no. 1, pp. 30-46.

Meho, L.I. & Yang, K. 2007, "Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar", *Journal of the American Society for Information Science and Technology,* vol. 58, no. 13, pp. 2105-2125.

Meulman, J. 1998, "Optimal scaling methods for multivariate categorical data analysis", *SPSS white paper: http://www.spss.com,* .

Meulman, J.J. & Heiser, W.J. 1989, "IBM SPSS Categories", manual.

Millidge, J. 2006, *El libro de los zumos y batidos,* Ediciones Robinbook, Barcelona.

Milligan, G.W. 1980, "An examination of the effect of six types of error perturbation on fifteen clustering algorithms", *Psychometrika,* vol. 45, no. 3, pp. 325-342.

Mondragon, S., Company, P. & Vergara, M. 2005, "Semantic differential applied to the evaluation of machine tool design", *International Journal of Industrial Ergonomics,* vol. 35, no. 11, pp. 1021-1029.

Montague, R. 2006, *Why Choose This Book?: How We Make Decisions,* Dutton Adult.

Morey, L.C., Blashfield, R.K. & Skinner, H.A. 1983, "A Comparison of Cluster Analysis Techniques Withing a Sequential Validation Framework", *Multivariate Behavioral Research,* vol. 18, no. 3, pp. 309.

Murphy, S. 2005, "Batidos: Jugosos y caprichosos", Kliczkowski.

Nagamachi, M. 1995, "Kansei Engineering - a New Ergonomic Consumer-Oriented Technology for Product Development", *International Journal of Industrial Ergonomics,* vol. 15, no. 1, pp. 3-11.

Nagamachi, M. 2002, "Kansei engineering as a powerful consumer-oriented technology for product development", *Applied Ergonomics,* vol. 33, no. 3, pp. 289-294.

Nagamachi, M. 2002a, "Kansei Engineering in Consumer Product Design", *Ergonomics in Design,* vol. 10, pp. 5-9(5).

Nagamachi, M. 2008, "Perspectives and the new trend of Kansei/affective engineering", *The TQM Journal,* vol. 20, no. 4, pp. 290-298.

Nagamachi, M., Nishino, T., Okuo, H., Ito, K. & Itami, T. 2007, "An application of kansei engineering and rough set model to designing a comprehensive ball pen", *First Conference on Affective Design and Kansei Engineering.*, eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 18-20 June, 2007.

Naka, T., Ishida, K., Fujii, H. & Takagi, A. 2007, "Evaluation indicators on water-use sustainability in irrigated paddy farming", *Japan International Research Center for Agricultural Sciences,* vol. 41, no. 3, pp. 233-233-245.

Nichols, D.P. 1999, "My coefficient alpha is negative", *SPSS Keywords,* vol. 68, pp. 1-4.

Nishino, T., Nagamachi, M. & Tanaka, H. 2005, "Variable Precision Bayesian Rough Set Model and Its Application to Human Evaluation Data", *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing,* , pp. 294-303.

Nishino, T., Nagamachi, M. & Tanaka, H. 2006, *Variable Precision Bayesian Rough Set Model and Its Application to Kansei Engineering*.

Nishino, T., Nagamachi, M., Okuo, H., Ito, K. & Itami, T. 2007, "Rough set model and its application to kansei engineering", *First Conference on Affective Design and Kansei Engineering.*, eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 18-20 June, 2007.

Nishino, T., Satsuta, R., Uematsu, M., Sugihara, S. & Nagamachi, M. 2008, "Identification of Customers' Latent Kansei Needs and Product Design By Rough Set Based Approach", *Second Conference on Affective Design and Kansei Engineering.*, eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 20-22 August, 2008.

Nishisato, S. 2006, *Multidimensional Nonlinear Descriptive Analysis,* Chapman & Hall/CRC.

Norman, D.A. 1998, *The Design of everyday things,* The MIT Press, London.

Norman, D.A. 2004, *Emotional Design: Why we love (or hate) everyday things,* Basic Books, New York.

Norman, D.A. 2007, *The design of future things*, Basic Books, New York.

Oba, M., Onoda, S. & Komoda, N. 2000, "Evaluating the quantitative effects of workflow systems based on real cases", *33rd Hawaii International Conference on System Sciences-Volume 6*, pp. 6031.

Ogawa, T., Nagano, M. & Yukawa, T. 2009, "Study on comfortable massage method of massage chair considering professional masseur's technique", *International Journal of Biomedical Engineering and Technology,* vol. 2, no. 3, pp. 217-233.

Okada, R. & Castillo, W. 2007, "Cross cultural web comfortability and the study on the influence of cultural profiles in Chile", *Sccc 2007: Xxvi International Conference of the Chilean Society of Computer Science, Proceedings,* , pp. 142-149.

Olsson, U. 1979, "Maximum likelihood estimation of the polychoric correlation coefficient", *Psychometrika,* vol. 44, no. 4, pp. 443-460.

Osgood, C.E., Suci, G.J. & Tannenbaum, P.H. 1957, *The measurement of meaning,* Univ of Illinois Pr.

Pawlak, Z. 1982, "Rough sets", *International Journal of Computer & Information Sciences,* vol. 11, no. 5, pp. 341-356.

Pawlak, Z. 1999, "Decision rules, Bayes' rule and rough sets", *Lecture notes in computer science,* , pp. 1-9.

Pearce, K.F. & Coleman, S.Y. 2008, "Modern-day perception of historic footwear and its links to preference", *Journal of Applied Statistics,* vol. 35, no. 2, pp. 161-178.

Pedrycz, W. & Vukovich, G. 2004, "Noninvasive data analysis: a web of information granules", *Soft Computing-A Fusion of Foundations, Methodologies and Applications,* vol. 8, no. 10, pp. 657-662.

Peralta Coll, E. & Gómez Muñoz, A. 2007, *Aplicació de la metodologia Kansei en la valoració de sucs de fruit*a, degree dissertation, Faculty of Mathematics and Statistics (FME), Technical University of Catalunya (UPC).

Petiot, J.F. & Yannou, B. 2004, "Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics", *International Journal of Industrial Ergonomics,* vol. 33, no. 6, pp. 507-525.

Picard, R.W. 2000, *Affective computing,* The MIT Press.

Poels, K. & Dewitte, S. 2006, "How to Capture the Heart? Reviewing 20 Years of Emotion Measurement in Advertising.", *Journal of Advertising Research,* vol. 46, no. 1, pp. 18.

Predki, B., Słowiński, R., Stefanowski, J., Susmaga, R. & Wilk, S. 2009, *ROSE - Software Implementation of the Rough Set Theory*.

R Development Core Team 2010. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Rapp, S., Jessen, M. & Dogil, G. 1994, "Using Rough Sets Theory to Predict German Word Stress", *Lecture Notes in Computer Science,* , pp. 398-398.

Reichheld, F.F. 2006, *The ultimate question: Driving good profits and true growth,* Harvard Business School Pr.

Ripley, B.D. 2007, *Pattern recognition and neural networks,* Cambridge University Press, Cambridge; New York.

Rohrer, C. 2009, "User Experience Research Methods in 3D: What to Use When and How to Know You're Right", *Keynote presentation at BayCHI – The San Francisco chapter of ACM SIGCHI*.

Rousseeuw, P.J. 1984, "Least median of squares regression", *Journal of the American statistical association,* vol. 79, no. 388, pp. 871-880.

Roy, R., Goatman, M. & Khangura, K. 2009, "User-centric design and Kansei Engineering", *CIRP Journal of Manufacturing Science and Technology,* vol. 1, no. 3, pp. 172-178.

Sakata, T., Watanuki, S., Sakamoto, H., Sumi, T., Kim, Y., Takakura, J., Maehara, K. & Taniguchi, K. 2007, "Objective evaluation of Kansei by a complementary use of physiological indexes, brain wave and facial expressions for user oriented designs", *First Conference on Affective Design and Kansei Engineering.*, eds. S.M. Dahlgaard-Park & J.J. Dahlgaard, Linköping University Electronic Press, , 18-20 June, 2007.

Schifferstein, H.N.J. & Hekkert, P. 2007, *Product experience,* Elsevier Science Ltd.

Schütte, S. & Eklund, J. 2003, *Product Development for Heart and Soul,* Linköping University, Department for Quality and Human Systems Engineering, Linköping.

Schütte, S. & Eklund, J. 2005, "Design of rocker switches for work-vehicles - An application of Kansei Engineering", *Applied Ergonomics,* vol. 36, no. 5, pp. 557-567.

Schütte, S. 2005, *Engineering Emotional Values in Product Design: Kansei Engineering in Development*, Linköping University.

Schütte, S. 2006, "Developing the Space of Product Properties Supporting Kansei Engineering Procedure", *Kansei Engineering International,* vol. 5, no. 4, pp. 11-19.

Schütte, S., Eklund, J., Axelsson, J. & Nagamachi, M. 2004, "Concepts, methods and tools in Kansei engineering", *Theoretical Issues in Ergonomics Science,* vol. 5, no. 3, pp. 214.

Shi, F., Sun, S. & Xu, J. 2009, "Bayesian inference of Kansei knowledge by using variable precision rough sets", *Journal of Computational Information Systems,* vol. 5, no. 1, pp. 35-40.

Shiba, S., Walden, D. & Graham, A. 2001, *Four practical revolutions in management: systems for creating unique organizational capability,* Productivity Press.

Shimizu, Y., Sadoyama, T., Kamijo, M., Hosoya, S., Hashimoto, M., Otani, T., Yokoi, K., Horiba, Y., Takatera, M., Honywood, M. & Inui, S. 2004, "On-demand production system of apparel on the basis of Kansei engineering", *International Journal of Clothing Science and Technology,* vol. 16, no. 1, pp. 32-42.

Slezak, D. & Ziarko, W. 2003, "Variable precision bayesian rough set model", *Lecture notes in computer science,* vol. 2639/2003, pp. 312-315.

Ślezak, D. & Ziarko, W. 2005, "The investigation of the Bayesian rough set model", *International Journal of Approximate Reasoning,* vol. 40, no. 1-2, pp. 81-91.

Such Pérez, M.J. 2003, *La Ingeniería Kansei como modelo de simulación del fenómeno de la percepción. Aplicación en el sector del mobiliario de oficina*., Universidad Politécnica de Valencia.

Sweeney, R.E. 1972, "A Transformation for Simplifying the Interpretation of Coefficients of Binary Variables in Regression Analysis", *The American Statistician,* vol. 26, no. 5, pp. 30.

Jindo, T. & Hirasago, K. 1997, "Application studies to car interior of Kansei engineering", *International Journal of Industrial Ergonomics,* vol. 19, pp. 105-114(10).

Tanaka, Y. 1979, "Review of the Methods of Quantification", *Environmental health perspectives,* vol. 32, pp. 113.

Tanoue, C., Ishizaka, K. & Nagamachi, M. 1997, "Kansei Engineering: A study on perception of vehicle interior image", *International Journal of Industrial Ergonomics,* vol. 19, pp. 115-128(14).

Tay, F.E.H. & Shen, L. 2002, "Economic and financial prediction using rough sets model", *European Journal of Operational Research,* vol. 141, no. 3, pp. 641-659.

Tomico, O., Pifarré, M. & Lloveras, J. 2006, "Experience Landscapes: A Subjective Approach to Explore User-Product Interaction", *Proceedings of the 9th International Design Conference DESIGN 2006*, pp. 393.

Tractinsky, N. 1997, "Aesthetics and apparent usability: empirically assessing cultural and methodological issues", *Proceedings of the SIGCHI conference on Human factors in computing systems*ACM New York, NY, USA, , pp. 115.

Tractinsky, N., Katz, A.S. & Ikar, D. 2000, "What is beautiful is usable", *Interacting with Computers,* vol. 13, no. 2, pp. 127-145.

Tsumoto, S. & Ziarko, W. 1996, "The application of rough sets-based data mining technique to differential diagnosis of meningoenchepahlitis", *LECTURE NOTES IN COMPUTER SCIENCE,* , pp. 438-447.

van Lottum, C., Pearce, K. & Coleman, S. 2006, "Features of Kansei Engineering Characterizing its Use in Two Studies: Men's Everyday Footwear and Historic Footwear", *Quality and Reliability Engineering International,* vol. 22, no. 6, pp. 629-650.

Velicer, W.F. & Jackson, D.N. 1990, "Component Analysis versus Common Factor Analysis: Some issues in Selecting an Appropriate Procedure", *Multivariate Behavioral Research,* vol. 25, no. 1, pp. 1.

Venables, W.N. & Ripley, B.D. 2002, *Modern applied statistics with S,* 4th edn, Springer-Verlag, New York.

Wang, J.C. & Wu, C.F.J. 1992, "Nearly orthogonal arrays with mixed levels and small runs", *Technometrics,* vol. 34, no. 4, pp. 409-422.

Ward, J.H.,Jr. 1963, "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association,* vol. 58, no. 301, pp. pp. 236-244.

Wilson, A. 2003, *Zumos y batidos,* H Kliczkowski-Onlybook, Madrid.

Wu, C. & Hamada, M. 2000, *Experiments: planning, analysis and parameter design optimization,* John Wiley and sons, New York etc.

Yang, S.M., Nagamachi, M. & Lee, S.Y. 1999, "Rule-based inference model for the Kansei Engineering system", *International Journal of Industrial Ergonomics,* vol. 24, no. 5, pp. 459-471.

Ying, W. & Yan, C. 2006, "The Kansei research on the style of women's overcoats", *The Proceedings of the IMACS Multiconference on Computational Engineering in Systems Applications*IEEE, Beijing, China, 4-6 Oct. 2006, pp. 6.

Zhai, L., Khoo, L. & Zhong, Z. 2009a, "A dominance-based rough set approach to Kansei Engineering in product development", *Expert Systems with Applications,* vol. 36, no. 1, pp. 393-402.

Zhai, L., Khoo, L. & Zhong, Z. 2009b, "A rough set based decision support approach to improving consumer affective satisfaction in product design", *International Journal of Industrial Ergonomics,* vol. 39, no. 2, pp. 295-302.

Ziarko, W. 1993, "Variable precision rough set model.", *Journal of computer and system sciences,* vol. 46, no. 1, pp. 39-59.

Origin of pictures

I have made every effort to give proper attribution to all pictures not created by me. The following is a list of sources for all Figures. In general, both the website and the date accessed are shown:

Figure	Source	Month accessed
2.8	http://en.wikipedia.org/wiki/File:Windows1.0.png	August 2009
2.9	http://en.wikipedia.org/wiki/File:Windows_3.11_workspace.png	August 2009
2.10	Left: http://en.wikipedia.org/wiki/File:Windows_XP_SP3.png Right: http://en.wikipedia.org/wiki/File:Windows_Vista_Desktop.png	August 2009
2.11	http://www.sizlopedia.com/2007/01/31/windows-vista-promotional-song-wow-is-now	August 2009
2.12	Screenshots from www.microsoft.com/design	August 2009
2.13	David Wright (in flickr.com). http://www.flickr.com/photos/dhwright/2621796406/in/set-72157600798616168/	September 2009
2.14	Screenshot from http://www.volvocars.com/intl	January 2006
2.15	Screenshot from http://www.volvocars.com/intl	February 2009
2.16	Left: http://1.bp.blogspot.com/_L8BTRutV0uk/SAyo8ahRMnI/AAAAAAAAEE0/Ho5gQ5I9Nes/s1600-h/Volvo+Ad+1989.jpg Right: http://www.marketingmagazine.co.uk/news/search/923179/Volvo-puts-emotional-appeal-ahead-functional-ability	September 2009
2.17	Screenshots from http://en.wikipedia.org/wiki/Pepsi	September 2009
2.18	Screenshots from http://www.nikeid.com	September 2009
2.19	http://en.wikipedia.org/wiki/File:Roomba_original.jpg	September 2009
2.20	Left: Eirik Newth (in flickr.com). http://www.flickr.com/photos/eiriknewth/282258671/ Right: Julie Vazquez (in flickr.com). http://www.flickr.com/photos/juliesjournal/2664349693	October 2009
3.3	Left: Dave Dyet (in Stock.XCHNG). http://www.sxc.hu/photo/1223460 Right: Lonnie Bradley (in Stock.XCHNG). http://www.sxc.hu/photo/348661	December 2009
3.9	Michael Sauers, in flickr.com. http://www.flickr.com/photos/travelinlibrarian/193383382/	November 2009
3.10	PrEmo by Pieter Desmet, marketed by Susagroup through www.premo-online.com, http://designandemotion.org/content/afbeeldingen/dummie6.jpg	November 2009
3.11	Steinhauen Watches (in flickr.com) Top: http://www.flickr.com/photos/steinhausen/3796907993 Bottom: http://www.flickr.com/photos/steinhausen/3796979873	October 2009
4.2	Pictures taken from the book El libro de los zumos y batidos, by Judith Millidge (Ediciones Robinbook, Barcelona, 2006).	
5.3	Screenshot from Google maps, http://maps.google.com	February 2010
5.4	Screenshot from Google maps, http://maps.google.com	February 2010
8.5	All photos from Stock.XCHNG. From left to right, photo 1: ontanu mihai, http://www.sxc.hu/photo/1159804; photo 2: Evgeniya Bulva, http://www.sxc.hu/photo/871697; photo 3: Luz Alvarez, http://www.sxc.hu/photo/1012519; photo 4: Catherine Sardar, http://www.sxc.hu/photo/877981; photo 5: Scott Snyder, http://www.sxc.hu/photo/727715	November 2009