Journal of Computing and Information Technology 06/2002; 10(2): 103-113

Experimental Methodology for the Evaluation of the 3D Visualization of Quantitative Information: a Case Study Concerning SEEG Information

Beatriz Sousa Santos[1,2], Jean-Louis Dillenseger[3], Carlos Ferreira[4,5]

1-Dept. Electrónica e Telecomunicações, Univ. Aveiro, Portugal

2- Instituto de Engenharia Electrónica e Telemática de Aveiro, Portugal

3- Laboratoire du Traitement du Signal et de l'Image, Univ. de Rennes I, France

4- Secção Autónoma de Gestão e Engenharia Industrial, Univ. Aveiro, Portugal

5- Centro de Investigação Operacional, Univ. de Lisboa, Portugal

ABSTRACT

The visual analysis of Stereoeletroencephalographic (SEEG) signals in their anatomical context is aimed at the understanding of the spatio-temporal dynamics of epileptic processes. The magnitude of these signals may be encoded by graphical glyphs, having a direct impact on the perception of the values. This problem has motivated an evaluation of the quantitative visualization of these signals, specifically to the influence of the coding scheme of the glyphs on the understanding and the analysis of the signals.

This work describes an experiment conducted with human observers in order to evaluate three different coding schemes used to visualize the magnitude of SEEG signals in their 3D anatomical context. In advance we had no clue to which of these schemes would provide better performances to the human observers, since we found in the literature theories that could support different answers. Through this experiment we intended to study if any of these coding schemes allows better performances in two aspects: accuracy and speed.

A protocol has been developed in order to measure these aspects. The results that will be presented in this work were obtained from 40 human observers. The comparison between the three coding schemes has first been performed through an Exploratory Data Analysis (EDA). The statistical significance of this comparison has then been established using nonparametric methods. The influence on the observers' performance of some other factors has also been investigated.

Keywords:

Evaluation – Visualization – Empirical evaluation – Epilepsy

I. INTRODUCTION

Epilepsy is the result of abnormal brain electric activities that mainly appear as synchronous (paroxystic) discharges within large populations of neurons belonging to brain structures implied during seizures. Investigation methods used in epileptology are aimed at defining and understanding the organization of the epileptogenic zone (from the areas originating the discharges to those secondarily affected by their propagation). Among these methods, Stereoeletroencephalography (SEEG) **(**Bancaud et al., 1970) provides signals, recorded with intra-cerebral electrodes, which bring major information on the dynamics of processes inside the brain structures. The visual analysis of SEEG signals is aimed at understanding the spatio-temporal dynamics of epileptic processes. One solution for this spatio-temporal analysis can be given by the fusion of the signals and the anatomy on a same referential (Rocha et al.,1996) (figure 1). In this representation the SEEG signals are displayed on the 3D location of their measurement points (e.g. on plots along depth electrodes), where the external anatomy and the position of the depth electrodes give the spatial reference. The 3D characteristic of this representation induced the choice of some 3D glyphs to encode the SEEG value and all these structures are displayed with the help of 3D-computer graphics. This encoding will have a direct impact on the perception of the value.

According to Tufte (Tufte,1987), there are considerable ambiguities in how people perceive a 2D surface (and perhaps moreover in 3D) and convert it in a one-dimensional number; changes in physical area do not reliably produce proportional changes in perceived area, thus using two dimensions to show one-dimensional data is a weak technique. However, according to Bertin (Bertin,1998), the stimulus to the perception of size is the variation of surface. Due to these different theories and also because the way human observers perceive quantitative information from a visualization can be very important in scenarios where they have to make decisions based on the visual analysis of very large data sets, as the diagnosis in epilepsy, we have considered this problem as deserving further investigation through a controlled experiment.

The goal of the present work will be the evaluation of this encoding technique (3D glyphs) in regard to the objectives of the medical task. The evaluation of procedures in scientific visualization is generally known as a difficult task due to the large amount of aspects that interfere on the final interpretation (Nielson,1996). So, for a simplification purpose, we have constrained our study to the evaluation of the technique for the quantitative visualization of the signals and more especially to the influence of the coding scale of the glyphs on the quantitative understanding and the analysis of the signals.

In this paper, we will first introduce the objectives and the specification of the planed evaluation. An experimental protocol for the evaluation of the performance in accuracy and speed of three different coding schemes, for representing the magnitude of the signals, will then be described. As the first statistical results showed that the assumption of normality of the data had to be rejected, some rank-based nonparametric statistical procedures (applied on the data extracted from this experiment) will finally be presented and discussed.

II. Objectives and context of the experiment

The objectives of the experiment presented in this work are related to the evaluation of the performance of human observers in extracting quantitative information from the visual representation of the signal coded through a glyph without the presence of the scale.

The choice of the glyph and the visual variables (size, color, texture, form, etc.) coding quantitatively the SEEG value was a critical point of the experiment. Bertin (1998) suggests that only the variation of a size may be used to encode quantitative data, thus in our application the chosen glyph was a sphere (its pattern is invariant with the viewing point) having a size proportional to the magnitude of the signals to encode. The coding scheme of the glyphs must state the same organization level as the values. The 3D characteristics of the sphere allows three types of proportionality between the magnitude of the signal and the size of the sphere (figure 2):

#1 - magnitude $\propto$ Radius of the sphere (R);

#2 - magnitude $\propto$ Projected surface of the sphere (πR^2);

#3 - magnitude $\propto$ Volume of the sphere ($(4/3)\pi R^3$);

where R is the radius of the sphere.

Through this experiment we intend to study specifically if any of these coding schemes (#1, #2, #3) allows better performances for the human observers using them, in the referred conditions. We have considered that a coding scheme can be compared to the others in two aspects: accuracy and speed. A coding scheme will be considered better than the others if it allows a more accurate and/or faster “measure” of the magnitude of the signal in a statistically significant way.

In order to restrain the effect of external factors, our evaluation must be very strongly delimited in time, space, visual variables and amount of information (Bertin,1998; Tufte,1987):

- Time - time variation of the signals is of great importance for the analysis; however this variation would hold most of the observer's attention (the human brain is highly specialized in motion understanding). Moreover, the time variable in animations is not perceived as linear. To avoid these facts, our evaluation is performed on static frames;
- Space - SEEG signals are displayed on the 3D location of their measurement points; the external anatomy and the depiction of the depth electrodes give the anatomical location of these points. But, as shown in figure 1, the orientation of electrodes can induce some perspective or superposition artifacts; thus on our evaluation we have used parallel projection and a viewing direction perpendicular to the electrodes (figures 3 and 4);
- Visual variables – only the size of the spheres encod the values. The other visual variables (color, texture, shape, etc.) remain neutral and constant throughout the experiment;
- Amount of information - this factor interferes in two aspects, on the number of signals displayed simultaneously and on the quantification of these signals. To limit the first factor, only a few spheres (4 at most) are displayed on each evaluation step so the overlapping of spheres is limited. The choice of the quantification of the scale is more crucial. The scale must present some perception step between the values but must still hold a quantitative nature. We chose to quantify the scale of the values to integers in the interval [0,10]. The range of the coding scheme is kept similar for all coding schemes: from 0 to a maximum radius encoding value 10 (see figure 2).

III. The experiment

When we plan an experiment (Cochran,1983; Preece et al.,1994; Dix et al.,1998), it is necessary to think about the purpose of the experiment, the subjects, the methodology, what specific data should be collected and what statistical tests should be applied to the collected data.

A. Variables.

In order to evaluate the accuracy and speed of the coding schemes, we have chosen concerning the:

- *independent variable*: the coding scheme (#1, #2 and #3) is the main independent variable;

- *dependent variables*: the accuracy is measured by the perceived magnitude error and the speed by the decision time (approximated by the reaction time).

It would be desirable that all measured changes in the dependent variables were due to an experimental effect (i.e., due to a change in the independent variable), however other variables can also have some influence. We were able to identify a few such variables (gender, age, profession, nationality, etc. defining an observer profile) and have used them to characterize each subject.

B. Subjects

One of the main goals of our experiment is to establish which coding scheme should be used to present information from SEEG signal to neurophysiologists which need to used it; however the experimentation presented here was mainly performed in order to establish a first basic experimental protocol and statistical methods. On the other hand we intend also to study the more general issue of conveying quantitative information through this kind of visual coding. We have chosen subjects having education and scientific levels similar to neurophysiologists. Thus, most of the subjects that have participated in the experiment were PhDs or post-graduated students in sciences or engineering.

C. Experimental methodology

A repeated measures design was used (Preece et al.,1994); all the subjects appear in all the experimental conditions. There are no problems of subject allocation, however we must be careful with the order in which the subjects perform in these conditions (Keselman and Keselman,1993). In our case three experimental conditions exist, corresponding to the tasks performed using the three coding schemes (we have called a task, the evaluation of the sequence of glyphs encoded by the same coding scheme). As experimental methodology, a within-group methodology was chosen (Dix et al.,1998; Shneidermen,1998); the independent variable coding scheme was placed within-groups and all the observers performed the same task under all the conditions. In order to compensate for possible influence of certain side variables (as learning effects and interference due to the use of different schemes, nervous behavior in the first task or fatigue in the last task), the six possible sequences of tasks were used. The assignment of sequences to observers was performed randomly. For each coding scheme three different images were used (i.e. nine images were used for the complete experiment).

In order to minimize the influence of some external conditions we chose to keep them from changing. Thus:

- all the test images were generated in the same way, varying only the coding scheme;
- the point of view was chosen so that the spheres were at equal distance from the observer;

- during the application of the experiment, the viewing conditions have been as similar as possible for all the observers (screen, viewing distance, ambient light, etc.).

D. Protocol

A S/W package, for Windows platforms, was developed to allow an easy application of the protocol defined for the experiment. All observers received a simple explanation about the context and aims of the experiment and what was expected from them. After they were fully informed and have agreed in proceeding, they were asked the information needed to define their profiles and then the protocol started.

The protocol is divided in three similar parts (one for each coding scheme) where observers exercise themselves in using a certain coding scheme before observing all the images corresponding to the same coding scheme. This training is interactive, consists of two screens (similar to the one shown in figure 3) and ends only when the observer decides he/she is ready to proceed.

After training for each coding scheme, the observers are shown three different images containing a certain number of spheres and are asked to evaluate the magnitude of one of the spheres (as shown in figure 4). The accuracy of the answer and its speed are registered in a file.

Before applying the experiment to the complete subject population, a pilot study was performed with 8 observers having a similar profile as the observer population (Cochran,1983; Preece et al.,1994). This study allowed testing the protocol and resulted in some adjustments.

E. Collected data

For each observer, the following variables were used to establish the observer profile: age (<25; [25,55]; >55), nationality, gender, familiarity with 3D video games or 3D synthetic images; profession; number of years of specialty and familiarity with 3D medical imaging (for medical doctors).

For each magnitude assessment, by an observer, the following variables were measured:

i) Main output (dependent) variables:

- the accuracy is estimated by the error in perception of the value, E = real magnitude - perceived magnitude. Note that $E \in Z$;
- the speed is estimated by the perception time, T = time the observer takes to evaluate the magnitude (approximated by the reaction time). T is expressed in seconds.

In the rest of the paper we will call E_n and T_n, the variables of respectively the errors and the perception time for coding scheme #n.

ii) Side variables concerning the images: Sequence of coding schemes; order of the task (first, second or third); number of spheres; spheres are occluding or not; size of the evaluated spheres.

Variables concerning the spheres were collected in order to evaluate a possible influence of the number and type of displayed spheres on the perception of magnitude. The order of the task is expected to give some information on the effects of boredom, fatigue or learning. Familiarity with 3D images is collected in order to establish a possible non-formal training in 3D perception.

During the experiment the experimenter has observed the behavior of the observers and, after the experiment, each observer was asked if he or she had any preferences about the coding schemes and why.

IV. RESULTS

The results presented in this work were obtained applying this first experimental set to 40 human observers. These observers exhibited a relative homogeneous profile (faculty staff and post-graduation students; 25% of females). The aim of this first step was mainly to verify the protocol and the statistical analysis. Nevertheless, we expected it would produce some results about the performances of observers using the three coding schemes, since the data obtained from such a number of observers could already be considered as statistically significant.

A statistical analysis has been performed on the recorded data in order to evaluate the part played by the coding scheme on the performance of the observers. But, after the application of the protocol to the observers, some of them have mentioned that their practice, the size and the location of the spheres displayed on each image, as well as their own preferences to a specific coding scheme, could have influenced their performances. Thus, we have carried out some further statistical analysis on the data corresponding to other collected variables.

In a first stage, we tested the assumption of normality of the distributions of the collected data. The performed goodness-of-fit test rejected this assumption, so we have used rank-based nonparametric hypothesis and procedures for the data analysis. All these statistical analysis were performed using the commercial s/w package MINITAB (Minitab, 1998).

A. Evaluation of the coding schemes

The framework of the statistical analysis used on the obtained data corresponding to the main output variables (error and time values) was the following:

- the first analysis was an Exploratory Data Analysis (EDA) (Hoaglin et al.,1983) in order to get an overview of certain data characteristics (such as ranges, asymmetries, existence of outliers, etc.). It allowed a preliminary comparison among the three coding schemes and helped on the choice of other statistical techniques to be used to further analyze the obtained data. It allowed also to get an idea of the influence of other factors on the observers' performance;
- the equality or differences between the schemes has then been statistically tested. Since the equality hypotheses have been rejected, we investigated the data to determine the cause;

A.1. Overview of the Data corresponding to the output variables

To analyze the overall behavior of the errors and times per coding scheme, a well-known EDA technique, the box plot (Hoaglin et al.,1983), was used. These plots display the maximum and minimum values as well as a central box indicating the location of the 50% central values (between the lower and upper quartile, in our case the 39^{th} and the 90^{th} elements from the sequence of 120 samples in ascending order).

The box plots obtained for the variables E_1, E_2, E_3, T_1, T_2, T_3 are shown in figures 5, 6.

The errors corresponding to coding schemes #1 and #2 seem very similar. They have the same minimum, median and 50% of central values (between 0 and 1), the only difference being the existence of a maximum value (3) for coding scheme #2 (considered as an outlier). On the other hand, coding scheme #3, presents a larger range; it must be noted that the minimum values for coding schemes #1 and #2 are considered outliers (just one and two (-2) values, respectively) and this is not the case for coding scheme #3 (with seven (-2) values). Moreover, the 50% central values of coding scheme #3 are spread between –1 and 1. Note that the median and the lower quartile values are coincident both for E_1 and E_2 (the 30th and the 60th values are zero), which is not the case for E_3. Analyzing the raw data, it can be verified that 66, 62 and 42 from 120 values for E_1, E_2 and E_3, respectively, are equal to zero. Finally the box plot indicates a large asymmetry concerning the distribution of E_1 and E_2 (a first sign for non normality).

The minimum, maximum and median values of the perception times obtained for coding scheme #1 are smaller than for any of the other coding schemes; this is also the case for the 50% of central values. However the great difference observed in the ranges of the three coding schemes may not be so significant since the (worst) maximum values of coding schemes #2 and #3 are considered outliers.

A.2. Comparison among schemes

Concerning accuracy.

The comparison can be stated as following: if the proportions of failures (among the categories) and successes in the perceived magnitude is equal when using any scheme, then it is possible to consider that scheme #1 is as good as scheme #2 and as scheme #3. Statistically this can be stated as the null hypothesis: H_{0e}: $p_{e\#1}=p_{e\#2}=p_{e\#3}$ *versus* H_{1e}: $p_{e\#1}\neq p_{e\#2}\neq p_{e\#3}$ (where $p_{e\#n}$ is the proportion of the several categories of failures and successes for coding scheme #n) and tested using the χ^2 test for homogeneity on contingency tables, due the amount of ties between error values (see table II) (Everitt, 1992).

Table I shows the contingency table of the failures and successes of the perceived magnitude when using any scheme. Since the statistic $X^2=11{,}06 > \chi^2_{(2);0.05}=5{,}99$, H_{0e} was rejected with a level of significance $\alpha=5\%$.

If we look more precisely on the distributions by investigating the equality proportions of failures among the categories, i.e. –3, -2, -1, 1, 2 and 3 and successes of the perceived magnitude when using any scheme (see the contingency table on Table II), H_{0e} is still rejected with an α of 5%, since the statistic $X^2=28{,}44>\chi^2_{(12);0.05}=21{,}06$.

The analysis of table I shows that c.s.#3 has more failures than c.s.#2 and c.s.#1, and table II shows that this is particularly relevant concerning errors of type (-2) and (-1).

Concerning speed

The comparison can be stated as following: if the median (me) of perception times is equal when using any scheme, then it is possible to consider that scheme #1 is as good as scheme #2 and as scheme #3 in speed. This can also be stated as the null hypothesis: H_{0t}: $me_{t\#1}=me_{t\#2}=me_{t\#3}$ *versus* H_{1t}: $me_{t\#1}\neq me_{t\#2}\neq me_{t\#3}$, and tested using the Kruskal-Wallis test (a nonparametric analog for the one-way ANOVA) (Gibbons, 1997). It requires only that the underlying distribution of each of the populations be identical in shape and is primarily designed to detect differences in “location” among the populations.

In our case (Table III), H_{0t} was rejected with an α of 5% since the test statistic $H=13{,}32>\chi^2_{(2);0.05}=5{,}99$.

We used a pairwise comparison method based on the average ranks (Freund et al., 1993) to determine where the differences are. We inferred that the locations of the distributions differ if the mean rank difference between two coding schemes exceeds the critical value S, which is S=25,9 for this data set. Analyzing table III we can say that there is insufficient evidence to declare different locations between coding schemes #1 and #2 (15,7<25,9). The locations of the distributions of times for coding schemes #1 and #3 (47,9>25,9) and for coding schemes #2 and #3 (32,3>25,9) may be declared different.

B. Observers' preferences

The relation between the preferences expressed by the observers concerning coding scheme and their performances with the preferred coding scheme was investigated. Thirty-five from the forty observers have expressed a preference by a certain coding scheme; from these, twenty have performed better when using the preferred coding scheme (the other fifteen have not). To study if there is an association between the performance and the preference expressed by the observer we have performed a Chi-square test on a 3x2 contingency table crossing preference of the observer on coding schemes #1,2,3 with best performance or not (Everitt, 1992). Since we have obtained an observed value for the statistic of X^2=0,22 <$\chi^2_{(2);0.05}$=5,99, the null hypothesis of independence between preference and performance is not rejected for an α of 5%.
This means that the proportion of best performances, for this data set, did not differ depending on the preference, which illustrates the fact, known to the Human Computer Interaction community, that we cannot rely solely on users' preferences if the goal is to improve the performance (Nielsen, 1993).

C. Influence of other factors

According the comments of the observers, some parameters could have influenced their performance. More precisely the practice of the observers and the size of the spheres have been mentioned frequently.
In the following analysis, we defined the observers' performance by only the occurrence of errors (failures) or not (successes) regardless to the coding schemes; the failures, successes and times have been accumulated for all the observations in the same circumstances. When a different number of observations was registered, we have used relative frequencies of successes, failures and time rates (obtained dividing the former values by the number of observations). We have also searched for statistical support to establish if the observed proportions of failures and successes could be due to random causes (in which case the suspected influence was not confirmed), or not. With this purpose, equality of proportions tests were performed on contingency tables (Everitt,1992) to analyze samples drawn from different populations to see if these populations have the same proportions of elements in a certain category. More specifically, to test the null hypothesis that the probability of some specified event is the same for each population.

C.1. Practice of the observers

To study specifically the influence of the practice they have accumulated along the experiment, we have computed the total number of errors made by all the 40 observers, for the first, second and third tasks they have performed. Table IV shows the corresponding number of failures and successes, as well as the accumulated times.

Analyzing table IV, we can observe that both the number of error and the accumulated times, obtained by all the observers decrease from the first to the second, to the third tasks, i.e., the observers increase their performances (specially concerning the times). This seems to imply that they are becoming more familiar with the application and/or the type of task they have to perform.

However, the equality of proportions test asserts the independence of the observers' performances from the order of the task. The differences among performances obtained in the three tasks seem not to be due to a learning effect but to random causes. This means that generally we have to be careful with informal analysis of data.

C.2. Size of the evaluated spheres

To study the influence of the size on the observers' performance, we have divided the 11 possible values of the spheres into three classes, small, medium and large, according to the signal values they encode (≤3, between 4 and 6, or ≥7). Table V displays the number of failures and successes, their corresponding rates and number N of evaluated spheres. The overall performance of the observers in the evaluation of small and large spheres is better than in the evaluation of medium spheres. This result had already been anticipated from the observation of the observers' behavior and remarks during the experiment and has been reinforced by a test of the proportion of failures and successes performed on a contingency table, which rejected the equality of proportions. This confirms the existence of a dependence of the observers' performances from the size of the sphere.

V. DISCUSSION

After applying the experiment to the 40 observers, critical reviews of the procedure and the results were performed. The review of the experimental procedure was based on the statistical results obtained and on the experience the experimenter herself had accumulated from observing the subjects performing the experiment as well as in interviewing them after they have completed the tasks. This review considered the following issues:

- observer preparation - the instructions given to the observers, as well as the amount of task practice seemed adequate;
- task duration - the duration of the tasks seemed too small when compared to the preparation time; thus it seemed possible to increase significantly the length of the tasks, without the risk of producing fatigue or boredom in the observers;
- impact of independent variables - some of the users referred that the sequence in which the coding schemes were used could make a difference, possibly introducing learning or confusion effects. This was not confirmed by the statistical analysis.
- viewpoint - some observers have suggested that the chosen point of view (parallel projection and a viewing direction perpendicular to the electrodes) bring back the experiment to a 2D case, which could have an influence on the results. A new experiment with a more general point of view could solve this doubt.

To establish exactly what has been found out by this experiment, the critical review of the results was performed and the following points were considered relevant:

- size effect - the difference between coding scheme #3 and the others seems rather significant also from a practical perspective. The observers not only make more errors with coding scheme #3 but also can make much larger errors; this can be important;
- observers' preference – in several cases the observers claimed some preferences or made remarks (e.g. poorer performances when occluded spheres, learning effects, better performances using coding scheme #1), which were not validated by the tests. This reinforces the utility of a real formal evaluation of each human computer interaction, and not just a recording of the users' comments;
- consistency between dependent variables - the results seem consistent in that one of the coding schemes was found worst both on accuracy and speed. However, some inconsistency may be related to the fact that many observers clearly preferred coding scheme #1, which was not found as significantly different from coding scheme #2. This preference is contradictory to the assessment of Bertin (1998) who suggested that an observer is more sensible to surfaces for the extraction of quantitative data. That preference can perhaps be due to the specific way coding schemes are shown to the observers (as shown is figure 2);

- the fact that coding scheme #1 was not discriminated from coding scheme #2, through the previous statistical analysis, could also be explained by the small sample of observations (three images per coding scheme), the observers' profile and the number of independent variables.

VI. CONCLUSIONS

This work described an experiment (the first from a set of experiments we intend to perform) integrating a study concerned with the evaluation of the influence of glyphs on the understanding and analysis of SEEG signals visualized in anatomical context. The aim of this case study was to experimentally assess the effect of the coding scheme chosen to map a variable on the observers' capacities of perceiving fast and accurately quantitative information. This effect may be very important in scenarios where observers have to take important decisions based on the visual analysis of large data sets, as it is the case of neurologists using SEEG visualizations to make epilepsy diagnosis. For this reason, we chose to make use of a statistically supported methodology.

An evaluation methodology specific for our case has been designed and an experimental protocol has been implemented. The obtained results seem to indicate that one of the three tested coding schemes is worse than the others, both on accuracy and speed. The study performed about the influence, on the observers' performance, of other factors (as task order and size of the spheres) has indicated that the size of the evaluated spheres has possibly an important impact; moreover it is a good illustration of the need of formal data analysis since the informal analysis of the influence of the learning effect suggested a different conclusion from the one obtained using statistical tests. The study performed on the influence of observer's preferences on their own performances has confirmed the generally known fact that if we want to improve the users' performances using a user interface (from which a visualization is a particular case) we cannot solely rely on users' preferences, we have to apply other methods, namely controlled experiments and formal statistical data analysis.

The experiment described allowed to establish a basic protocol and provided several important hints on how to perform further experiments and analyse the obtained results. As a consequence, another experiment, using basically the same protocol is currently being performed. The new protocol differs from the previous one, mainly in the number of images shown per coding scheme. This second experiment will involve a greater number of observers having different user profiles (including neurophysiologists and observers with formal 3D training) and the obtained results are expected to allow the clarification of the results obtained until now. We also expect this experiment to allow establish relationships among some of the characteristics of the observers' profiles and the performances with some of the coding schemes.

REFERENCES

Bertin J., *Sémiologie graphique, les diagrammes, les reseaux, les cartes.* 3me ed., Éditions de l'École des Hautes Études en Sciences Sociales, Paris, 1998 (can be found in english: *Semiology of graphics*. The University of Madison Press, Madison, 1983).

Bancaud J., R. Angelergues, C. Bernouilli, A. Bonis, M. Bordas-Ferrer, M. Bresson, P. Buser, L. Covello, P. Morel, G. Szikla, A. Takeda, J. Talairach, Functional stereotaxic exploration (SEEG) of epilepsy. *Electroencephalogr. Clin. Neurophysiol*. 28(1) (1970), 85-86.

Cochran W. G., *Planning and analysis of observational studies*. John Wiley, 1983.

Dix A., J. Finley, G. Abowd, B. Russell, *Human Computer Interaction*. 2nd Ed., Prentice Hall, 1998.

Everitt B. S., *The analysis of contingency tables*. 2nd ed. Chapman and Hall, 1992.

Freund R., W. Wilson , *Statistical methods*. Academic Press, 1993.

Gibbons J., *Nonparametric methods for quantitative analysis*. 3rd Ed, American Sciences Press, 1997.

Hoaglin D., F. Mosteller, J. Tukey, *Understanding robust and exploratory data analysis*. John Wiley and Sons, 1983.

Keselman H., J. Keselman, Analysis of repeated measurements. Presented at *Applied Analysis of Variance in Behavioral Science*, (1993), Marcel Dekker, 105-146.

Minitab User's Guide 2: Data analysis and Quality Tools. MINITAB Inc., Statistical Software, 1998

Nielsen J., *Usability Engineering*p AP Professional, Academic Press, 1993.

Nielson G. M., Theme issue introduction: Challenges in visualization research. *IEEE trans. On Visualization and Computer Graphics*, 2 (2), (1996).

Preece J., Y. Rogers, H. Sharp, D. Benyon, S. Holland, T. Carey, *Human Computer Interaction*. Addison Wesley, 1994.

Rocha C. J.-L. Dillenseger, J.-L. Coatrieux, Multi-array EEG signals mapped with 3-Dimensional images for clinical epilepsy studies. *Lecture Notes in Computer Science 1131, Visualization in Biomedical Computing*, Höhne and Kikinis Eds., Springer Verlag, (1996), 467-476.

Shneidermen B., *Designing the user interface, strategies for effective Human-Computer Interaction.* 3rd ed., Addison Wesley, 1998.

Tufte E., *Visual display of quantitative information*. Graphics Press, 1987.

Tufte E., *Envisioning information*. Graphics Press, 1990.

CAPTIONS

Figure 1- SEEG in anatomical context.

Figure 2 - Three coding schemes used in the experiment.

Figure 3 - Interactive training for one of the coding schemes.

Figure 4 - What is the magnitude represented by this sphere?

Figure 5- Box plot for the errors

Figure 6- Box plot for times (T_1, T_2, T_3) corresponding to coding schemes #1, #2 and #3

Table I- Contingency table used to test the differences among coding schemes (#1, #2, #3) concerning the Error (E_1, E_2, E_3).

Table II - Contingency table used to test the differences among coding schemes (#1, #2, #3) concerning the Error (E_1, E_2, E_3).

Table III - Kruskal-Wallis (ANOVA by ranks) to test the differences among coding schemes concerning the Time (T_1, T_2, T_3).

Table IV- Number of failures, successes and accumulated times for the 40 observers corresponding to the first, second and third tasks.

Table V - Total number of failures and successes obtained for small, medium or large sphere.

FIGURES AND TABLES

Figure 1

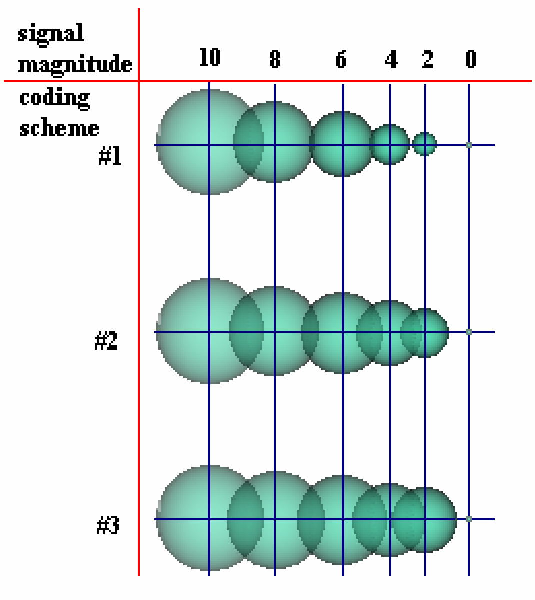

Figure 2

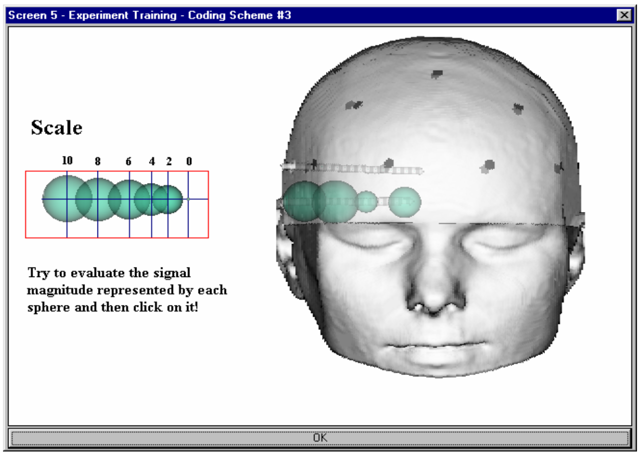

Figure 3

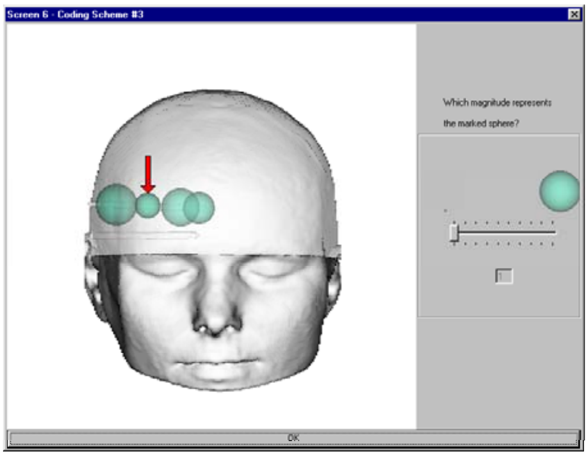

Figure 4

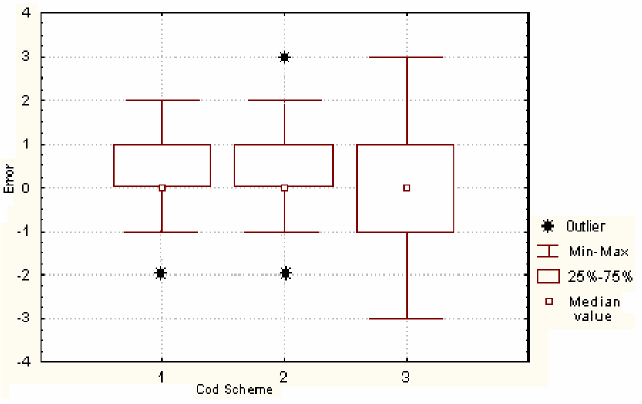

Figure 5

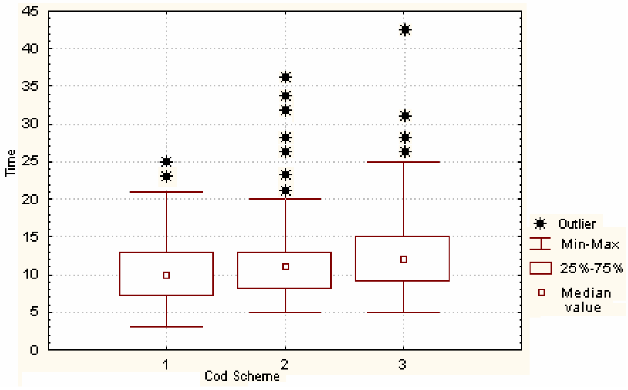

Figure 6

	c.s. #1	c.s. #2	c.s. #3	Total
Failure	54	58	78	190
Success	66	62	42	170
Total	120	120	120	**360**

Table I

	-3	-2	-1	0	1	2	3	Total
c.s. #1	0	1	7	66	35	11	0	**120**
c.s. #2	0	2	13	62	24	16	3	**120**
c.s. #3	1	7	22	42	32	13	3	**120**
Total	**1**	**10**	**42**	**170**	**91**	**40**	**6**	**360**

Table II

	N obs	Sum of ranks	Mean rank	Mean rank difference between		
				c.s.#1 and c.s.#2	**c.s.#1 and c.s.#3**	**c.s.#2 and c.s.#3**
c.s. #1	120	19117,0	159,3	15,7	47,9	32,3
c.s. #2	120	20995,5	175,0			
c.s. #3	120	24867,5	207,2			

Table III

	Failures (errors)	Successes (no errors)	Acc. Times (s)
First task	67	53	1619
Second task	64	56	1327
Third task	59	61	1244

Table IV

Size of the sphere	Failures	Successes	Num. Obs (N)	Failures/N	Successes/N
Small	49	79	128	0,38	0,62
Medium	81	38	119	0,68	0,32
Large	60	53	113	0,53	0,47

Table V