KEER2010, PARIS | MARCH 2-4 2010
INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

RATING SCALES IN KANSEI ENGINEERING

-MODIFICATONS FOR AN EUROPEAN CONTEXT

Schütte, Simon [a], Eklund, Jörgen [b]

[a] *Linköping University, Department for Management and Engineering, Division of Machine design, SE-58183 Linköping, Sweden, simon.schutte@liu.se*
[b] *Royal Institute of Technology (KTH), Skolan för Teknik och Hälsa, Division of Ergonomics, ALFRED NOBELS ALLÉ 10, SE-14152 HUDDINGE*

ABSTRACT

Kansei or Affective Engineering is a relatively new research field in the West. Linköping University in Sweden has been pioneering research and application within Kansei Engineering since 1999. Several companies have cooperated in different Kansei studies, for example Toyota/BT, Volvo, Saab, Scania. In some of the early studies, there were difficulties to apply the methodology. Reasons for this included shortage of competence within affective design in European organizations, but also the fact that incitements for improvement of product design seemed to be lacking. One of the problems applying the methodology was that European participants in the studies did not accept to make a vast number of ratings during long time periods. Japanese researchers used 5-point semantic differential questionnaires and up to 300 rating-scales per participant and product. In order to overcome this problem, a modification of the rating scales was developed. Hence, data reduction methods such as factor analysis, affinity diagram and Pareto charts were tested and validated in order to reduce the number of Kansei words and product samples used, without compromising the validity of the result. This approach eventually resulted in the development of a universal Kansei Engineering Software (KESo) which reduces the time needed for each Kansei Engineering study. Another objective for our research has been to develop and validate methods for incomplete data collection, i.e. a prioritization of product attributes. Another result from the cooperation with industrial companies is that the companies strongly emphasised that product development studies need to be less time and resource consuming.

Keywords: **Kansei Engineering Software, KESo, Efficient Data collection**

1. INTRODUCTION

Increasing competition makes product manufacturers more aware of the necessity to satisfy also the emotional and affective needs of their customers. Product design and engineering design to an increasing extent include affective aspects through suitable methodology for this. A pioneer in this field was Professor Mitsuo Nagamachi who developed a methodology, Kansei Engineering (KE), in the early 70ies in Japan. This methodology is now widely spread in Japan and Korea, and in addition, it is applied in several other countries. Kansei Engineering is defined as a product development methodology, which translates customers' and users' feelings, impressions and emotions into concrete design parameters (Nagamachi, 1989; Nagasawa, 2002). It can be applied on physical products as well as on services. There are many research groups developing affective methodology and there are many companies and consultants applying it in their product development. There are also several other methods that have been developed within the field of Affective Engineering or Affective Design, many of which have been summarized in the ENGAGE program (http://www.designandemotion.org, 2009).

2. KANSEI ENGINERING IN NORTHERN EUROPE

In 1999 the first conctacts to Japanese researchers and Linköping university were established. The reason was that the company of BT-Industries (now Toyota/BT Handling Equipment) wanted to improve the "driving feeling" of their trucks. First attempts were done in improving ergonomic aspects in order to improve efficiency. An improvements in subjective ratings for drivability were achieved. Still, no suitable method for measuring such important property was available. Since Kansei Engineering was applied, measuring the driving feeling and improving it purposive was possible. New models were introduced to the market at on patent was granted. Since then other sectors such as European car- and lorry manufacturers, household appliance industry and food industry were followed.

Initially, the research team had great difficulties with the application of the Japanese rating scales. Improving the scales and adapting it to the individual needs of the companies and the European context solved that problem. In 2005 Schütte presented a Ph.D. thesis on the topic proposing a general model on Kansei Engineering.

3. SCALE TYPES

Very commonly used methods in subjective assessments are rating methods. They are widely known and provide relatively rich information when compared to other methods collecting affective data. Moreover, data from a great number of participants may be collected with a minimum of resources.

Several different rating methods can be mentioned. Thurstone's Paired comparison technique, Thurstone's 'Equal-Appearing Intervals' or Likert's 'Summated Rating' method also known as 'Likert scale', are commonly used. However these scaling methods are suitable if one wants to evaluate a number of different entities. It is less common that one wishes to assess a group of entities just for one attribute alone. Doing so the Semantic Differential Scales (SD-scales) are more useful (Guilford, 1971). In Kansei Engineering contexts, most (if not all) of those scales above have been used in practical studies. Each of them have their advantages and drawbacks.

3.1. Semantic Differential Scales

In the first half of the 20th century several researchers aimed to analyse the relationship between words, their meaning and underlying ideas, which in this context can be called Kansei. It became apparent that such studies needed powerful instruments for quantification and measurement. Several researchers conducted studies on this problem and proposed different models. Thorndike and Lorge (1944), Cason and Cason (1925) and Zipf (1949) 'Zipf's law' made *similar frequency-of-usage* counts in order to detect laws and relations to its meaning. The underlying theory is that associations on the semantic level appear to be organised in such a way that 'few words and expressions have a higher probability of occurrence whereas many have low probabilities of occurrence' (Osgood, Suci and Tannenbaum, 1957). Inspired by the different political ideologies, which became evident in World War II and the following Cold War, Osgood developed a method to measure the emotional content of a word more objectively. He called this method 'Semantic Differential technique', which more than 30 years later became one of the foundations of Kansei Engineering. His assumption was to divide any expression into two parts:

- The object, 'which is a pattern of stimulation which evokes reactions on the part of an organism', and
- The sign, 'which is any pattern of stimulation which is not the object but yet evokes reactions relevant to the 'object' conditions under which this holds lying the problem for theory'

(Osgood et al., 1957)

For example, the spoken word 'hammer' is not the same stimulus as the object hammer. The former is a pattern of sound waves and the latter a combination of visual, olfactorical and tactual sensations. The word hammer elicits a type of behaviour, which is in some manner relevant to the object hammer. This means that the spoken or read word ‘hammer’ is the sign for the object ‘hammer’. Osgood’s approach is expressed in simplest terms by the question**: Under what conditions does something which is not an object become a sign of that object?**’(Osgood et al., 1957)

To answer the question above, Osgood and his colleagues (Stagner and Osgood, 1946) conducted surveys by means of questionnaires. His subjects were supposed to rate signs (words) of objects like Pacifist, Russian, Germans, Dictator or Neutrality (remember that the experiments were conducted during World War II) on bipolar scales. These scales were defined with a number of contrasting adjectives at each end on which the participants checked that position which best represented the direction and intensity according their point of view. An example of the 7-point rating scale type used is shown in Figure 1.

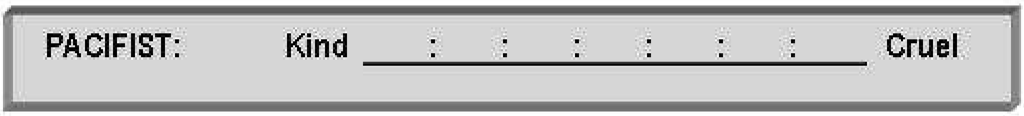

Figure 1: Example of a 7-point rating scale, originally used by Stagner and Osgood (1946).

The data collected can be stacked in a three dimensional raw store data matrix, as it can be seen in Figure 2.

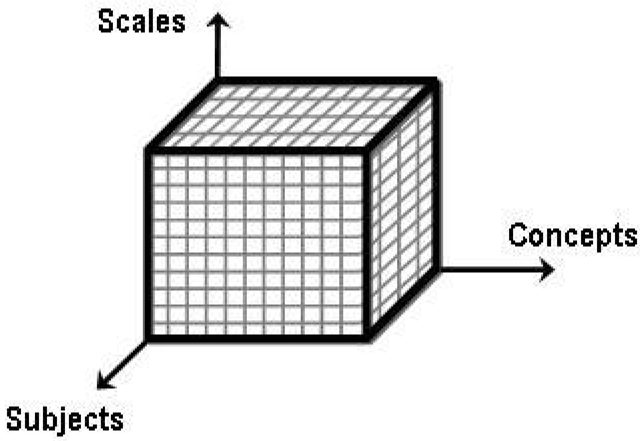

Figure 2: Raw store data matrix, obtained when a group of subjects (x-axis) judges a sample of concepts (y-axis) against a set of semantic scales (z-axis). Each cell contains a number from 1 to 7, representing the judgement of a particular concept on a particular scale by a single subject. (adapted from Osgood and Suci (1969)).

In a following step, the 3 dimensional matrix from Figure 2 is converted to a matrix of inter-correlations by summing together both subjects and concepts. This allows an easy comparison of every scale with every other scale to which the total data contributes. It also avoids spuriously low variability of judgements on single concepts. In addition, factor analysis can easily be run with the data conditioned in that way.

Together with the interrelation matrix, the factor analyses answer questions on how the different word pairs are related to each other; in which way they affect the understanding of a meaning of a certain word; and how to facilitate upcoming experiments. Furthermore, comparing the rotated matrix of the factor analysis from many different experiments led to the discovery of the existence of a common pattern (Carroll, 1959). It could clearly be seen that all examined word pairs span a three dimensional orthogonal vector space as presented in Figure 3. Osgood called this space the semantic space.

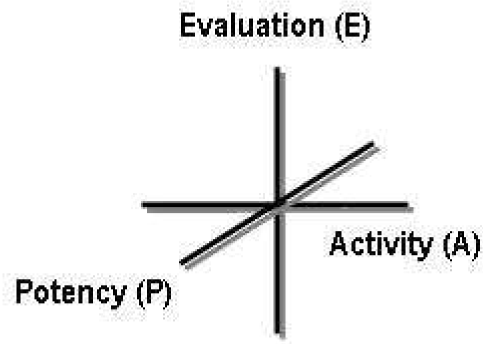

Figure 3: The Semantic Space.

Considering the word pairs in the individual factors it was possible to identify a pattern and name these factors.

- Evaluation (E) usually contains word-pairs like:
  good-bad, timely-untimely, kind-cruel, beautiful-ugly, successful-unsuccessful, important-unimportant, true-false, wise-foolish, etc. All these word-pairs have in common a possibility to evolve into a better or worse stage.

- Potency (P) usually contains word-pairs like:
  large-small, hard-soft, masculine-feminine, strong-weak, etc. These pairs characterise a potential, a capacity for change.

- Activity (A) is characterised by word pairs like:
  active-passive, fast-slow, hot-cold, sharp-dull, angular-rounded, etc. This factor indicates the grade and speed of change.

Applying these factors into the semantic space, as is seen in

Figure 3, these factor-names become the names of the axis. Now it is possible to project every concept in the semantic space and give it an individual position. E.g. a dictator would score high on the potency axis (hard, strong, etc.), low on the evaluation axis (bad, cruel, ugly, etc.) and receives low positive values on the activity axis (active, fast…)

3.2. Scale types in Kansei Engineering

3.2.1. Ordinal scales vs. Ratio scales

Osgood (1957) uses 7 point Semantic Differential Scales gathering for evaluation. Nagamachi and many Japanese researchers use this type of scale (Nagamachi, 1989; Ishihara, Ishihara and Nagamachi, 2000). However, many use a 5 point scale. Also Sinclair (1990) provides a 5 point SD scale. The advantages of SD-scales are that they are recognisable by the participants since many of them have prior experience of working with them (Guilford, 1971).

One problem especially with 5 point-scales is that the type of distribution of the data is difficult to determine. Moreover, the 5 point scale sometimes is experienced as to narrow, in particular when a neutral point is located in the middle. Participants are experiencing the extremes 1 and 5 as overly extreme statements and the remaining three points are not sufficient for making a proper estimation (Schütte, 2005). Therefore it might be better to choose a 7-point scale in those cases. Küller (1975) uses 7-point scales in his method for semantic descriptions of environments (SMB) for similar tasks as in Kansei Engineering . A 7-point scale allows more sensitive ratings, while it is as comprehensive and quick to use as a 5-point scale. However, the problems with determination of the data distribution due to the low number of discrete steps remain.

In medical science, another type of scale is used, the so called Visual-Analogue Scale (VAS), sometimes also called 'Quality of Life Scale' according to its application. It is basically a 100mm horizontal strip, with extreme statements at both ends (Figure 4 (a)). The participants mark their estimation with a cross on it. Despite the fact that this scale possesses discrete steps, the sheer number of them (100) makes it appear as continuous for the participants. It is therefore very sensitive and has no technical details such as numbers or lables that can confuse the study participants. One disadvantage is that it is not commonly known and therefore not easily understood by all participants. Moreover, it is not completely linear (even if this effect is much smaller than in the other scales presented). If ratio properties are assumed, this scale offers the advantage that more statistical methods may be applied. This type of scale has been used in Kansei Engineering context several times.

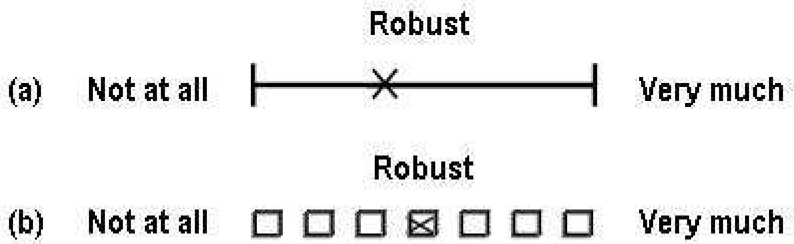

Figure 4: Rating scales used in the experiment in this work. Above a 100mm VAS scale, below a 7 degree modified SD-scale.

In fact it is possible to convert a 5 or 7 degree scale into a AVI scale and vice versa.

Another aspect in this context is that most statistical treatment methods and in particular the methods used in Kansei Engineering require data from continuous scales, i.e. an interval or ratio scale must be deployed (compare Guilford (1971)). However, the SD-scales used here do in fact deploy an ordinal scale. Consequently, the SD-data could not be used e.g. for factor analysis as Osgood (1969) does. Also the data must have a bivariate normal distribution, which e.g. is not the case in the 5 point SD scale used by many researchers (

Figure 5(b)). Nevertheless, experience shows that even these 'dirty' ordinal scales lead to similar conclusions as data from interval scales, and Visual Analogue Scales.

3.2.2. Indexing the extremes

In order to understand what the subject is supposed to do, the scales are often named at their extremes. These are called the anchors. Choosing the labels can have a crucial impact on the results. The labels must be easy to understand for the subject and have to refer to the object of the study.

In Kansei Engineering every Kansei Word is attached to an individual scale. The way the extremes are handled is done differently by different researchers. This is due to cultural differences, deviating experiences or for practical reasons. Osgood et al. (1957) uses synonym and antonym for spanning the range of rating (compare

Figure 5 (a)). This allows reducing the number of ratings to a minimum since both words are rated simultaneously. On the other hand, it sometimes is difficult to find words having exactly the opposite meaning. As an example the word comfort can be mentioned. It is shown that discomfort has a different meaning and can therefore not be used as an antonym on such type of scales (Zhang, Helander and Drury, 1996). Moreover, this type of data is difficult to handle by Quantification Theory Type I.

Nagamachi and many other Japanese researchers use the Kansei Word as an extreme on the left side of the scale, whereas he adds a 'not at all' on the right side (Nagamachi, 2001) (compare

Figure 5 (b). Doing so, liberates from the force to find opposite meaning and makes the scale easy to understand and quick to complete for the participant. Disadvantages are then again a skewed distribution. People experience the scale as un-balanced and see the neutral value more to the left side of the scale (REF?). Hence, (Schütte, 2005) chose to combine the advantages of both scales at the same time excluding most of the problems. The Kansei Word is placed on top of the middle of the scale, while the anchors are labeled as 'not at all' and 'very much' (compare 4(b)). This constellation delivers good data distribution. Together with a 7-point or VAS scale it is one of the most comprehensive solutions for the subjects. Küller (1975) uses such scales in the Semantic Description of Environment (SMB) method. Nevertheless, some disadvantages still remain. The extremes are in many cases considered to be indefinite which in turn means that the distances are not considered to be completely equal. Even if this effect is slighter than in the other cases, even this scale must be called an ordinal scale. Another problem is that subjects which have no opinion or understanding regarding a certain Kansei Word, feel forced to check 'somewhere in the middle' biasing the result. Therefore, if the mean value of a distribution is around the middle value of the scale, it could be because the word either is meaningless for the object evaluated or subjects did not understand the word properly.

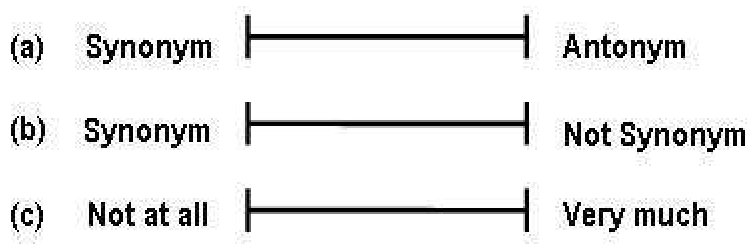

Figure 5: Typical scales used for semantic evaluations. Scale (a) is Osgood's original SD scale (Osgood et al., 1957), (b) is a scale used by many Japanese Kansei Engineering researchers (Ishihara, 2001) and (c) is a modified SD scale by (Küller, 1975).

3.2.3. Multiple ratings on scales

All of the scales presented above have in common that only one rating is made on each of them. However, in a few Kansei Engineering studies multiple ratings have been done on the same scale. In a study on ware house trucks, Schütte (2005) developed a questionnaire where the different truck types are rated as A, B and C on the <u>same</u> scale. In addition, the participants were asked to rate the 'ideal' truck on the same scale, where 'ideal' was defined as 'most suitable for your personal work situation'. The order of the three rated products was altered randomly, but the comparison with the 'ideal' value was always asked for in the end. An example of the modified software used in the main study can be seen in Figure 3.

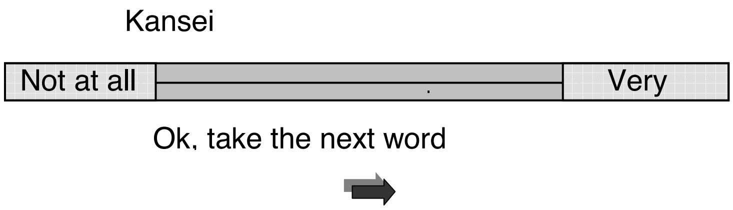

Figure 6: Computerized data collection for main study.

The reason for this was that the study participants should be able to relate their ratings to each other and being able to compare the different truck types with an hypothetical ideal value. This meant that it was not possible to use the absolute value of the rating for evaluation but it was possible to calculate the differences between the truck models and the ideal.

3.2.4. Minimizing the number of ratings and its impact on data quality

According to (Osgood and Suci, 1969) many emotional words need to be examined in order to find out the true meaning of the words. Then the number of words can be reduced using e.g. factor analysis. A thorough analysis using Semantic Differental Method therefore requires long time to conclude. Küller (1975) spent about 20 years for extracting 8 Semantic dimension describing the affective impact of colour setting of student homes. This takes too much time for usage in product development context. Hence, Nagasawa (1997) states that Kansei Engineering usually uses quicker methods for determining the semantic dimensions in a product. However he does not mention how this is done. Schütte (2005) went through a number of Japanese studies, and found out that a typical number of Kansei Words in Japanese studies is around 300 words. This number has to be multiplied with the number of products to be evaluated to get the total number of ratings a participant has to do. Applying this in Sweden, Schütte (2005) found out that most participants are not able to maintain concentrated. Consequently the data quality is declining the more rating that have to be performed. A good compromise between data quantity and quality, according to the authors' experience, is 200 to 300 ratings for each participant. Given a typical number of products being evaluated is 10-20 products, the maximum number of Kansei Words is around 10-15 for each product. This number is of course significantly lower than the 300 Words in Japanese studies.

Hence, methods must be found to reduce the total number of Kansei Words in a suitable way. Schütte suggests a quick factor analysis using only 40-50 participants. Alternatively, Affinity diagram (compare (Bergman and Klefsjö, 1994)) has proven to be a quite reliable tool delivering relatively accurate input to Kansei Engineering studies given the short amount of time necessary to apply.

4. COMPUTERISED DATA COLLECTION

One relevant hinder for introducing Kansei Engineering methodology into industry was in many cases the need of expertise in the areas of statistics, cognitive ergonomics and product development. Many companies the authors were co-operating with felt insecure about the validity of the results since they perceived that the use of methods was not transparent. An often repeated complaint was that they could not do the analysis on their own and had to employ expert consultants on it. More concretely they required tools which would eliminate the need of expertise, as well as reducing the time consumption of a study.

1.1. Kansei Engineering Software (KESo)

In a first step a new client software tool was developed at Linköping University in 2005. In 2009 the latest version of KESo is available as a net-based development tool (www.kanseiengineering.net). The software generates data collection sites on the internet. Collected data will be analyzed regarding the relation between product properties and emotional words (Kansei Engineering Words). From this mathematical models are derived supporting the design of new products.

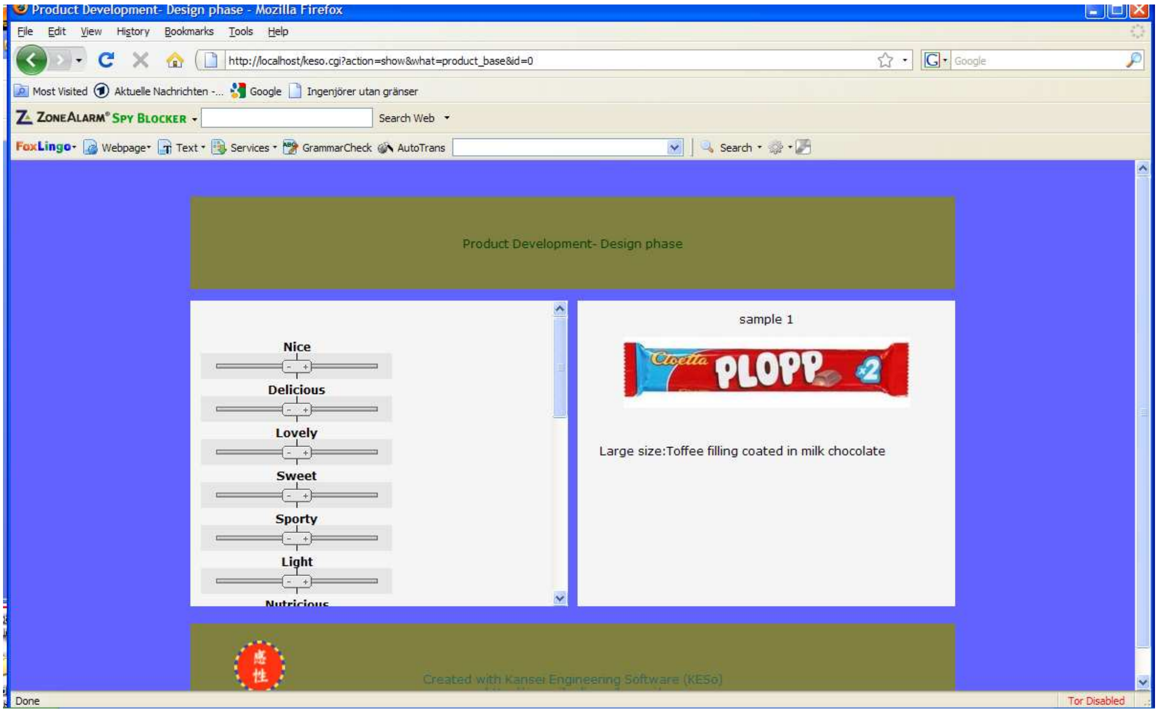

Figure 7: Data collection view of KESo (Schütte, 2006) .

KESo software uses the theories around the semantic space. Visual Analogue Scales (VAS) with the extremes "not at all" and "very much" and the Kansei Word on top proved to deliver valid results and at the same time being comprehensive enough for using in an internet environment (Schütte, 2007). One product is presented at a time to the participants. Then the participants are supposed to evaluate their affective response on a number of VAS scales (see Figure 7). With this data, factor analysis can be carried out revealing the meaning of the Kansei Words. Furthermore, the relationship between the Kansei Words and the specific product properties can be quantified. For this two methods can be used: Quantification Theory Type I (QT1) (Komazawa and Hayashi, 1976) and Rough Set Analysis (RSA) (compare e.g. (Nishino, Nagamachi and Ishihara, 2001)).

5. DISCUSSION AND CONCLUSION

This paper shows that Kansei Engineering as applied in Japan in many cases does not work in European contexts. The reasons are multiple and in many cases unreserached. However, it can be seen in many studies that standard 5 point rating scales used by Osgood and many Japanse researchers are often considered to be to insensitive. Also it is in many cases difficult to find words for the scales extreme points possessing opposite meaning. Putting just one word on one side of a scale and the term "not at all" on the other side produces scued distributions. A good compromise is in fact an AVI-scale with the Kansei Word on top and the expressions "not at all" and "very much" at the extremes (Figure 6).

Another problem for European study participants is the duration of the studies. The longer time is required to fill out the forms the more difficult it gets for the participants to stay concentrated. In turn the data quality is compromised. I has been shown experimentally that a time duration of 20 minutes is a working compromise between good quality of the data and enough data to draw good conclusions. Hence, the number of Kansei Words and the number of properties evaluated and products presented must be reduced in a reasonable manner. Possible methods for this purpose are Factor Analysis, Affinity Diagram and Pareto-Charts.

A major point of criticism from industrial users of Kansei Engineering method has been the need of expertise in the areas of experimental design, statistics and mechanical design. Automated systems gathering Kansei Engineering raw data has been developed for this purpose. Do these systems solve the problems addressed before? In some applications good results have been achieved using automated systems. However, still if the data is not clear-cut and difficult to interpret, expertise is needed. One solution applied at Linköping university is cooperation with industry. Most tasks can be done by industrial personal, difficult questions are solved by the university expert group.

However, Kansei Engineering methodology is still heavily reliant on mathematical tools. The models derived are reductionist in nature with all its consequences. Methods making it possible to draw conclusions from incomplete data collection or very complex products with a great number of internal dependencies would make a step towards filling the gap between mathematic model and reality. These methods would preferably be qualitative methods rather than quantitative.

6. REFERENCES

Bergman, B. and B. Klefsjö (1994). Quality. Lund, Studentlitteratur.

Carroll, J. B. (1959). Review of The Measurement of Meaning. Semantic Differential Technique. J. G. Snider and C. E. Osgood. Chicago, Aldine Publishing: 96-115.

Cason, H. and E. B. Cason (1925). Association tendencies and learning ability. J.exp.Psychol.: 167-189.

Guilford, J. P. (1971). Psychology Methods. New York, McGraw-Hill Publishing Company.

Ishihara, S. (2001). Kansei Engineering Procedure and Statistical Analysis. Workshop at International Conference on Affective Human Factors Design. Singapore.

Ishihara, S., K. Ishihara, et al. (2000). Kansei Analysis and Product Development of Hair Treatment. Ergon-Axia 2000, Warsaw, Poland, Central Institute for Labour Protection.

Komazawa, T. and C. Hayashi (1976). A statistical method for quantification of categorical data and its applications to medical science. F. T. de Dombal and F. Gremy, North-Holland Publishing Company.

Küller, R. (1975). Semantisk Miljö Beskrivning (SMB). Stockholm, Psykologiförlaget AB Liber Tryck Stockholm.

Nagamachi, M. (1989). Kansei Engineering. Tokyo, Kaibundo Publishing Co. Ltd.

Nagamachi, M. (2001). Workshop 2 on Kansei Engineering. International Conference on Affective Human Factors Design, Singapore, 2001.

Nagasawa, S. (1997). Kansei evaluation using fuzzy structural modelling. Kansei Engineering 1. M. Nagamachi. Kure, Kaibundo Publishing Co., Ltd.: 119-125.

Nishino, T., M. Nagamachi, et al. (2001). Rough set analysis on Kansei evaluation of color. The International Conference on Affective Human Factors Design, Singapore, Asean Academic Press.

Osgood, C. E. and G. J. Suci (1969). Factor analysis of meaning. Semantic differential technique - a source book. C. E. Osgood and J. G. Snider. Chicago, Aldine Publishing Company**:** 42-55.

Osgood, C. E., G. J. Suci, et al. (1957). The measurement of meaning. Illinois, University of Illinois Press.

Schütte, R. (2006). Kansei Engineering Software (KESo). Linköping.

Schütte, S. (2005). Engineering emotional values in product design- Kansei Engineering in development. Institution of Technology. Linköping, Linköping University.

Schütte, S. (2007). Collecting data for Kansei Engineering Studies. the First European Conference of Kansei Engineering, Helsingborg, Sweden.

Sinclair, M. A. (1990). Subjective assessment. Evaluation of Human Work- A practical ergonomics methodology. Wilson, J. R. and E. N. Corlett. London, Taylor and Francis**:** 58-88.

Stagner, R. and C. E. Osgood (1946). Impact of war on a nationalistic frame of reference. Journal of soc. Psychology. **14:** 389-401.

Thorndike, D. E. and I. Lorge (1944). The teachers workbook of 30,000 words. New York, Bureau of Publications, Teachers College, Columbia University.

Zhang, L., M. G. Helander, et al. (1996). "Identifying Factors of Comfort and Discomfort in Sitting." Human Factors **38**(3): 377-389.

Zipf, G. K. (1949). The psychobiology of language. New York, Houghton Mifflin.