# Revealing marketing preferences using electroencephalography

Juho Aaltonen, Leifu Chen, Joonas Laitinen,
Salomo Lassfolk, Ali Salloum, Timo Tommila

This study was completed as a part of *Life Science Technologies Project Course* (JOIN-E3000) at
Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland.

***Abstract*— Electroencephalography provides tools to measure physiological and neural signals to better understand customers' preferences and decisions. This can help the businesses to improve their advertising, product development and other marketing strategies.**

**We applied the neuromarketing tools to study the effects of advertisement language, the origin of the company and the adjective describing the company. Two of these factors could be affected by the marketing experts: Should the product be marketed in the local or more universal language? What adjective produces the largest positive response among certain population? Neuromarketing can answer these type of questions.**

**In this project, we planned the whole study from the scratch. We studied the brain responses of six persons to gain more insights around the aforementioned themes. Despite of the small sample size, we found with our multiple statistical methods clear differences in brain responses when language, origin or the adjective was changed. Interestingly, the magnitude and direction of the response depended heavily on the subject. This is either because of small sample size or the unique nature of human minds.**

***Key words:* Neuromarketing, EEG, Neural correlates, N400 component, Bootstrap, GFP, Cluster-based permutation test, Spatiotemporal cluster test**

## 1. INTRODUCTION

Neuromarketing uses neuroimaging techniques to understand consumer behavior to stimuli used in marketing [1][2]. Traditional tools used for marketing research fail to reveal the subconscious behavior mechanisms of consumers [3]. Neuromarketing can be used to reveal hidden preferences of consumers, leading to more effective advertisements and possible changes to products in development [4].

On our study, brain responses from adjective — brand word combinations stimuli were recorded using electroencephalography (EEG). For each brand, matching and conflicting adjectives were used to see if there are any differences in the responses. There were Finnish and non-Finnish brands, and the adjectives were in English and Finnish.

Hypotheses for this study were:

- Conflicting adjective — brand combinations have stronger responses than matching adjective — brand combinations
- Responses of Finnish brands equal to non-Finnish brands
- Responses have the same strength when the test is done in Finnish or in English

The hypotheses were tested by analyzing EEG data collected from our subjects. Results are then demonstrated with different methods.

## 2. METHODS

### 2.1. Brands

For the study, 60 Finnish and 60 non-Finnish brands were chosen by following criteria: length of the brand's name should be three or more letters long, and it should include only letters (excluding Scandic characters, German umlauts, etc.). Only brands with single-word name were accepted.

For each of the 120 brands, one matching and one conflicting adjective were chosen both in English and in Finnish. Adjectives were chosen using similar criteria than with the brands. Also, the adjectives were chosen based on their rationality with the brand in question. Approximately half of the brand — adjective pairs had an adjective before the brand and the other half had the order other way around. Both adjectives and brands were written with lowercase letters to avoid unwanted emphasis to capital letters. An excerpt from the brands and adjectives is presented on the table I.

### 2.2. The study subjects

The study was performed for six subjects (median age 21.5 and standard deviation 4.9). The subjects

TABLE I

A few excerpts from the brand lists in English.

| Brand | Matching adjective | Brand first? | Conflicting adjective | Brand first? |
|---|---|---|---|---|
| abloy | safe | Yes | risky | Yes |
| fiskars | sharp | No | fragile | Yes |
| lego | creative | No | impractical | No |
| visa | handy | Yes | useless | No |

were all Finnish university students (three females and three males) and they all were right handed, also every subject spoke Finnish as their native language. The study subjects were asked to answer a survey before the experiment (see 2.3). This survey gathered information about sex, age, family's socioeconomic background, study field, phone model, interest towards design, fast-food consumption, importance of quality versus cheap prices and the subject's favorite and least favorite brands. Sex, ages, study fields and family backgrounds from this survey are listed on the table II.

After the experiment, the subjects were asked to tell which brands (see 2.1) and adjectives were unfamiliar. The subjects were also asked for a general feedback of the experiment.

TABLE II

Summary of the study subjects. Based on surveys.
* Note: N = non-academic, A = academic

| Subject | Sex | Age | Study field | Background* |
|---|---|---|---|---|
| Subject 1 | Female | 21 | Architecture | N middle class |
| Subject 2 | Female | 21 | Architecture | A middle class |
| Subject 3 | Female | 33 | Technology | N middle class |
| Subject 4 | Male | 21 | Medicine | A middle class |
| Subject 5 | Male | 22 | Technology | A high class |
| Subject 6 | Male | 27 | Technology | N middle class |

### 2.3. *The experiment*

The experiment recordings were recorded in the shielded room of the Aalto Behavioral Laboratory (ABL) with 32-channel EEG (BrainProducts GmbH). The electrodes were placed using the 10-20 system. Recordings were done against a reference electrode *FCz*, and the ground electrode was *AFz*. In addition to EEG recording, electrooculography (EOG) was recorded using four channels: three on the right eye (over, below and at the corner of the eye) and one on the left eye (at the corner of the eye). For EOG, reference and ground electrodes were placed on middle points of right and left clavicle bone, respectively. Sampling rate of the recording was 1000 Hz.

The experiment was divided into three parts: English test, Finnish test and opinion test. On the English test, 60 brands (Finnish and non-Finnish) with two adjectives in English were chosen from brand data (explained on 2.1) by random selection. Random selection was chosen to reduce the effects of possible bad brand selections. The stimuli were created this way for each study subject separately. 120 stimuli were divided into six subgroups. The stimuli were presented to the subject by using Python-based PsychoPy-software (version 3.0) [5] and 24" stimulus monitor. A length of one stimulus was 1.5 seconds and a pause between two stimuli was 2 seconds, also there was a 60 second pause between each subgroup.

The Finnish test was similar to the English test, except the adjectives were in Finnish, but the brands were same with the English test. The Finnish test was performed approximately 5 minutes after the English test. The opinion test was performed approximately 5 minutes after the Finnish test. On the opinion test 100 stimuli from the Finnish and the English tests were chosen by random. The subject was asked to give her/his opinion about the brand — adjective combination: agree, disagree or neutral opinion. The subject was able to move to the next combination at her/his own pace, and the responses were recorded with a keyboard. The opinion part was done without using EEG or EOG.

### 2.4. *Data preprocessing*

Data from the experiments (2.3) were processed using MNE Python (version 0.17.1) [6]. First, the data were visually inspected, and the space outside the stimuli was cropped. After that the cropped data was band-pass filtered with high-pass value of 1 Hz and low-pass value of 40 Hz. The bad noisy channels were chosen by visual inspection and by the experiment notes. These channels were then interpolated from the surrounding channels with MNE Python's *interpolate_bads*-function. Time interval from -0.2 second to 0 second was used to apply the baseline correction.

Independent component analysis (ICA) was performed using MNE Python's *fastica* method to remove EOG and other artifact components, such as noisy channels in which interpolation and filtering did not help. 25 components were computed for each data set, and components from the EOG artifacts were excluded automatically. Other artifacts components were excluded manually. Manual excluding was based on components low, or nonexistent, epoch variance. Some ICA components are presented in the figure 1 as an example.

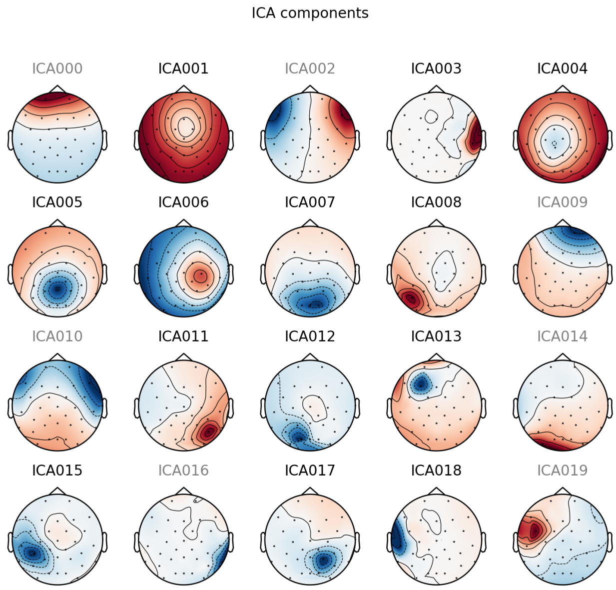


Fig. 1. A part of ICA components from the English test of subject 4 as an example. The components with a grey number are the excluded artifacts components.

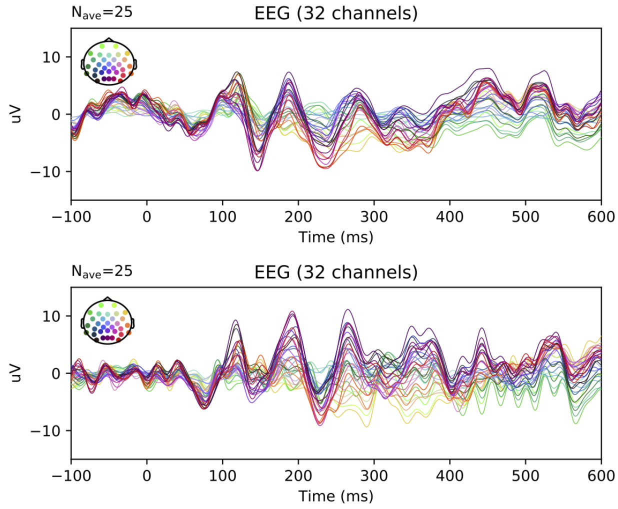


Fig. 2. Averaged ERPs of subject 4 on Finnish test with Finnish brands. Figure above has matching adjective and figure below has conflicting adjective. The N400 is detectable on the latter one.

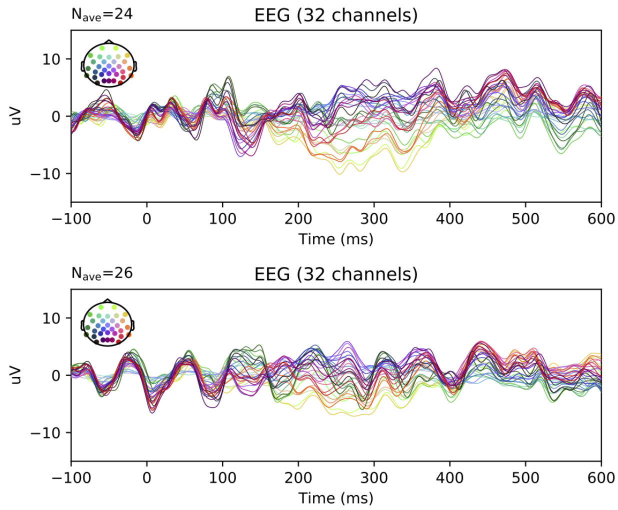


Fig. 3. Averaged ERPs of subject 4 on English test with Finnish brands. Figure above has matching adjective and figure below has conflicting adjective. The N400 is detectable on the latter one.

The data from the opinion test was analyzed for each subject, and it was used to correct some stimuli to correct groups (according to subject in question). Stimuli which had neutral opinions were excluded. Also, stimuli with unknown brands, adjectives, favorite and least favorite brands based on the surveys (2.2) were excluded. The matching or conflicting adjective — brand pairs which were not excluded were converted to epochs to be used for the analysis. Epochs had a time frame from 0.2 seconds before to 0.7 seconds after the stimuli. This whole preprocessing pipeline was done for both English and Finnish tests for each subject separately.

## 3. RESULTS

### 3.1. Event-related potentials

As several papers in language research (for example [7, 8, 9]) have claimed, a stimulus, e.g. visual or auditory, with an contradictory content causes a large negative peak around 400 ms called N400 component. A stimulus with congruent content does not cause the N400 component. Reason why the N400 appears with stimuli with a contradictory content is still unclear.

At this study the congruent contents respond to matching adjective — brand pairs, and vice versa contradictory contents equals to conflicting adjective — brand pairs. Averaged event-related potentials (ERP) from above-mentioned stimuli groups, were used to study the possibility of the N400 components.

A bit surprisingly, when non-Finnish brands were shown, no negative peak around 400 ms was detectable from any subject in the Finnish test nor in the English test. This was repeated with Finnish brands except with one subject. The subject 4 had a negative, detectable, peak at 400ms with Finnish brands on both, Finnish and English tests, as shown in the figures 2 and 3, respectively. These negative peaks resemble the N400 component in above-mentioned studies, so we will assume that the peak on the tests of subject 4 was the N400 component. In the fig. 3 from the English test, the possible N400 component is seen more clearly than in the similar figure from Finnish test.

Why the N400 appeared only with one study subject and why it wasn't seen with non-Finnish brands is unclear. Anyhow conclusions about the matching versus conflicting adjective — brand pair response differences at 400 ms cannot be made.

### 3.2. *Bootstrapping*

We applied bootstrapping method to study the averaged global field power graphs of each subjects. Bootstrapping is a statistical tool to estimate the true sampling distribution by sampling a data set with replacement. Figure 4 shows one of the global field power (GFP) plots, where one can compare visually how the origin of the company affects the response. Whether the largest response was caused by domestic or foreign business varied between the subjects.

We detected the consequences of P200 component also in all global field power plots. The high spike at about 200 milliseconds after the external stimulus is linked to basic and low-level perception. Therefore, we do not believe that our experimental conditions have influence on the P200 region.

Naturally, the bootstrapping approach provides us a chance to compute many statistics such as mean, p-values and confidence intervals. The p-values are listed in Table III. To illustrate the power of bootstrapping, we visualized the bootstrap distributions in one plot. The figure 6 displays how for certain subject 1 all the four English tests had weaker response than the Finnish ones, thus located in left. The largest mean and variance was related to the Finnish test where a foreign company with a conflicting adjective was showed to the subject. There existed little overlapping between the bootstrap distributions. But again, this same plot varied widely between the subjects.

TABLE III

P-VALUES FROM BOOTSTRAP HYPOTHESIS TESTING.

| Test / Subject no. | Adjective | Origin | Language |
|---|---|---|---|
| Subject 1 | 0.0 | 1.0 | 0.0 |
| Subject 2 | 0.009 | 1.0 | 1.0 |
| Subject 3 | 1.0 | 0.0 | 0.0 |
| Subject 4 | 0.0 | 1.0 | 0.498 |
| Subject 5 | 1.0 | 0.0 | 0.053 |
| Subject 6 | 0.73 | 0.0 | 0.0 |

### 3.3. *Cluster analysis*

The cluster-based permutation test was applied to solve the multiple comparison problem. In our case,

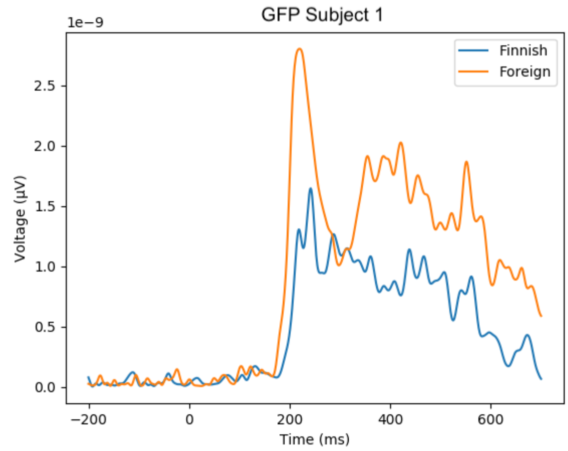


Fig. 4. Averaged global field power curves show clearly how the foreign companies produced stronger responses for the subject 1.

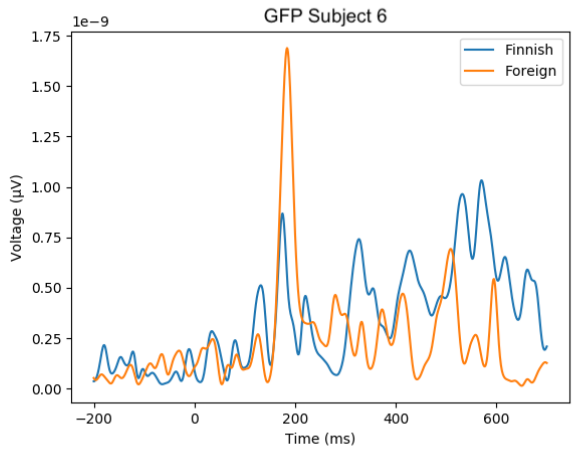


Fig. 5. The subject 6 had larger responses towards Finnish companies in the conscious cognitive processing region.

it is used to study the response between different kinds of adjective-brand word pairs. As we mentioned before (see 2.1), all the word pairs were expected to be divided into various categories. Some of them were predetermined in the experiment design phase, such as Finnish brand with matching adjective, Finnish brand with conflicting adjective, non-Finnish brand with matching adjective, non-Finnish brand with conflicting adjective. The rest of the categories were collected from post-experiment surveys, which included favourite brand with matching adjective, favourite brand with conflicting adjective and unknown brand.

For each subject, MNE Python built-in *permutation_cluster_test* function was used to test if the if the evoked response is significantly different between categories. The null hypothesis (H0) of the permutation test is that the data in the experimental con-

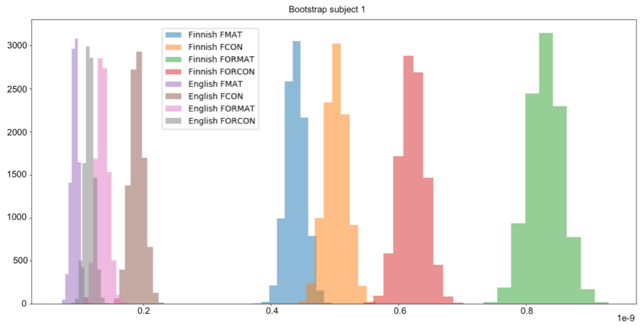


Fig. 6. The bootstrap distributions visualized for each control for subject 1. B = 10 000.

ditions come from the same probability distribution (i.e. the data in the different conditions cannot be distinguished) while the alternative hypothesis (H1) is that the data in the experimental conditions do not come from the same probability distributions (i.e. the data are different). It is a common practice to say that data in the two experimental conditions are significantly different if the p-value is smaller than the typical critical alpha-level (0.05). In Table IV, the significant results are reported based on the p-value. For example, in the first row, the cluster-based permutation tests revealed a significant difference between Finnish brand with matching adjective and Finnish brand with conflicting adjective condition for subject 5. However, we are not able to generalize the difference of two different conditions across different subjects. Additionally, we can visualize the difference (see Figure 7) and conclude that, in the latency range from 600 ms to 700 ms post-stimulus, the difference was most pronounced. We also have to mention that, even for the same subject, the latency range are different in each case. For example, the pronounced difference between non-Finnish brand with matching adjective and non-Finnish brand with conflicting adjective, was in the latency range between 550 ms to 650 ms (see Figure 8).

### *3.4. Spatial analysis*

The spatial difference was studied using two methods: permutation t-test and spatiotemporal cluster test [10]. Permutation t-test was applied to find out the sensors that are statistically significant under certain control. Based on the results we decided to apply the latter test, as it was more applicable to our problem. The cluster test was performed using MNE Python's *spatio_temporal_cluster_test*-function.

Spatiotemporal cluster test was used for studying distinctive effects between different types of stimuli by comparing pairs by their responses. The analysis was started by 100 permutations. After the selection of significant clusters, the analysis was continued further with the number of 1100 permutations.

Clear differences were found between the English and the Finnish tests from five out of six study subjects in the response strength. For example, in the figure 9 from the subject 4, a difference was seen in all three sensor locations: left, midline and right. Most clearly it was seen on the right side's sensors from 200 ms to 450 ms, and on the left side's sensors from 250 ms to 550 ms.

Spatiotemporal cluster test was also tried on adjective comparison. Only one of the study subjects had any statistical difference when matching - and conflicting adjective — brand pairs were compared. This difference was around 600 ms.

TABLE IV
P-VALUES FROM CLUSTER-BASED PERMUTATION F-TEST.
* NOTE: FIN = FINNISH, NON-FIN=NON-FINNISH, CONF = CONFLICTING, FAV=FAVOURITE

| Test Subject no. | Condition 1 | Condition 2 | p-value |
|---|---|---|---|
| Subject 5 | Fin Match | Fin Conf | 0.002 |
| Subject 5 | Fin Match | Fav Match | 0.044 |
| Subject 5 | Fin Match | Unknown | 0.04 |
| Subject 5 | Fin Conf | Fav Match | 0.037 |
| Subject 5 | Fin Conf | Fav Conf | 0.019 |
| Subject 5 | Fav Conf | Unknown | 0.011 |
| Subject 5 | Non-Fin Conf | Fav Match | 0.025 |
| Subject 5 | Non-Fin Conf | Fav Conf | 0.05 |
| Subject 5 | Non-Fin Match | Non-Fin Conf | 0.032 |
| Subject 2 | Non-Fin Match | Fav Match | 0.045 |
| Subject 2 | Non-Fin Conf | Unknown | 0.015 |

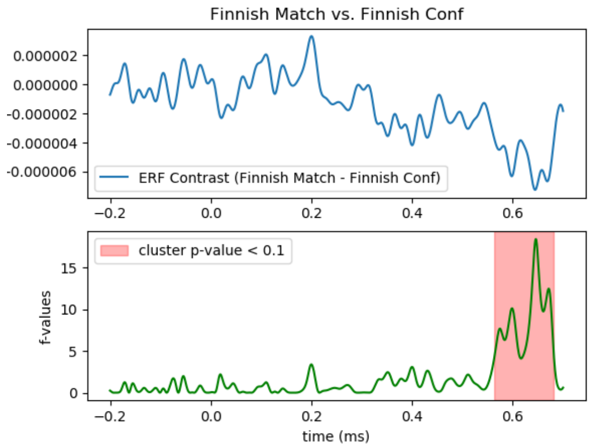


Fig. 7. Subject 5's evoked response is significantly different between Finnish brands with matching adjective and Finnish brands with conflicting adjective.

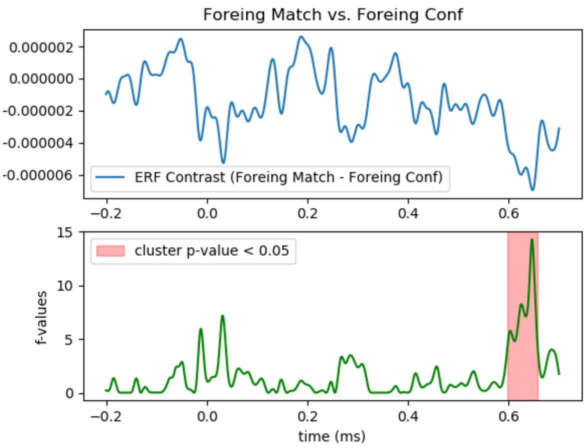


Fig. 8. Subject 5's evoked response is significantly different between non-Finnish brands with matching adjective and non-Finnish brands with conflicting adjective.

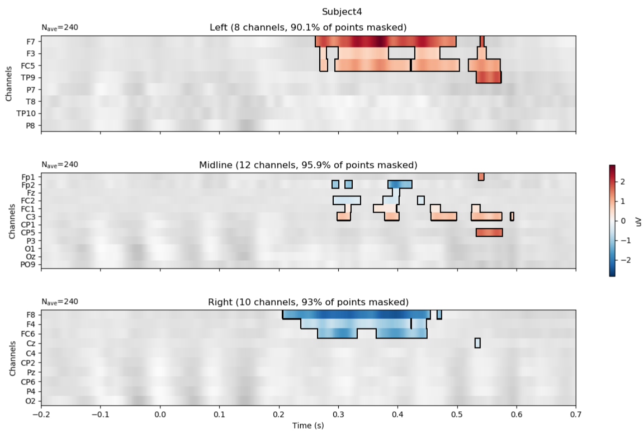


Fig. 9. Map of differences, between the English test and the Finnish test, on the response strength of spatiotemporal temporal cluster test on subject 4. Left side's sensors on top of the figure, midline sensors on middle and right side's sensors on the bottom of the figure.

## 4. DISCUSSION

Based results provided in chapter 3, we can see some statistical difference in responses to different types of stimuli. When using bootstrapping method for each hypothesis, difference was found with three out of six subjects. That means that from our six Finnish students that we used subjects 50 % showed difference in global field power when studying the difference that language (Finnish vs English), origin of the brand used (Finnish vs non-Finnish) and adjective (matching to the brands image vs conflicting to the brands image). However we avoid doing general claims, due to the fact our sample size was really small and the results were not consistent.

We used cluster-based permutation test to address the multiple comparison problem. For each subject, we have tested 15-20 category comparison, which depends on the number of categories we found. Among more than 100 comparisons, 11 of them revealed the significant difference (see IV). Surprisingly, 9 out of 11 significant comparison are from the response of subject 5. Therefore, no results could be generalized based on our tests. This depend on the multiple factors, e.g. the length of the time window or the threshold chosen to select the samples of a cluster. Therefore, we may achieve better performance by improving the sensitivity for widespread clusters. To improve the sensitivity efficiently, we could have more trails or increase the threshold for cluster inclusion.

Spatiotemporal cluster test was used to study the effect of language and adjectives. As described on 3.4, five out of six study subjects responded differently to the native language (Finnish) and to English. The effect of the adjective was not seen as powerful as the effect of the language. To find better results more permutations could have been done.

## 5. CONCLUSION

We studied, how the brains respond to different adjective — brand combinations using EEG. The purpose of the study was to discover differences in the responses between matching and conflicting adjectives related to the brands, and also if there is any significance to the language of the adjective presented or the origin of the brand.

As discussed we do not want to make any strong claims for any hypothesis. Bootstrapping showed that there are too much variation between study subjects that we could confirm or disprove our hypothesis. On some subjects results may apply. The cluster-based permutation test, on the other hand, did not show any results that can be used to make conclusions. Interestingly spatiotemporal cluster test gave us the only strong result, the human brains react with different power to the native language (now Finnish) and to English stimulus. The small sample size and the unique nature of human mind are the two most essential factors that will explain the difference in results between the subjects and the test.

This research gave us preparedness to do corresponding research in the future. We learned valuable things about running an experiment and working with study subjects. This research broadened our knowledge in the field of Neuromarketing and gave us an opportunity to put our previous knowledge in to the practice. The phase were we worked with the data, taught us much about data analysis and data science.

## 6. APPENDIX

Github repository for the scripts that were used in this project. The repository consists of data-analysis scripts, the midterm and final presentations, list of used brands with adjectives and the stimulus generators.

## 7. ACKNOWLEDGMENTS

The authors would like thank our instructor Dovile Kurmanaviciute for all the help provided during the study and the course. And also Veli-Matti Saarinen from ABL for the help provided for the measurements. And finally Tomi Laurila for organizing this course.

## REFERENCES

[1] Nick Lee, Amanda J Broderick, and Laura Chamberlain. “What is ’neuromarketing’? A discussion and agenda for future research”. In: *International Journal of Psychophysiology* 63.2 (2007), pp. 199–204.

[2] Vlăsceanu Sebastian. “New directions in understanding the decision-making process: neuroeconomics and neuromarketing.” In: *Procedia-Social and Behavioral Sciences* 127 (2014), pp. 758–762.

[3] Gemma A Calvert and Michael J Brammer. “Predicting consumer behavior: using novel mind-reading approaches”. In: *IEEE pulse* 3.3 (2012), pp. 38–41.

[4] Dan Ariely and Gregory S Berns. “Neuromarketing: the hope and hype of neuroimaging in business”. In: *Nature Reviews Neuroscience* 11.4 (2010), p. 284.

[5] Jonathan Peirce et al. “PsychoPy2: Experiments in behavior made easy”. In: *Behavior research methods* (2019), pp. 1–9.

[6] Alexandre Gramfort et al. “MNE software for processing MEG and EEG data”. In: *Neuroimage* 86 (2014), pp. 446–460.

[7] Anna M Beres. “Time is of the essence: A review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research”. In: *Applied Psychophysiology and Biofeedback* 42.4 (2017), pp. 247–255.

[8] Ellen F Lau, Colin Phillips, and David Poeppel. “A cortical network for semantics:(de) constructing the N400”. In: *Nature Reviews Neuroscience* 9.12 (2008), p. 920.

[9] Marta Kutas and Kara D Federmeier. “Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP)”. In: *Annual Review of Psychology* 62 (2011), pp. 621–647.

[10] Eric Maris and Robert Oostenveld. “Nonparametric statistical testing of EEG- and MEG-data”. In: *Journal of Neuroscience Methods* 164.1 (2007), pp. 177–190.