SEGMENTATION IN KANSEI ENGINEERING STUDIES BASED ON THE EMOTIONAL RESPONSE.

Lluís MARCO-ALMAGRO[a], Xavier TORT-MARTORELL[a] and Simon SCHÜTTE[b]

[a] *Universitat Politecnica de Catalunya | BarcelonaTech, Spain, lluis.marco@upc.edu*
[b] *Linköpings Universitet, Sweden, simon.schutte@liu.se*

ABSTRACT

Socioeconomic and demographic data of participants are often collected when performing a Kansei Engineering study for stratification purposes. This paper offers an alternative stratification procedure, directly based on the emotional response that participants give to prototypes. This approach can deliver groups that are more meaningful for the Kansei Engineering study. Socioeconomic and demographic data (among other kinds of data) can then be used to characterize the obtained emotional groups.

Keywords: *hierarchical cluster analysis, customer segmentation, Kansei engineering study*

1. INTRODUCTION

Quite often in Kansei Engineering (KE) 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 done with this data in many occasions. 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 for a KE study often 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". However, it would be possible to find women fitting this description with very diverse tastes related to a product.

There are, in fact, two reasons that justify the advantages of segmenting participants of a KE study based on their emotional response:

1. A conceptual reason: As we have stated, it makes more sense to use data from the Kansei Engineering study to make segmentations. The socioeconomic or demographic data usually collected for each person can then be used to try to typify the *a posteriori* groups that come from the emotional segmentation.

2. A practical reason: If the emotional segmentation is not performed and data from all participants are analysed together, chances are that properties that do have an influence in the emotional response are not detected as significant. Basically, if two different groups react on a Kansei word in an opposite way, both effects could be compensated and thus that property not seen as affecting that Kansei word.

So the idea of this paper is the following: finding groups directly based on the emotional response on products (using data from the KE study), and not on other external data. These groups can then be characterized based on socioeconomic or demographic data, if possible. The rest of the paper develops a method, based on cluster analysis, to perform this emotional segmentation.

2. A SUGGESTED METHOD FOR AN EMOTIONAL SEGMENTATION, WITH AN EXAMPLE

The method suggested for performing an emotional segmentation comprises two steps:

1. A cluster analysis is conducted for each of the Kansei words, having participants in the study as individuals for the clustering. In this way, groups of participants are created, and these groups will most probably be different for each response.

2. A new cluster analysis is conducted on the groups found in the previous step, having Kansei words (responses) as individuals for the clustering. In this way, a global set of groups of participants is achieved.

An example will be used to better illustrate the details of the suggested method.

2.1. An example to illustrate the method.

The example shown in this section and used to illustrate the suggested segmentation method is adapted from a study first appeared in the doctoral thesis of Marco-Almagro (2011). The choice of domain for this Kansei Engineering study are fruit juices, and specifically its presentation just before being drunk. Prototypes were presented with photographs on a computer screen (so only sight was considered in this study, there was no touching or tasting of the products).

The set of emotional responses of interest (semantic space) were the following: refreshing, healthy, exotic, seductive, natural, relaxing, and tasty. For the space of properties, five different properties (factors) were used, each one at two possible values (Table 1).

Table 1: Properties and levels used in the juices experiment

Property	Levels
Straw	Yes / No
Decoration	Yes / No
Ice	Yes / No
Container	Glass / goblet
Colour	Yellow / orange

The set of prototypes for the study were built following the design matrix shown in Table 2.

Table 2: Properties and levels used in the juices experiment

	Straw	Decoration	Ice	Container	Colour
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

Figure 1 shows the 16 fruit juices used in the KE study.

Figure 1. The 16 prototypes created for the fruit juice experiment

A total of 19 participants rated each one of the juices on a 7 scale Likert scale for each one of the Kansei words. The data collection was done at a neutral room, all participants at the same time, looking at one fruit juice after the other in a random order, until the 16 juices were rated.

2.2. A first cluster analysis Kansei word by Kansei word.

A hierarchical cluster analysis is done for each one of the Kansei words (7 in the case of our example). For each Kansei word, the objects are all the prototypes (there are n stimuli, 16 juices in our case) and the variables are the participants (there are p participants, 19 participants in our case).

In a hierarchical cluster analysis, a distance of each object to all other objects is calculated (Manly, 2005). Groups are then created usually by agglomeration: all objects start alone in groups of one. The closest groups are gradually merged until finally all objects are in a single group. A cluster analysis needs:

The definition of a distance: the most common distance is the Euclidean distance (the one you have when using a rule on a paper). In our case, we are interested not in the absolute distance between two participants, but on the distance of their rating profile. So our suggestion as distance between participants *i* and j is the following:

$$(,)=1-\qquad(,)$$

If both participants have exactly the same rating profile, correlation will be 1 (so distance will be 0). In the case that one profile is completely opposite to the other, correlation will be –1 and the distance will be the maximum (2).

In order to compute correlation between participants i and j, instead of using the common Pearson coefficient of correlation, the polychoric coefficient of correlation seems more appropriate. The polychoric correlation assumes there are two continuous normally distributed latent variables, but they become ordinal variables when observed (Holgado– Tello et al., 2010). This is the case with Kansei Engineering studies, where an underlying continuous variable is present (so sometimes visual analogue scales are used to measure it), but often 5, 7, or 9-point ordinal scales are used.

A linkage method. Once observations are joined in groups in a cluster analysis, a linkage method is needed to know how to calculate the distance between clusters. Although there are several linkage methods, the Ward's linkage method seems to be the most appropriate for KE data (Ward, 1963).

Figure 2 shows the dendrogram from the hierarchical cluster analysis for the Kansei word Refreshing. A reasonable criterion to decide the number of groups is cutting the dendrogram 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. In this case, we have decided to split the participants in two groups.

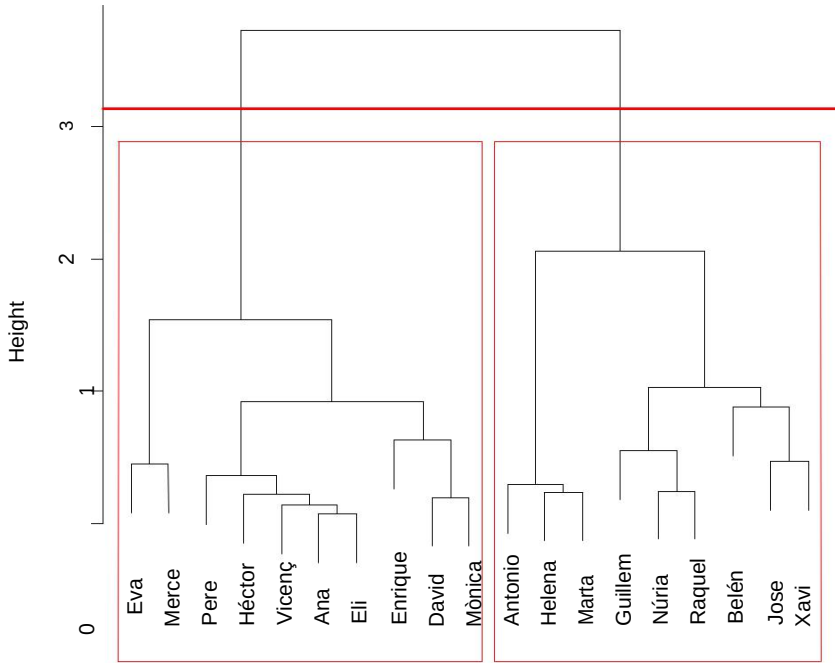

Figure 2. Dendrogram for the Kansei word Refreshing

This hierarchical cluster analysis should be performed for each one of the Kansei words in the study. Figure 3 shows the created groups for each one of the Kansei words in the study.

	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

Figure 3. Groups for each participant and Kansei word in the fruit juice experiment

When analysing results in the synthesis phase using, for instance, QT1, one can detect the differences between groups. Take again, for instance, Kansei word Refreshing. Figure 4 shows that having ice in the juice affects the perception of Refreshing. However, this effect is higher in group B (in blue) than in group A (in red). We can also detect that colour has an effect on this Kansei word, but in opposite ways depending on the group. Notice that the fact the effect of Colour cannot be detected when analysing all participants together, as the opposing effects will be compensated.

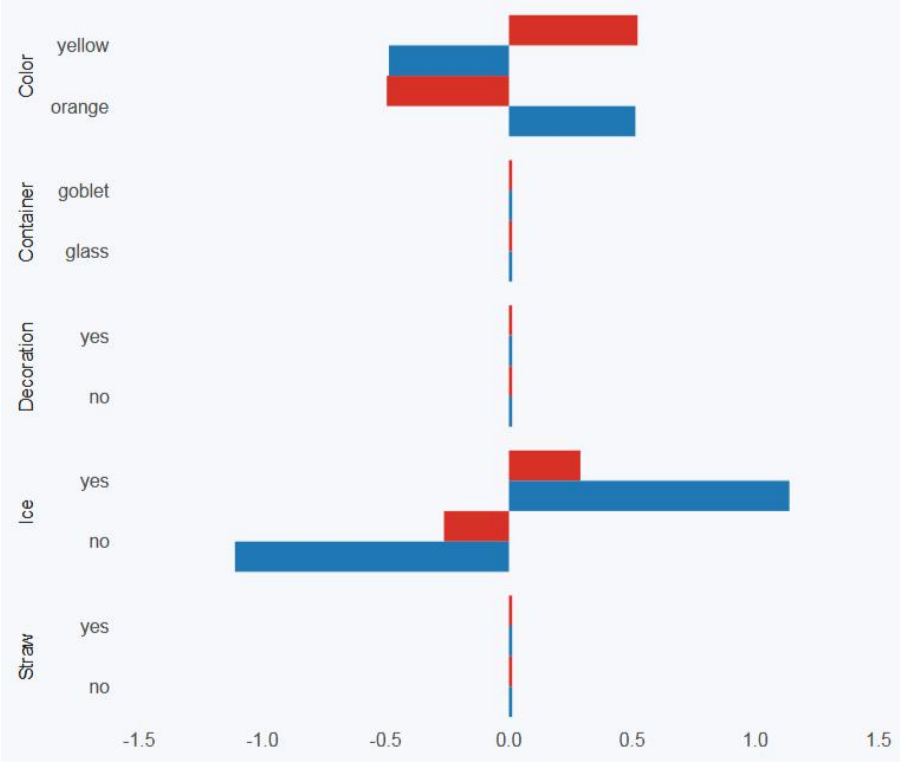

Figure 4. QT1 results for Kansei word Refreshing in the fruit juices experiment

2.3. Clustering groups of participants

After the emotional segmentation of participants done Kansei word by Kansei word, the aim now is performing a global grouping of participants, taking into account information from all Kansei words together.

The global clustering of participants will be done, again, using a hierarchic cluster analysis. Ward's linkage method will be used again. 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.

A suggested distance could be the following:

(,) = h

For example, (,) = 0, as columns Ana and Eli in Figure 3 are exactly the same (all A). (,) = 2, as they fall in different groups in Kansei words Refreshing and Relaxing.

Figure 5 shows the result of the hierarchical cluster analysis for all participants.

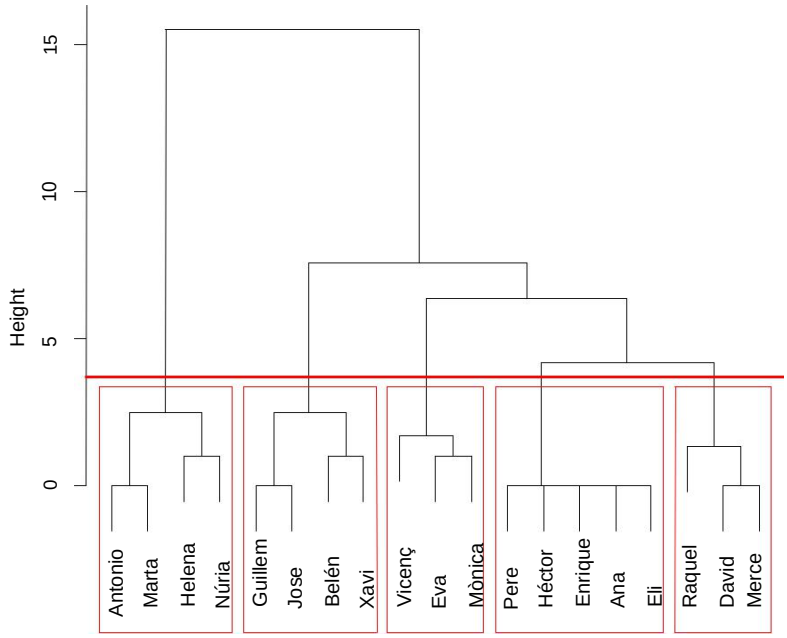

Figure 5. Dendrogram grouping subjects

In this case, 5 different groups of participants have been selected.

3. CONCLUSIONS

This paper has suggested a way to cluster participants of a Kansei Engineering study based on the emotional response that users have to the product under examination. Although socioeconomic and demographic data can be useful to stratify, the suggestion is using these data to characterize groups coming from the emotional segmentation.

The emotional segmentation is conducted in two steps. In the first step, participants are grouped in each Kansei word. In the second step, a cluster analysis is performed with the groups of participants for each Kansei word, thus having a global grouping of participants.

This global grouping could be later used as the basis for further studies (either quantitative or qualitative).

REFERENCES

Holgado–Tello, F. P., 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*, *44*(1), 153-166.

Manly, B. F., & Alberto, J. A. N. (2016). *Multivariate statistical methods: a primer*. CRC Press.

Marco Almagro, L. (2011). Statistical methods in Kansei engineering studies. Universitat Politecnica de Catalunya – BarcelonaTech. Doctoral Thesis.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, *58*(301), 236-244.