sensors

MDPI

Article

The Sample Size Matters: To What Extent the Participant Reduction Affects the Outcomes of a Neuroscientific Research. A Case-Study in Neuromarketing Field

Alessia Vozzi [1,2,*], Vincenzo Ronca [1,2], Pietro Aricò [2,3], Gianluca Borghini [2,3], Nicolina Sciaraffa [2,3], Patrizia Cherubino [2,3], Arianna Trettel [2], Fabio Babiloni [2,3,4] and Gianluca Di Flumeri [2,3]

[1] Department of Anatomical, Histological, Forensic & Orthopedic Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy; vincenzo.ronca@uniroma1.it

[2] BrainSigns srl, Via Lungotevere Michelangelo, 9, 00192 Rome, Italy; pietro.arico@uniroma1.it (P.A.); gianluca.borghini@uniroma1.it (G.B.); nicolina.sciaraffa@uniroma1.it (N.S.); patrizia.cherubino@uniroma1.it (P.C.); arianna.trettel@brainsigns.com (A.T.); fabio.babiloni@uniroma1.it (F.B.); gianluca.diflumeri@uniroma1.it (G.D.F.)

[3] Department of Molecular Medicine, Sapienza University of Rome, Viale Regina Elena, 291, 00161 Rome, Italy

[4] Department of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China

[*] Correspondence: alessia.vozzi@uniroma1.it

check for updates

Citation: Vozzi, A.; Ronca, V.; Aricò, P.; Borghini, G.; Sciaraffa, N.; Cherubino, P.; Trettel, A.; Babiloni, F.; Di Flumeri, G. The Sample Size Matters: To What Extent the Participant Reduction Affects the Outcomes of a Neuroscientific Research. A Case-Study in Neuromarketing Field. *Sensors* **2021**, *21*, 6088. https://doi.org/10.3390/s21186088

Academic Editors: Filippo Zappasodi and Yvonne Tran

Received: 29 May 2021
Accepted: 7 September 2021
Published: 10 September 2021

Abstract: The sample size is a crucial concern in scientific research and even more in behavioural neurosciences, where besides the best practice it is not always possible to reach large experimental samples. In this study we investigated how the outcomes of research change in response to sample size reduction. Three indices computed during a task involving the observations of four videos were considered in the analysis, two related to the brain electroencephalographic (EEG) activity and one to autonomic physiological measures, i.e., heart rate and skin conductance. The modifications of these indices were investigated considering five subgroups of sample size (32, 28, 24, 20, 16), each subgroup consisting of 630 different combinations made by bootstrapping n (n = sample size) out of 36 subjects, with respect to the total population (i.e., 36 subjects). The correlation analysis, the mean squared error (MSE), and the standard deviation (STD) of the indexes were studied at the participant reduction and three factors of influence were considered in the analysis: the type of index, the task, and its duration (time length). The findings showed a significant decrease of the correlation associated to the participant reduction as well as a significant increase of MSE and STD ($p < 0.05$). A threshold of subjects for which the outcomes remained significant and comparable was pointed out. The effects were to some extents sensitive to all the investigated variables, but the main effect was due to the task length. Therefore, the minimum threshold of subjects for which the outcomes were comparable increased at the reduction of the spot duration.

Keywords: applied neurosciences; EEG; HR; GSR; sample size; signal processing

1. Introduction

In the field of behavioural neurosciences, as well as in every scientific discipline, sample size definition is a crucial factor for ensuring the reliability and accuracy of the results obtained in a study, as well as the replicability of the study itself [1]. In fact, the statistical power and the reproducibility of a study are strictly related to the number of participants included in the analysis [2]. In order to prevent and mitigate consequences of a poor statistical power, it is generally recommended to estimate the proper sample size. To this end, different studies recently tried to propose methods to easily calculate the required sample size while designing the experimental protocol [3–8]. Additionally, sometimes it is not only strictly a problem of sample size, intended as the number of participants, but also the concerns of trial numbers, i.e., the number of tasks and repetitions should be carefully considered [9]. Backe and colleagues pointed out the relation between the number of trials necessary to perform studies involving human participants and the sample size. They proposed a tool based on power contours for calculating the optimal combination of trials and number of participants [10]. Additionally, Guttmann-Flury [11] pointed out a method for a priori sample size determination based on Montecarlo simulations on Electroencephalographic (EEG) data. Therefore, the inclusion of an adequate number of participants in neuroscientific studies is an important issue, and some authors reported alarming data about the potential risks of employing a small sample size [1]. This is the case of Button and colleagues [2], who claimed that many of the works published in neuroscientific domain are not reproducible due to their small sample size and the low statistical power associated with false-positive effects counted in the results. The reduction of the sample size is denounced as a deliberate action, driven by the aim of researcher or organisations which want to increase the prestige associated with the publishing of their works at any cost [12]. A smaller sample size, in these cases, is chosen to undermine the statistical power of the study, enhancing the presence of false-positive effects which endorse the desired results.

Though this behaviour must be countered, there are cases in which the reduction of the sample size is necessary for organisational and logistic needs. In fact, thanks to the increasing shared knowledge and recent scientific evidence produced by the neuroscientific community, behavioural neuroscience is experiencing an exponential growth in terms of applied research [13–15]. Recently, new terms have been coined, such as Neuroergonomics, Neuromarketing, Neuroaesthetics, and others, to include the new neuroscientific disciplines that aim at investigating neurophysiological correlates of humans' behaviour at work [16–18], including driving a car [19–24], watching television [25–27], having a sensorial experience [28–30], and in any other activity of daily life. In the context of applied neurosciences, when physiological signals are acquired in real environments, several factors work against the ease of participant recruitment. The whole experimental session of each subject requires time, and the operators have to welcome the participant, explain the protocol, ask for informed consent, place the devices in the correct way, ensure the quality of the signals, and finally start the experimental tasks during which biosignals are recorded. Therefore, participants should account for taking one or two hours off from their working day for taking part in the experiment. At the same time, expert operators should be available to spend several days to perform the experiment, usually outside their working place (i.e., outside the laboratories). Furthermore, if the study is conducted in real environments, out from the labs, it is necessary to concentrate all the recordings in few days because generally the facilities that host the experiments are not available for long. Therefore, the number of slots at disposal to screen each participant is not so high. In the last period, moreover, the on-going Coronavirus Disease 2019 (COVID-19) pandemic [31] made participant recruitment even more difficult. During the study, both participants and operators take the risk of reaching the place in which the experiments are performed, they must carefully observe all the sanitising practices, avoid contacts, and maintain social distancing. The time spent on respecting these procedures renders the duration of the protocol even longer than usual. The reduction of the sample size, in these cases, would allow for the performing of experiments without compromising the experimental protocols and ensuring at the same time the good sanitizing practices and any other health preventive measure.

In this context, premising that the best practise is to include an optimal sample size, it is interesting to explore what error affects the outcomes of a study when researchers rely on a lower number of subjects. In other words, in those cases when for different reasons it is not possible to enlarge the sample size anymore, the researchers should be at least aware of the potential error affecting the results in order to interpret them properly. With this aim, the present study took into consideration a case study coming from the neuromarketing domain. It was analysed through an existing dataset of thirty-six participants while performing a neuromarketing experimental protocol in order to explore the modification of three neurophysiological indexes, frequently employed in applied neurosciences, when the participants decrease, pointing out the conditions for which the indices remain comparable, and quantifying the tendency of measurements variability and related error depending on the sample size. Five subgroups of subjects' numerousness (i.e., sample size) from the total sample of participants (36, hereafter 'Reference population') were considered: 32, 28, 24, 20, and 16 subjects respectively (sub-groups). In particular, each subgroup consisted of 630 different combinations (since 630 is the maximum number of different combinations of 32 elements from a set of 36) of the whole 'Reference population'. The outcomes associated with the entire population were considered as reference since 36 is within the range of sample sizes generally involved in this kind of study. This number was deemed appropriate to assume the results were reliable, i.e., representative of the theorical population, in line with the scientific literature [32]. These subgroups were chosen since: (i) in the literature there is a large consensus considering [16–32] the optimal range of participants in neuroimaging studies [33]; (ii) in order to perform Analysis of Variance, the number of cases (i.e., 630) has to be much higher than the number of groups (i.e., five). The experimental task consisted of watching four TV spots with different lengths (30, 30, 20, and 15 s) during which electroencephalographic (EEG) signal and autonomic measures, i.e., heart rate (HR) and galvanic skin response (GSR), were acquired. The data analysis explored the correlation and the error between the outcomes for the total sample size and the considered subgroups of subjects. Furthermore, the data distribution over the experimental sample was explored in terms of standard deviation of the indexes, for each second, at the participants' reduction. The hypothesis at the base of this study was to find a decrease of correlation as well as an increase of error and data dispersion moving from 32 to 16 subjects with respect to the total sample size (36 subjects). In addition, the data analysis was conducted considering three influencing factors: the kind of index, the task, and the time (i.e., task length), assuming that they could affect the outcoming modifications at the sample size reduction.

In conclusion, all the possible sub-sample sizes from 34 up to 10 were considered in order to model the decreasing curve of sample size-correlation relationship.

Before going ahead with the methods, it is important to clarify that the analysis was focused on investigating the alterations induced by sample size reduction on neurophysiological indicators averaged on the whole available sample (group analysis), without taking into consideration the sensitivity of the resulting indicators (i.e., the indicators computed on smaller samples) with respect to the phenomenon itself (i.e., if the sample size was reduced the probability of wrongly rejecting the null hypothesis—type I error—or wrongly accepting the null hypothesis—type II error—would increase or decrease). This concern is addressed in Section 4.

2. Materials and Methods

2.1. Experimental Sample and Design

The dataset analysed in this study included EEG, GSR, and PPG (photopletismographic) signals acquired from 36 subjects (37.39 ± 10.67 years old), gender balanced, recruited on a voluntary basis. These subjects signed the informed consent and the related information sheet, in which the study was explained, before participating in the experiment. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000, and it received the approval of Sapienza University of Rome ethical committee (nr. 2507/2020).

The task consisted of watching a TV video, almost 12 min in length, based on a documentary and with two commercial breaks in the middle. Such breaks consisted of twelve different TV commercials selected as experimental stimuli, six per break, presented in a randomised order. This means that the same spot was seen by the participants in different moments of the experience in order to avoid any habituation effect. Subjects were comfortably seated in front of the screen during the entire protocol. The documentary was chosen to be relaxing and emotionally neutral [34], in contrast with the high emotional and engaging contents of TV advertising, and it was used as individual baseline. The task was designed for neuromarketing research, and thus one of the new branches of applied neuroscience [26,35] as depicted in the introduction, with the aim of evaluating the response to some target advertising. However, for the purpose of the present study, only four commercials over twelve were considered in the analysis. Assuming that larger spot lengths would favour correlation analysis, we took the longer (t = 30 s) spots. Actually, 30 s is the most common time duration employed for TV commercial advertising (source: Nielsen Report, 2013 [36]). Then, we selected two shorter spots (20 and 15 s) taking care to select non-contiguous ones in order to have four spots distributed along the whole commercial break.

2.2. Neurophysiological Data Recording

The cerebral activity was recorded by means of a portable EEG system (BEmicro and Galileo software, EBneuro, Firenze, Italy). Ten Ag/AgCl electrodes were placed according to the 10–10 International System, at the following channels: Fp2, Fpz, Fp1, AF7, AF5, AF3, AFz, AF4, AF6, and AF8, all referred to the earlobes (REF channel). EEG activity was collected at a sampling rate of 256 Hz while the impedances were kept below 10 kΩ by employing conductive gel to adapt them. The raw signals were fully analysed offline through MATLAB software (MathWorks Inc., Natick, MA, USA). The power-supply interference was removed with a notch filter at 50Hz. Moreover, signals were bandpass filtered by a 5th-order Butterworth filter [2,30] Hz) in order to remove interferences in frequencies not of interest. Independent component analysis (ICA) was then performed, in particular by employing the SOBI algorithm [37], to remove Independent Components related to ocular and muscular activity. On average, 2 ± 1 components were manually selected and removed by the same neurophysiology expert. The reconstructed EEG signal was then segmented into 1-s-long epochs with 0.5 s of overlap in order to avoid any “boundary effect”, and three additional criteria for detecting artifacts based on the signals’ amplitude and trend [38,39] were applied in order to remove those portions of data still affected by artifacts that had not been corrected before.

At the beginning of the EEG recording, the participants were invited to close their eyes for 60 s. Such a time interval was then employed for the calculation of the individual alpha frequency (IAF), i.e., the peak of the signal power spectrum within the traditional alpha frequency range (7–12/13 Hz). According to Klimesch and Doppelmayr [40], the IAF was then used to define the individual EEG frequency bands of each participant as a function of the IAF itself ([IAF–x, IAF + x] Hz). In particular, for the purposes of this study two EEG frequency bands have been investigated, the whole alpha [IAF–4Hz, IAF + 2 Hz] and the upper alpha [IAF–2Hz, IAF + 2 Hz] [41]. Then, the EEG signals for each subject were filtered in these specific sub-bands and the global field power (GFP) [42] was computed in order to obtain a synthetic indicator of the synchronous de-/activation of brain cortical areas.

In terms of autonomic measures, Galvanic Skin Response (GSR) and Photopletismographic (PPG) signal were recorded with a Shimmer3+ (Shimmer Sensing, Ireland) device applied to the non-dominant hand. Two electrodes were placed on the palmar side of the middle phalanges of the second and third fingers according to scientific best practice [43], with the PPG sensor over the thumb. These signals were acquired with a sampling rate of 64 Hz. From the PPG signal, the Heart Rate (HR) was obtained by means of the Pan-Tompkins algorithm [44]. On the other side, the tonic component of the skin conductance (Skin Conductance Level, SCL) was estimated with the LEDAlab software and the Continuous Decomposition Analysis [45].

At this point, from the available biosignals pre-processed as described so far, three indices with different natures were considered in the analysis.

1. Index 1: general descriptor of the frontal cerebral activity. It was computed as the inverse of the upper-alpha GFP over all the frontal electrodes. It is generally related to the attentional processes [46,47];
2. Index 2: descriptor of the frontal alpha asymmetry [48]. It was computed as the difference between the GFP over right (Fp2, AF4, AF8) and left (Fp1, AF3, AF7)

electrodes. Such a "neurometric" is frequently employed in EEG studies [49,50], and it is known as the Approach-Withdrawal index [51];

3. Index 3: descriptor of the autonomic response, namely, the emotional index (EI), computed as the combination between the SCL and the HR measures, as described by Vecchiato and colleagues [52].

Each Index is a time-series with a time resolution of 1 s for each task and for each subject. In general, individual indexes were then averaged among the subjects in order to estimate the population index.

2.3. Data Analysis and Statistics

Among all the 12 tasks (i.e., TV spots), four in particular were chosen for the analysis: two 30-s-long (the maximum length of the tasks in the present protocol) in order to verify the task-independency of the results, one 20-s, and another 15-s-long in order to also investigate the effect of task duration. At first, for each of the selected spots and for each of the three indices, 25 subgroups of subjects' numerousness (from 10 to 34 participants) taken from the 36 participants at the study were considered in the analysis. For each of these subgroups, 630 different random combinations of subjects were selected using a bootstrap, from the initial population. It was carried out to avoid any possible effect associated to the choice of subjects. The number of combinations for each subgroup (630) was chosen since it is the maximum number of possible combinations considering a subset of 34 subjects from the initial sample of 36 (Equation (1)).

$$C_{(36,34)} = \frac{36!}{(34!(36-34)!)} = 630 \tag{1}$$

For coherence, this number of combinations was extended to the other 24 subgroups of subjects by randomly selecting 630 out of all the possible combinations for that specific case.

Keeping in mind that for each task, each index corresponds to a t-long (with t = 30, 20, 15) vector (i.e., the population index '*v*'), for every subgroup, for each spot, and for each index, we computed:

1. Group-mean values of the index (function of t, i.e., the task duration), for each of the 630 combinations. We so obtained 630 vectors '*v630*' (thus resulting in a matrix 630 x t);
2. Pearson correlation between each '*v630*' (630 x t) and the vector 'v' (1 x t), containing the mean values of the index computed over the entire population (36 subjects);
3. Mean Squared Error (MSE) to describe the error committed considering each '*v630*' rather than '*v*' along each task (within-task variability):

$$MSE = \frac{\sum_{i=1}^{t} (v630_i - v_i)^2}{t} \tag{2}$$

In this formula '*v630*' represents the vector of the observed values while '*v*' is assumed as the vector of predicted values (i.e., the unique possible vector of each index if considering the full population).

4. The standard deviation of the 630 values assumed by the vectors '*v*630', for every second of the task itself (between-groups variability):

$$STD = \sqrt{\frac{\sum_{n=1}^{630} \left(v630_n - \overline{v630}\right)^2}{630}} \tag{3}$$

The effect of the sample size reduction was explored considering the variables described in the previous section.

In order to ensure the statistical power of the analysis, only five subgroups with a steady reduction of the sample size were considered. Specifically, they were subgroups of 32, 28, 24, 20, and 16 subjects.

In the figure below (Figure 1), it is possible to visually appreciate, for a representative spot and index, the increasing variability of the index itself when decreasing the sample size. In fact, they represented the maximum (Max) and the minimum (Min) values of the index among the 630 combinations of the n-th sample size (n = 32, 28, 24, 20, 16), i.e., the '*v630*' vectors with respect to the '*v*' one (full sample size, i.e., ALL).

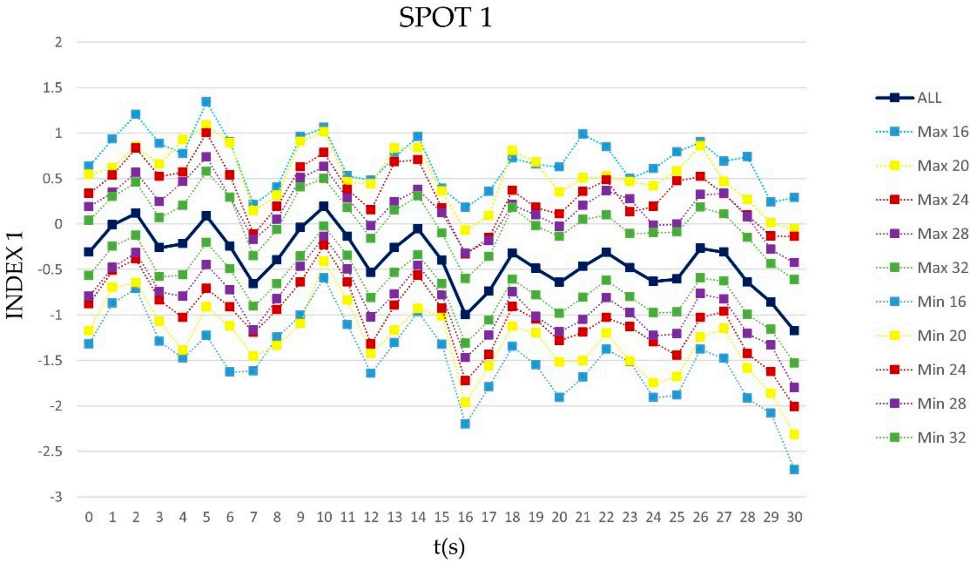

Figure 1. The graph represents the trend of '*v*', the mean value of Index1 computed over the entire population (36 subjects) during SPOT1 (ALL). The dashed lines show the trend of the maximum and minimum values (Max, Min) of the index among the '*v630*' for each second, in each subgroup of subjects (32, 28, 24, 20, 16).

The statistical analysis was performed on these data to assess the differences among the five sub-groups. At first, the Pearson correlation between each '*v630*' with respect to '*v*' was conducted for each subgroup of subjects and the number of significant correlations were reported. The p-value for these correlations was corrected through the Bonferroni method in order to avoid any improper deduction due to the type I error, the likelihood of which is higher with greater numbers of repeated statistical tests carried out. In particular, being that the multiple comparisons were equal to 630, the test was deemed significant for $p < 8 \times 10^{-5}$ [53]. The Shapiro Wilk test [54] of normality demonstrated that data (the Rho correlation coefficients, MSE, and STD values) were not Gaussian; hence, the non-parametric Friedman test [55] was performed to assess the difference between groups. Specifically, the subgroups of subjects (32, 28, 24, 20, 16) were considered as within factors for the analysis run over the Rho values resulting from the correlation analysis, the MSE and the STD values. The analysis was performed for each of the three considered indices and each of the four selected spots. Nemenyi post-hoc test [56], specifically conceived for non-parametric repeated measures ANOVA (i.e., the Friedman test), was applied to further analyse the significant effects and interactions.

In conclusion, the Rho correlation coefficient was computed and summarised for each index for each task for all the possible sub-sample sizes (from 34 up to 10) in order to visually highlight the decreasing curve of sample size-correlation relationship.

3. Results

3.1. The Effect of the Index

The analysis of the first spot (SPOT1) was conducted exploring the effect of the sample size reduction on the three different indices.

Concerning Index1, EEG frontal desynchronization, the correlation analysis showed that '*v630*' significantly correlated with the vector '*v*', associated with the entire population, considering the subgroups of 32 and 28 subjects. Reducing the sample size to 24, to 20, and to 16, two correlations over 630, 33 over 630, and 151 over 630 resulted as respectively not significant ($p > 8 \times 10^{-5}$). The Friedman test evidenced a significant decreasing effect of the mean rho values depending on the sub-groups, Figure 2a (Friedman chi-squared = 1868, p-value < 2.2×10^{-16}). The post-hoc analysis showed that rho significantly decreased from 32 to 16 subjects, assuming different values for each subgroup ($p < 0.05$).

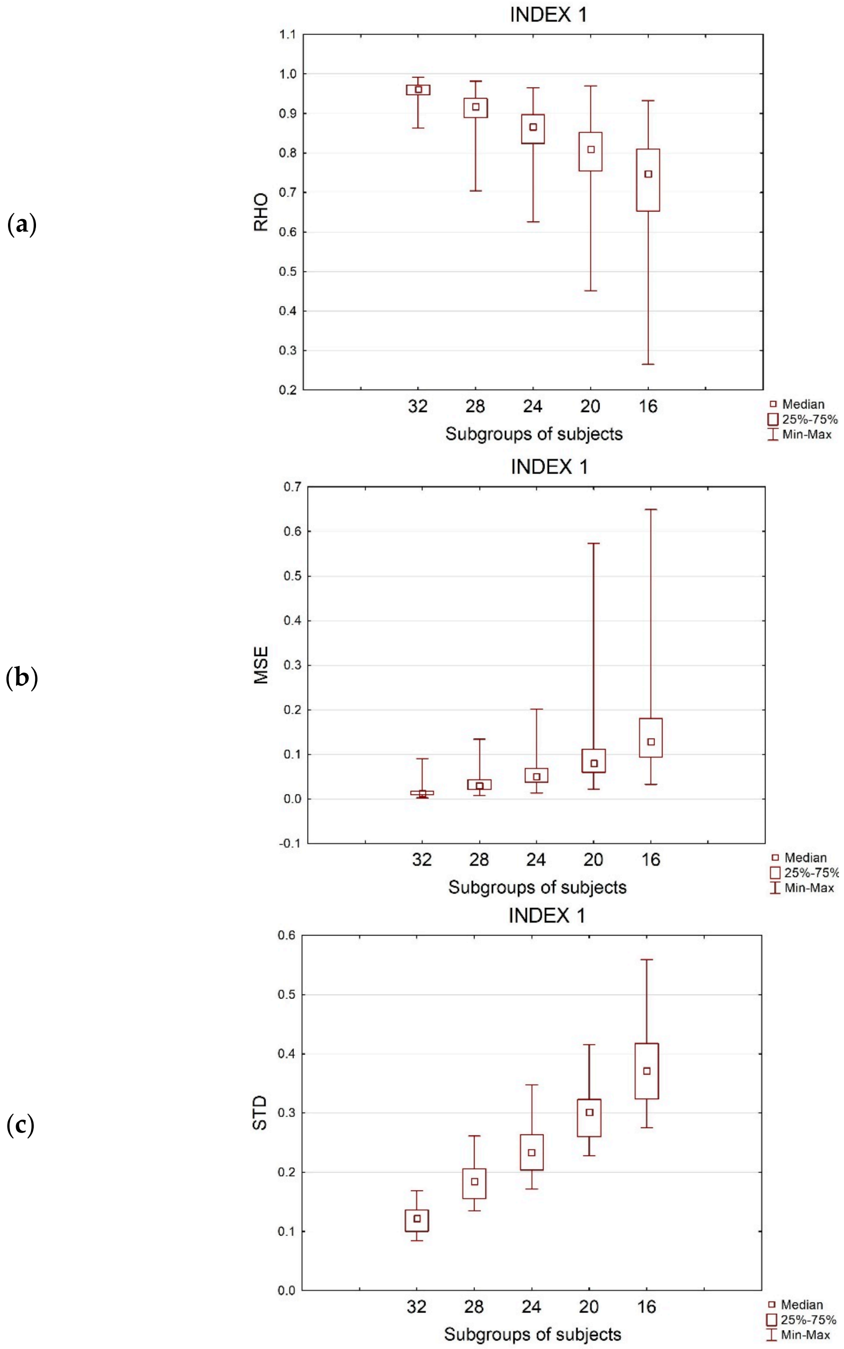

Figure 2. The box plots represent the effect of the subgroup of subjects on (**a**) Rho values, (**b**) MSE values, and (**c**) STD values, computed for Index 1, during the SPOT 1.

Additionally, the MSE showed dependency from the subgroups of subjects, Figure 2b (Friedman chi-squared = 1932.3, p-value < 2.2×10^{-16}): it was significantly lower for 32 subjects, increasing with the sample size reduction. The highest error was committed considering 16 subjects instead of 36. All the groups were significantly different as showed by the Nemenyi test.

The analysis of the STD evidenced the effect of the groups, Figure 2c (Friedman chi-squared = 124, p-value < 2.2 × 10^{-16}). It was lower for 32 subjects, increasing with the reduction of the sample size. The post-hoc analysis demonstrated that STD was significantly different for each subgroup.

Concerning Index2, EEG frontal asymmetry, all the correlations were significant for the subgroups 32 and 28 ($p < 8 \times 10^{-5}$). Concerning the subgroups of 24, 20, and 16 subjects, 13 over 630, 92 over 630, and 237 over 630 correlations were respectively not significant. The Friedman analysis on rho, MSE, and STD showed the effect of the groups, with a significant increase of rho (Friedman chi-squared = 1868, p-value < 2.2 × 10^{-16}), a significant increase of MSE (Friedman chi-squared = 1932.3, p-value < 2.2 × 10^{-16}), and a significant increase of STD from 32 to 16 subjects (Friedman chi-squared = 124, p-value < 2.2 × 10^{-16}). The post-hoc Nemenyi test evidenced that all the subgroups assumed different values of rho, MSE, and STD.

Considering Index3, Autonomic response, for every subgroup, only the subgroup 16 presented 12 not-significant correlations over 630 ($p > 8 \times 10^{-5}$). The Friedman analysis on rho, MSE, and STD showed the effect of the groups, with a significant increase of rho (Friedman chi-squared = 2235.5, p-value < 2.2 × 10^{-16}), a significant increase of MSE (Friedman chi-squared = 2232.1, p-value < 2.2 × 10^{-16}), and a significant increase of STD from 32 to 16 subjects (Friedman chi-squared = 124, p-value < 2.2 × 10^{-16}). The post-hoc Nemenyi test evidenced that all the subgroups assumed different values of rho, MSE, and STD.

The analysed variables for SPOT1 and the three indexes are summarised in Table 1.

3.2. The Effect of the Task

The analysis on SPOT2 (Table 2), with the same length of SPOT1 but different content, showed that all the correlations between the '*v630*' and '*v*' were significant for the subgroups 32, 28, and 24, considering Index1 ($p < 8 \times 10^{-5}$), while for the subgroups 32 and 28 for Index2 and Index3. The analysis evidenced that reducing the sample size at 24:10 and 49 over 630 correlations were not significant for Index2 and Index3, respectively. For a sample size of 20 subjects, 8, 90, and 138 over 630 correlations were not significant for the three indices, respectively. Moving on, 16 subjects, 49, 215, and 290 over 630 were not significant for the three indices, respectively. Friedman test on rho, MSE, and STD revealed a significant effect of the groups for the three indices, showing rho decreasing (Index1: Friedman chi-squared = 1879.4, p-value < 2.2 × 10^{-16}; Index2: Friedman chi-squared = 1800.9, p-value < 2.2 × 10 $^{-16}$; Index3: Friedman chi-squared = 1888.1, p-value < 2.2 × 10^{-16}), MSE increasing (Index1: Friedman chi-squared = 2124.8, p-value < 2.2 × 10^{-16}; Index2: Friedman chi-squared = 2151.2, p-value < 2.2 × 10^{-16}; Index3: Friedman chi-squared = 2182.7, p-value < 2.2 × 10^{-16}), and STD increasing (Index1: Friedman chi-squared = 124, p-value < 2.2 × 10^{-16}; Index2: Friedman chi-squared = Friedman chi-squared = 124, p-value < 2.2 × 10^{-16}; Index3: Friedman chi-squared = 124, p-value < 2.2 × 10^{-16}). The post-hoc test showed that all the comparisons between groups were significant.

Table 1. Number of significant correlations (#s.c.) and eventual decreasing as percentage of the total number of combinations (630), median ± std of rho, MSE, and STD for the three indices (I1, I2, I3) and the five subgroups of subjects (32, 28, 24, 20, 16) during SPOT1.

SPOT1(30 s)												
	#s.c. ($p < 8 \times 10^{-5}$)			**Rho**			**MSE**			**STD**		
	I1	**I2**	**I3**	**I1**	**I2**	**I3**	**I1**	**I2**	**I3**	**I1**	**I2**	**I3**
32	**630 ***	**630 ***	**630 ***	0.96 ± 0.02	0.94 ± 0.02	0.97 ± 0.01	0.012 ± 0.01	0.014 ± 0.01	0.001 ± 0.0005	0.12 ± 0.02	0.13 ± 0.01	0.03 ± 0.003
28	**630 ***	**630 ***	**630 ***	0.91 ± 0.04	0.89 ± 0.05	0.94 ± 0.02	0.029 ± 0.02	0.032 ± 0.02	0.002 ± 0.001	0,18 ± 0.03	0,19 ± 0.02	0.05 ± 0.004
24	628 (−0.3%)	617 (−2.1%)	**630 ***	0.87 ± 0.06	0.84 ± 0.07	0.90 ± 0.03	0,05 ± 0.03	0,053 ± 0.03	0,004 ± 0.002	0.23 ± 0.04	0.23 ± 0.03	0.07 ± 0.007
20	597 (−5.2%)	538 (−14.6%)	**630 ***	0.81 ± 0,08	0.76 ± 0.10	0.86 ± 0.04	0,079 ± 0.06	0,089 ± 0.05	0,007 ± 0.003	0.30 ± 0.05	0.31 ± 0.04	0.09 ± 0.007
16	479 (−23.9%)	393 (−37.6%)	618 (−1.9%)	0.74 ± 0.12	0.69 ± 0.12	0.79 ± 0.06	0.128 ± 0.08	0.136 ± 0.08	0.011 ± 0.005	0.37 ± 0.06	0.38 ± 0.04	0.11 ± 0.01

Table 2. Number of significant correlations (#s.c.) and eventual decreasing as percentage of the total number of combinations (630), median ± std of rho, MSE, and STD for the three indices (I1, I2, I3) and the five subgroups of subjects (32, 28, 24, 20, 16) during SPOT2.

SPOT2 (30 s)												
	#s.c. ($p < 8 \times 10^{-5}$)			**Rho**			**MSE**			**STD**		
	I1	**I2**	**I3**	**I1**	**I2**	**I3**	**I1**	**I2**	**I3**	**I1**	**I2**	**I3**
32	**630 ***	**630 ***	**630 ***	0.98 ± 0.007	0.95 ± 0.02	0.94 ± 0.02	0.02 ± 0.01	0.02 ± 0.01	0.001 ± 0.00	0.14 ± 0.02	0.13 ± 0.02	0.03 ± 0.004
28	**630 ***	**630 ***	**630 ***	0.96 ± 0.01	0.9 ± 0.04	0.88 ± 0.04	0.04 ± 0.01	0.04 ± 0.02	0.002 ± 0.00	0.21 ± 0.02	0.19 ± 0.03	0.05 ± 0.006
24	**630 ***	620 (−1.6%)	598 (−5.1%)	0.94 ± 0.03	0.84 ± 0.07	0.81 ± 0.07	0.07 ± 0.03	0.07 ± 0.03	0.004 ± 0.001	0.27 ± 0.04	0.26 ± 0.05	0.07 ± 0.008
20	622 (−1.2%)	540 (−14.3%)	492 (−21.9%)	0.91 ± 0.06	0.78 ± 0.10	0.74 ± 0.11	0.12 ± 0.05	0.11 ± 0.05	0.007 ± 0.003	0.34 ± 0.04	0.32 ± 0.06	0.08 ± 0.01
16	581 (−7.8%)	415 (−34.1%)	340 (−46%)	0.86 ± 0.08	0.70 ± 0.13	0.67 ± 0.14	0.18 ± 0.07	0.17 ± 0.07	0.01 ± 0.004	0.42 ± 0.06	0.40 ± 0.07	0.1 ± 0.02

3.3. The Effect of the Time

To explore the influence of the task time duration, or rather the length of the spot in this case, the last analysis was conducted on the number of significant correlations found for SPOT3 and SPOT4, respectively 20 and 15 s long (Table 3). The MSE and STD were not considered in this part of the study because they do not depend on the time. The analysis of SPOT3 showed that all the correlations were significant for the subgroups of 32 and 28 subjects for both Index1 and Index3 ($p < 8 \times 10^{-5}$). Concerning Index1, 12, 53, and 136 correlations over 630 were not significant in subgroups 24, 20, and 16, respectively, while for Index3, 2, 63, and 216 correlations were not significant in those subgroups. Considering Index2, the number of not-significant correlations increased from the subgroup of 28 subjects in which 23 correlations were not significant, while 115, 290, and 408 over 630 correlations were not significant for the subgroups of 24, 20, and 16 subjects, respectively. Friedman test on rho values revealed a significant effect of the groups for the three indices, showing a rho decreasing related to the reduction of the sample size (Index1: Friedman chi-squared = 1712.1, p-value < 2.2×10^{-16}; Index2: Friedman chi-squared = 1586.3, p-value < 2.2×10^{-16}; Index3: Friedman chi-squared = 2017.2, p-value < 2.2×10^{-16}). Post-hoc test showed that all the comparisons between groups were significant. The analysis of SPOT4, showed that all the correlations were significant for the subgroups of 32 subjects, for both Index1 and Index3 ($p < 8 \times 10^{-5}$). Concerning Index1 63, 214, 335 and 464 correlations over 630 were not significant in subgroups 28, 24, 20 and 16 respectively, while for Index3, 19, 122, 296 and 423 correlations were not significant in those subgroups. Considering Index2, the number of not-significant correlations increased, already from the subgroup of 32 subjects in which 7 correlations were not significant, while 82, 203, 342, 450 over 630 correlations were not significant for the subgroups of 28, 24, 20 and 16 subjects, respectively. Friedman test on rho values revealed a significant effect of the groups for the three indices, showing rho decrease related to the reduction of the sample size (Index1: Friedman chi-squared = 1510.9, p-value < 2.2×10^{-16}; Index2: Friedman chi-squared = 1263, p-value < 2.2×10^{-16}; Index3: Friedman chi-squared = 1617.6, p-value < 2.2×10^{-16}). The post-hoc test showed that all the comparisons between groups were significant.

Table 3. Number of significant correlations (#s.c.) and eventual decrease as percentage of the total number of combinations (630), median ± std of rho values for the three indices (I1, I2, I3), and the five subgroups of subjects (32, 28, 24, 20, 16) during SPOT3 and SPOT4.

	SPOT3 (20 s)						**SPOT4 (15 s)**					
	#s.c. ($p < 8 \times 10^{-5}$)			**Rho**			**#s.c. ($p < 8 \times 10^{-5}$)**			**Rho**		
	I1	**I2**	**I3**	**I1**	**I2**	**I3**	**I1**	**I2**	**I3**	**I1**	**I2**	**I3**
32	**630 ***	**630 ***	**630 ***	0.98 ± 0.01	0.95 ± 0.03	0.98 ± 0.01	**630 ***	623 (−1.1%)	**630 ***	0.96 ± 0.026	0.96 ± 0.04	0.97 ± 0.01
28	**630 ***	607 (−3.6%)	**630 ***	0.95 ± 0.03	0.88 ± 0.06	0.95 ± 0.02	567 (−10%)	548 (−13%)	611 (−3%)	0.92 ± 0.04	0.93 ± 0.05	0.94 ± 0.02
24	618 (−1.9%)	515 (−18.2%)	628 (−0.3%)	0.92 ± 0.05	0.83 ± 0.08	0.92 ± 0.04	416 (−33.9%)	427 (−32.2%)	508 (−19.3%)	0.87 ± 0.09	0.87 ± 0.13	0.89 ± 0.04
20	577 (−8.4%)	340 (−46%)	567 (−10%)	0.88 ± 0.07	0.76 ± 0.13	0.87 ± 0.06	295 (−53.2%)	288 (−54.3%)	334 (−46.9%)	0.80 ± 0.12	0.81 ± 0.19	0.84 ± 0.06
16	494 (−21.6%)	222 (−64.7%)	414 (−34.3%)	0.83 ± 0.10	0.70 ± 0.16	0.81 ± 0.07	166 (−73.7%)	180 (−71.4%)	207 (−67.1%)	0.74 ± 0.17	0.77 ± 0.24	0.78 ± 0.08

The number of significant correlations for the three indices and the four considered spots are summarised in Table 4.

Table 4. Number of significant correlations (#s.c.) and eventual decreasing as percentage of the total number of combinations (630) for the three indices (INDEX1, INDEX2, INDEX3) and the five subgroups of subjects (32, 28, 24, 20, 16) during SPOT1 (30 s), SPOT2 (30 s), SPOT3 (20 s), and SPOT4 (15 s).

	INDEX1				**INDEX2**				**INDEX3**			
	#s.c. ($p < 8 \times 10^{-5}$)				**#s.c. ($p < 8 \times 10^{-5}$)**				**#s.c. ($p < 8 \times 10^{-5}$)**			
	30 s	**30 s**	**20 s**	**15 s**	**30 s**	**30 s**	**20 s**	**15 s**	**30 s**	**30 s**	**20 s**	**15 s**
32	**630 ***	**630 ***	**630 ***	**630 ***	**630 ***	**630 ***	**630 ***	623 (−1.1%)	**630 ***	**630 ***	**630 ***	**630 ***
28	**630 ***	**630 ***	**630 ***	567 (−10%)	**630 ***	**630 ***	607 (−3,6%)	548 (−13%)	**630 ***	**630 ***	**630 ***	611 (−3%)
24	628 (−0.3%)	**630 ***	618 (−1.9%)	416 (−33.9%)	617 (−2.1%)	620 (−1.6%)	515 (−18.2%)	427 (−32.2%)	**630 ***	598 (−5.1%)	628 (−0.3%)	508 (−19.3%)
20	597 (−5.2%)	622 (−1.2%)	577 (−8.4%)	295 (−53.2%)	538 (−14.6%)	540 (−14.3%)	340 (−46%)	288 (−54.3%)	**630 ***	492 (−21.9%)	567 (−10%)	334 (−46.9%)
16	479 (−23.9%)	581 (−7.8%)	494 (−21.6%)	166 (−73.7%)	393 (−37.6%)	415 (−34.1%)	222 (−64.7%)	180 (−71.4%)	618 (−1.9%)	340 (−46%)	414 (−34.3%)	207 (−67.1%)

3.4. Rho-Sample Size Relationship

In conclusion, the average value of the Rho correlation coefficient between ‘*v630*’ and ‘*v*’ vectors, and its standard deviation among all the 630 combinations, were computed for all the possible sample sizes, from 34 up to 10 (generally considered as the minimum acceptable sample size) in order to model the function of the relationship between the correlation with the whole “Reference population (36 subjects) index” and the indexes resulting from reducing the sample size. Represented below (Figure 3) is the graphical representation of such a function for each spot for each index.

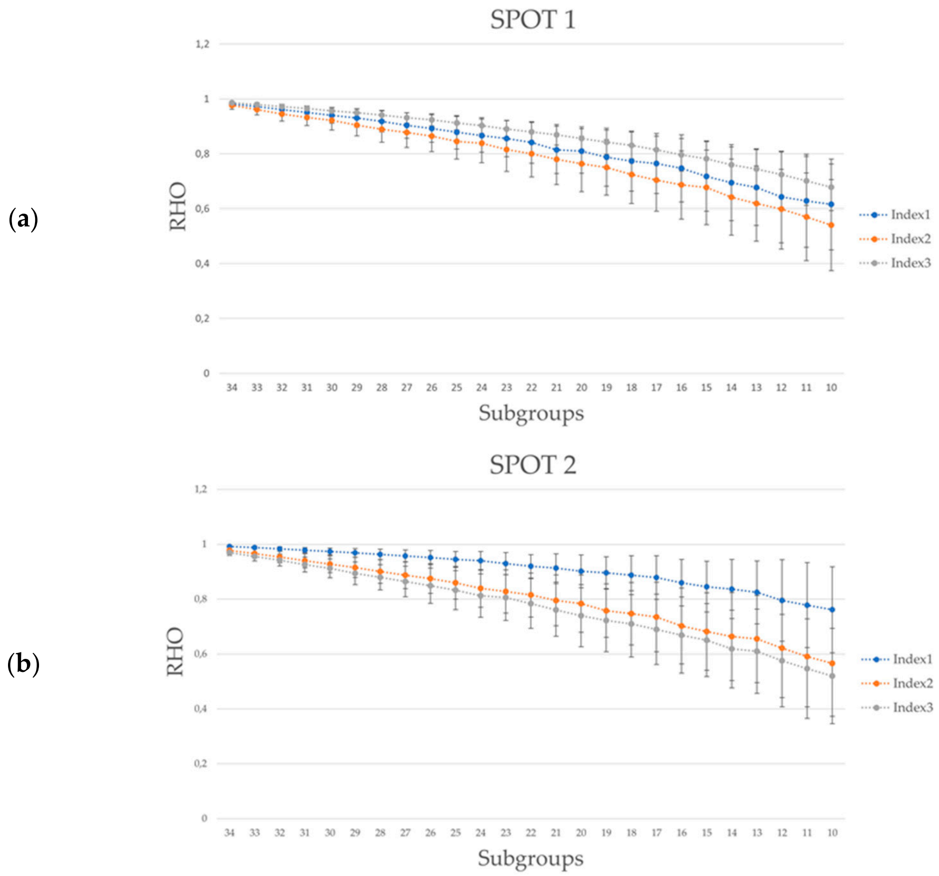

Figure 3. *Cont.*

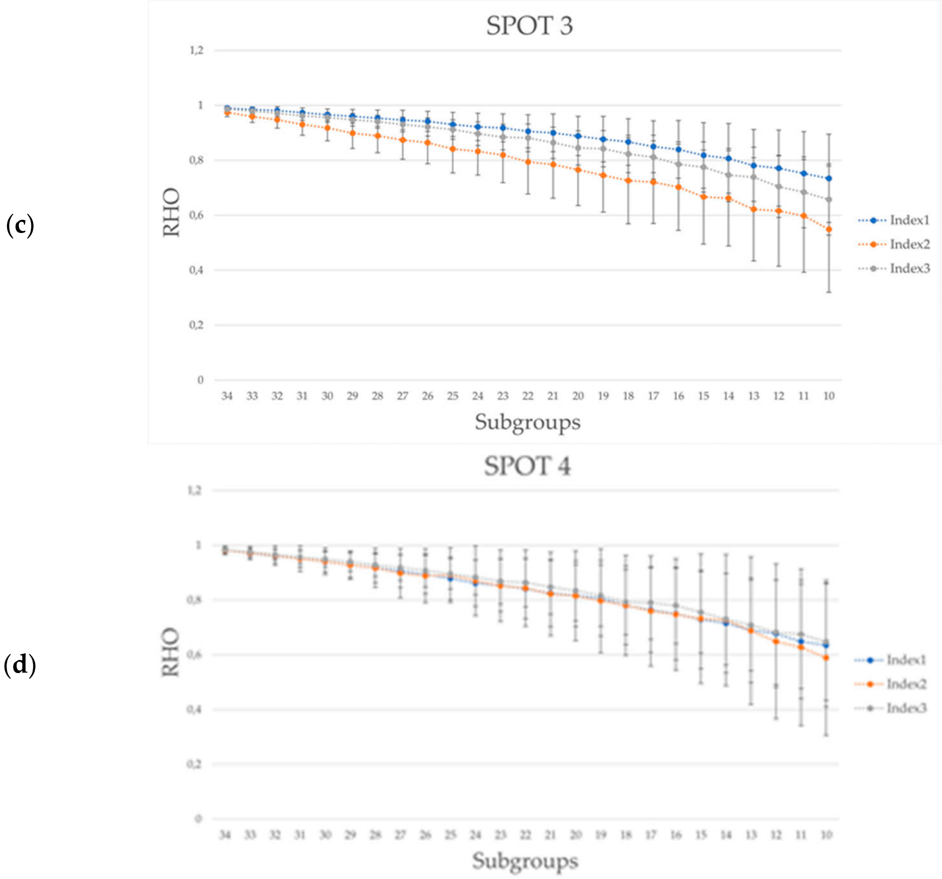

Figure 3. Graphical representation of the trend of rho correlation coefficient at reducing the sample size (subgroups) with respect to the 'full population index' (36 participants). Tasks (**a**) and (**b**) were 30 s long, Task (**c**) 20 s, and Task (**d**) 15 s long.

4. Discussion

In scientific literature, the reduction of the sample size is a much-debated topic. Despite many authors claiming the importance of including an optimal number of participants, applied research and even more studies performed in real contexts could be limited by several constraints, and so it could become problematic to involve large experimental samples of participants. After all, in the field of EEG research, studies published in peer-reviewed journals do not have the same sample size, even when they consider the same topic. For instance, Xu and Zhong, in their review on the EEG in educational research [57], pointed out that over the 22 screened papers, the sample size varied from 80 to 5 subjects, undermining the study outcome's interpretability and replicability. The present research aimed to investigate the modifications of the outcomes of a study in the context of applied neurosciences when the sample size was reduced. An initial sample of 36 subjects was considered as reference. This number was deemed appropriate to represent the sample since it is around the size of participants generally involved in this kind of study. This is confirmed by the review of Bazzani and colleagues [32], who reported that, over studies that used advertisings as stimuli, 33 was the average number of recruited participants. Five subgroups of subjects were randomly chosen 630 times (each n-th combination was different from the others) and the related outcomes were compared to the condition in which all the subjects were included. The first factor considered in the analysis was the dependency from the index, or rather whether the sample size reduction had different effects on indices with different natures (general EEG activity, 'neurometric', intended as a combination of EEG activity features, and autonomic biosignals-related metric). Considering the targeted task, thirty seconds long, and comparing the effects found on rho, MSE, and STD in response to the decreasing of the included sample, we observed that the patterns associated to the three indices were comparable. As expected, rho significantly decreased while the MSE and STD significantly increased, moving from 32 to 16 subjects for all the indices. This latter result is particularly relevant since in scientific literature, while dealing with research outcomes' discussion and their statistical power, effect size is generally represented through the Cohen's d coefficient [58], computed as the ratio between the 'μ' mean value of the effect, and the 'σ' standard deviation of the measures over the sample. Cohen [58] divided effect sizes into large, medium, and small, where large effect sizes are about one (the mean has the same size as the standard deviation), medium effect sizes are about a half, and small effect sizes are a quarter, while under one eighth, the effect sizes can be considered trivial [33]. This implies that with lower sample sizes, even if the effect "amplitude" remains comparable (i.e., the 'μ'), standard deviation (i.e., 'σ') increasing would negatively affect the effect size.

The number of significant correlations showed that, from 32 to 24 subjects, most of the comparisons were significant for all the three indices (Table 4). Therefore, from the first analysis, the reduction of the sample size from 36 to a maximum of 24 subjects seemed to generate comparable outcomes, independently from the nature of the considered index, committing an acceptable error (<0.1 on indexes with the 95% of distribution ranging between –3 and 3). During the second part of the study, we explored whether the findings remained consistent when the analysis was performed on a different task with same length (i.e., SPOT2). In this phase we pointed out that during SPOT2, rho, MSE, and STD values followed the same patterns observed in SPOT1: rho significantly decreased while MSE and STD significantly increased when the participants were reduced, for all the three indices. Furthermore, most of the correlations were significant considering the same threshold of 24 subjects, and the maximum deviation from 630 was observed in Index3, for which the decrease of significant correlations was equal to 5.1%. Hence, we concluded that the observed results were independent from the task when considering a spot of 30 s. The last factor investigated in the study was the effect of the time. In this phase, only the correlation was explored in the analysis since MSE and STD are not time-dependent variables. Considering spots shorter than 30 s, the correlation still decreased with the reduction of the participants but, contrarily to the SPOT1 and SPOT2, the threshold for which all the correlations were significant changed. Considering SPOT3, 20 s long, the threshold of significant correlation was 28 subjects for Index1 and Index3 but not for Index2, for which in 3.6 percent of cases (23/630), the correlation resulted as not significant. For SPOT4, the threshold increased for all the three indices: the analysis of the significant correlations showed most of the combinations were significantly correlated until 32 subjects instead of 24. These findings suggest that the modification of the outcomes related to the reduction of the sample size depended on the time, or rather on the length of the task employed in the study. It has been highlighted that the number of subjects for which the outcomes remained comparable increased with the decreasing of the spot duration, i.e., in general of the task.

If looking at the distribution of the Rho coefficients with respect to the different sample sizes (Figure 3), it has to be taken into consideration that in scientific literature the range 0.5–0.7 is considered as 'Moderate correlation', the range 0.7–0.9 as 'High correlation', and the range 0.9–1 as 'Strong correlation' [59]. Therefore, we can notice, independently from the index and the task length under 30 subjects, the correlation values moved from strong to high correlations, while under 18–20 subjects from high to moderate correlations, i.e., the probability that a subset of, for instance, 16 participants will represent the population behaviour over time is lower. This result is relevant considering that the statistical errors matter. As it is well known, statistical errors are classified in 'Type I error', i.e., the mistaken rejection of the null hypothesis ("false positive"), and 'Type II error', i.e., the mistaken acceptance of the null hypothesis ("false negative"). As stated in the introduction, this topic was not the focus of the study, since the experimental question was if the trends (significant or not) revealed by the study results coming from smaller samples are actually correlated with the "Reference population" behaviour (where 'Reference population' stands for the ideal case of the 36-sample dataset assumed as representative of the theorical population). Of course, the reduction of the sample size will undoubtedly increase the probability of a type II error, as commonly agreed upon in the literature and supported by our results (data standard deviation increasing with sample size decreasing). However, our results also shed light on the moderate correlation between a large sample and small sub-samples, meaning that it is not implicit that small samples are always representative of the population behaviour, thus increasing the probability of committing even type I error. In any case, the latter sentence has to be intended as a mere deduction, and was not demonstrated by the present study.

Therefore, it can be concluded that, as assumed, the correlations between the outcomes associated with the entire 'Reference population', considered as 36 observed samples, and the outcomes computed for different subgroups of subjects, decreased with the reduction of the sample size. As expected, the errors committed with the decreasing of the sample size increased as well as the dispersion of the data. However, it was possible to determine a threshold of subjects, above which almost all the comparisons were significant, the outcomes remained comparable, and the committed errors and the data dispersion were acceptable. This threshold was 24 subjects when the task lasted 30 s and 28 when the task length was reduced, even 32 when reduced up to 15 s. Further investigations will be necessary to explore whether these findings are replicable in the same and in other contexts, considering different kind of datasets. In addition, also the effect of the time needs to be examined further; for instance, it could be interesting to analyse whether the outcomes change considering different parts of the same task, or maybe with respect to specific events as well as in conditions discriminability, a common problem when employing machine-learning techniques [60,61]. In fact, in this study, time-varying functions (with a time resolution of 1 s) were employed, but in some research there is no need of achieving such a high time resolution of assessment: in some experimental protocols, lower time resolutions are enough (from some seconds to even minutes) [18,62,63]; alternatively, in other protocols, EEG indexes are employed as large averages on the entire task in order to investigate task discriminability [64,65]. In these cases, the lower time resolution would decrease the intra-subject variability, thus impacting the group average of the indexes, and therefore sample size variation could have different effects. Another common neuroscientific experimental approach envisages the analysis of Event-Related Potentials (ERPs) [24,66]; in this case, temporal resolution is even higher (milliseconds), but signal-to-noise ratio is increased by increasing the number of repetitions and computing multiple averages over them. Again, also in this kind of application, sample size variation could have different effects; therefore, neuroscientific field encloses several methodologically different applications.

This is the main limitation of the present research: the results of this study have to be considered as strictly related to this case study. The consistency of the results among the different tasks in the experiment and the outcomes of the research need to be investigated with larger samples (more than 36) and in other contexts to be generalisable. This should encourage further research to generalise the outcomes, i.e., ideally (but probably it would sound utopistic) to define a mathematical model able to predict the proper sample size of an experiment as a function of for example experimental design, time resolution, type of indicators, and so on. In addition, another topic that should be investigated more in depth is that one related to the probability of committing statistical errors. As discussed before, this study demonstrated why the probability of committing type II errors increases when decreasing the sample size. However, this study also hinted at the possibility of inferring wrong conclusions and committing type I errors with small sample sizes.

In conclusion, this study, of course, has not to be interpreted as an encouragement to reduce sample size in scientific research. However, in a context when participants recruiting is limited for several reasons, the present study aimed to provide an overview of sample reductions consequences in order to shed light on this concern when discussing research outcomes. It can be claimed that the reduction of the sample size does not necessarily affect the outcomes of research. However, it is necessary to carefully establish the degree of this reduction since there is a "threshold" above which the findings might lose their coherence and validity.

5. Conclusions

The present research aimed to investigate the modifications of the outcomes of a study in the context of applied neurosciences when the sample size was reduced, specifically up to 16 subjects from an original experimental sample of 36 subjects. Three different neurophysiological indexes have been considered, two based on EEG and one on autonomic—GSR and HR—activities, during four different tasks of decreasing time duration, i.e., from 30 to 15 s. This study highlighted how, for longer tasks (30 s), sample size decrease up to 24 participants did not dramatically affect the results, i.e., the computed neurophysiological indexes kept significant correlation with the same indexes computed on the whole population, of course with the compromise of increasing mean error and standard deviation. Under this "threshold" (24 subjects), results began to lose statistical power (no significant correlations) and measure variability (MSE and STD) still increased. These effects increased in terms of magnitude as short as the task was; in fact, with 15-long-tasks, the results suggested to consider 32 as the minimum sample size.

In the light of these results, although further investigation is required to assess these findings as generalisable, it is important to underline that, even in those applied research situations where it becomes difficult to have large sample sizes, it may still be unwise to conduct research with too small a sample because the results could be not representative of the population.

Author Contributions: Conceptualization, G.D.F., F.B. and A.V.; methodology, G.D.F. and A.V.; software, G.D.F., A.V., P.A., G.B. and N.S.; validation, A.V. and G.D.F.; formal analysis, A.V. and G.D.F.; investigation, G.D.F., P.C., A.V., V.R. and A.T.; resources, F.B. and A.T.; data curation, P.C., A.V. and V.R.; writing—original draft preparation, A.V. and G.D.F.; writing—review and editing, G.D.F., P.A., G.B., F.B., A.V. and A.T.; visualization, A.V. and G.D.F.; supervision, G.D.F.; project administration, G.D.F. and F.B.; funding acquisition, A.T., F.B. and G.D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-funded by European Commission through the Horizon2020 projects "WORKINGAGE: Smart Working environments for all Ages" (GA n. 826232); "SIMUSAFE: Simulator Of Behavioral Aspects For Safer Transport" (GA n. 723386); "SAFEMODE: Strengthening synergies between Aviation and maritime in the area of human Factors towards achieving more Efficient and resilient MODE of transportation" (GA n. 814961), "MINDTOOTH: Wearable device to decode human mind by neurometrics for a new concept of smart interaction with the surrounding environment" (GA n. 950998) and H2020-SESAR-2019-2 project: Transparent artificial intelligence and automation to air traffic management systems, "ARTIMATION", GA n. 894238. The individual grant BE-FOR-ERC for the project "BRAIN2GETHER" recognized by Sapienza University of Rome to Gianluca Di Flumeri is also acknowledged.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Sapienza University of Rome (protocol code nr. 2507/2020 and approved the 04/08/2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The aggregated data analysed in this study might be available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Larson, M.J.; Carbine, K.A. Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. *Int. J. Psychophysiol.* **2017**, *111*, 33–41. [CrossRef]
2. Button, K.S.; Ioannidis, J.P.; Mokrysz, C.; Nosek, B.A.; Flint, J.; Robinson, E.S.; Munafò, M.R. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **2013**, *14*, 365–376. [CrossRef] [PubMed]
3. Eng, J. Sample size estimation: How many individuals should be studied? *Radiology* **2003**, *227*, 309–313. [CrossRef] [PubMed]

4. Sanders, N.; Choo, S.; Nam, C.S. The EEG cookbook: A practical guide to neuroergonomics research. In *Cognitive Science and Technology*; Springer: Berlin, Germany, 2020; pp. 33–51.
5. Mao, Z.; Jung, T.P.; Lin, C.T.; Huang, Y. Predicting EEG sample size required for classification calibration. In Proceedings of the Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience—AC 2016, Toronto, ON, Canada, 17–22 July 2016; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer: Cham, Switzerland, 2016; Volume 9743, pp. 57–68. [CrossRef]
6. Anderson, S.F.; Kelley, K.; Maxwell, S.E. Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychol. Sci.* **2017**, *28*, 1547–1562. [CrossRef] [PubMed]
7. Sim, J.; Saunders, B.; Waterfield, J.; Kingstone, T. Can sample size in qualitative research be determined a priori? *Int. J. Soc. Res. Methodol.* **2018**, *21*, 619–634. [CrossRef]
8. Javanmard, A.; Montanari, A. Debiasing the lasso: Optimal sample size for Gaussian designs. *Ann. Stat.* **2018**, *46*, 2593–2622. [CrossRef]
9. Combrisson, E.; Jerbi, K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* **2015**, *250*, 126–136. [CrossRef]
10. Baker, D.H.; Vilidaite, G.; Lygo, F.A.; Smith, A.K.; Flack, T.R.; Gouws, A.D.; Andrews, T.J. Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychol. Methods* **2021**, *26*, 295–314. [CrossRef]
11. Guttmann-Flury, E.; Sheng, X.; Zhang, D.; Zhu, X. A priori sample size determination for the number of subjects in an EEG experiment. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Berlin, Germany, 23–27 July 2019; pp. 5180–5183. [CrossRef]
12. Nosek, B.A.; Spies, J.R.; Motyl, M. Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* **2012**, *7*, 615–631. [CrossRef] [PubMed]
13. Arico, P.; Borghini, G.; Di Flumeri, G.; Sciaraffa, N.; Babiloni, F. Passive BCI beyond the lab: Current trends and future directions. *Physiol. Meas.* **2018**, *39*, 08TR02. [CrossRef]
14. Di Flumeri, G.; Aricò, P.; Borghini, G.; Sciaraffa, N.; Di Florio, A.; Babiloni, F. The dry revolution: Evaluation of three different eeg dry electrode types in terms of signal spectral features, mental states classification and usability. *Sensors* **2019**, *19*, 1365. [CrossRef]
15. Ayaz, H.; Dehais, F. *Neuroergonomics: The Brain at Work and in Everyday Life*; Elsevier: Amsterdam, The Netherlands, 2018.
16. Mühl, C.; Jeunet, C.; Lotte, F. EEG-based workload estimation across affective contexts. *Front. Neurosci.* **2014**, *8*, 114. [CrossRef]
17. Berka, C.; Levendowski, D.J.; Lumicao, M.N.; Yau, A.; Davis, G.; Zivkovic, V.T.; Olmstead, R.E.; Tremoulet, P.D.; Craven, P.L. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* **2007**, *78*, B231–B244.
18. Di Flumeri, G.; De Crescenzio, F.; Berberian, B.; Ohneiser, O.; Kramer, J.; Aricò, P.; Borghini, G.; Babiloni, F.; Bagassi, S.; Piastra, S. Brain–computer interface-based adaptive automation to prevent out-of-the-loop phenomenon in air traffic controllers dealing with highly automated systems. *Front. Hum. Neurosci.* **2019**, *13*, 296. [CrossRef] [PubMed]
19. Karthaus, M.; Wascher, E.; Getzmann, S. Proactive vs. reactive car driving: EEG evidence for different driving strategies of older drivers. *PLoS ONE* **2018**, *13*, e0191500. [CrossRef]
20. Di Flumeri, G.; Borghini, G.; Aricò, P.; Sciaraffa, N.; Lanzi, P.; Pozzi, S.; Vignali, V.; Lantieri, C.; Bichicchi, A.; Simone, A.; et al. EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Front. Hum. Neurosci.* **2018**, *12*, 509. [CrossRef] [PubMed]
21. Islam, M.R.; Barua, S.; Ahmed, M.U.; Begum, S.; Aricò, P.; Borghini, G.; Di Flumeri, G. A novel mutual information based feature set for drivers' mental workload evaluation using machine learning. *Brain Sci.* **2020**, *10*, 551. [CrossRef] [PubMed]
22. Liu, X.; Zhang, J.; Yin, J.; Bi, S.; Eisenbach, M.; Wang, Y. Monte Carlo simulation of order-disorder transition in refractory high entropy alloys: A data-driven approach. *Comput. Mater. Sci.* **2021**, *187*, 110135. [CrossRef]
23. Chavarriaga, R.; Khaliliardali, Z.; Gheorghe, L.; Iturrate, I.; Millán, J.D.R. EEG-based decoding of error-related brain activity in a real-world driving task. *J. Neural Eng.* **2015**, *12*, 066028. [CrossRef]
24. Marucci, M.; Di Flumeri, G.; Borghini, G.; Sciaraffa, N.; Scandola, M.; Pavone, E.F.; Babiloni, F.; Betti, V.; Aricò, P. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Sci. Rep.* **2021**, *11*, 4831. [CrossRef]
25. Cherubino, P.; Martinez-Levy, A.C.; Caratu, M.; Cartocci, G.; Di Flumeri, G.; Modica, E.; Rossi, D.; Mancini, M.; Trettel, A. Consumer behaviour through the eyes of neurophysiological measures: State-of-the-art and future trends. *Comput. Intell. Neurosci.* **2019**, *2019*, 1976847. [CrossRef] [PubMed]
26. Morin, C. Neuromarketing: The new science of consumer behavior. *Society* **2011**, *48*, 131–135. [CrossRef]
27. Sciaraffa, N.; Borghini, G.; Di Flumeri, G.; Cincotti, F.; Babiloni, F.; Aricò, P. Joint analysis of eye blinks and brain activity to investigate attentional demand during a visual search task. *Brain Sci.* **2021**, *11*, 562. [CrossRef]
28. Amores, J.; Richer, R.; Zhao, N.; Maes, P.; Eskofier, B.M. Promoting relaxation using virtual reality, olfactory interfaces and wearable EEG. In Proceedings of the 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN 2018), Las Vegas, NV, USA, 4–7 March 2018; Volume 2018, pp. 98–101. [CrossRef]

29. Di Flumeri, G.; Aricò, P.; Borghini, G.; Sciaraffa, N.; Maglione, A.G.; Rossi, D.; Modica, E.; Trettel, A.; Babiloni, F.; Colosimo, A.; et al. EEG-based Approach-Withdrawal index for the pleasantness evaluation during taste experience in realistic settings. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Jeju, Korea, 11–15 July 2017; pp. 3228–3231. [CrossRef]
30. Yoto, A.; Moriyama, T.; Yokogoshi, H.; Nakamura, Y.; Katsuno, T.; Nakayama, T. Effect of smelling green tea rich in aroma components on EEG activity and memory task performance. *Int. J. Affect. Eng.* **2014**, *13*, 227–233. [CrossRef]
31. Velavan, T.P.; Meyer, C.G. The COVID-19 epidemic. *Trop. Med. Int. Health* **2020**, *25*, 278–280. [CrossRef] [PubMed]
32. Bazzani, A.; Ravaioli, S.; Trieste, L.; Faraguna, U.; Turchetti, G. Is EEG suitable for marketing research? A systematic review. *Front. Neurosci.* **2020**, *14*, 594566. [CrossRef] [PubMed]
33. Friston, K. Ten ironic rules for non-statistical reviewers. *Neuroimage* **2012**, *61*, 1300–1310. [CrossRef] [PubMed]
34. Vecchiato, G.; Toppi, J.; Astolfi, L.; Fallani, F.D.V.; Cincotti, F.; Mattia, D.; Bez, F.; Babiloni, F. Spectral EEG frontal asymmetries correlate with the experienced pleasantness of TV commercial advertisements. *Med. Biol. Eng. Comput.* **2011**, *49*, 579–583. [CrossRef]
35. Cartocci, G.; Cherubino, P.; Rossi, D.; Modica, E.; Maglione, A.G.; Di Flumeri, G.; Babiloni, F. Gender and age related effects while watching TV advertisements: An EEG study. *Comput. Intell. Neurosci.* **2016**, *2016*, 3795325. [CrossRef]
36. Nielsen. Advertising and Audiences: Making Ad Dollars Make Sense. Available online: https://www.nielsen.com/us/en/insights/article/2014/advertising-and-audiences-making-ad-dollars-make-sense/ (accessed on 27 August 2021).
37. Belouchrani, A.; Abed-Meraim, K.; Cardoso, J.F.; Moulines, E. A blind source separation technique using second-order statistics. *IEEE Trans. Signal. Process.* **1997**, *45*, 434–444. [CrossRef]
38. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef]
39. EEGLAB Wiki. 6. Reject Artifacts. Available online: https://eeglab.org/tutorials/06_RejectArtifacts/ (accessed on 28 May 2021).
40. Doppelmayr, M.; Klimesch, W.; Pachinger, T.; Ripper, B. Individual differences in brain dynamics: Important implications for the calculation of event-related band power. *Biol. Cybern.* **1998**, *79*, 49–57. [CrossRef]
41. Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Res. Rev.* **1999**, *29*, 169–195. [CrossRef]
42. Skrandies, W. Global field power and topographic similarity. *Brain Topogr.* **1990**, *3*, 137–141. [CrossRef] [PubMed]
43. Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures; Boucsein, W.; Fowles, D.C.; Grimnes, S.; Ben-Shakhar, G.; Roth, W.T.; Dawson, M.E.; Filion, D.L. Publication recommendations for electrodermal measurements. *Psychophysiology* **2012**, *49*, 1017–1034. [CrossRef]
44. Pan, J.; Tompkins, W.J. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *BME-32*, 230–236. [CrossRef] [PubMed]
45. Benedek, M.; Kaernbach, C. A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* **2010**, *190*, 80–91. [CrossRef] [PubMed]
46. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* **2012**, *16*, 606–617. [CrossRef]
47. Magosso, E.; De Crescenzio, F.; Ricci, G.; Piastra, S.; Ursino, M. EEG alpha power is modulated by attentional changes during cognitive tasks and virtual reality immersion. *Comput. Intell. Neurosci.* **2019**, *2019*, 7051079. [CrossRef] [PubMed]
48. Briesemeister, B.B.; Tamm, S.; Heine, A.; Jacobs, A.M. Approach the good, withdraw from the bad—A review on frontal alpha asymmetry measures in applied psychological research. *Psychology* **2013**, *4*, 261–267. [CrossRef]
49. Modica, E.; Cartocci, G.; Rossi, D.; Martinez Levy, A.C.; Cherubino, P.; Maglione, A.G.; Di Flumeri, G.; Mancini, M.; Montanari, M.; Perrotta, D.; et al. Neurophysiological responses to different product experiences. *Comput. Intell. Neurosci.* **2018**, *2018*, 9616301. [CrossRef]
50. Cartocci, G.; Modica, E.; Rossi, D.; Inguscio, B.; Arico, P.; Martinez Levy, A.C.; Mancini, M.; Cherubino, P.; Babiloni, F. Antismoking campaigns' perception and gender differences: A comparison among EEG indices. *Comput. Intell. Neurosci.* **2019**, *2019*, 7348795. [CrossRef]
51. Davidson, R.J.; Ekman, P.; Saron, C.D.; Senulis, J.A.; Friesen, W.V. Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology I. *J. Pers. Soc. Psychol.* **1990**, *58*, 330–341. [CrossRef]
52. Vecchiato, G.; Maglione, A.G.; Cherubino, P.; Wasikowska, B.; Wawrzyniak, A.; Latuszynska, A.; Latuszynska, M.; Nermend, K.; Graziani, I.; Leucci, M.R.; et al. Neurophysiological tools to investigate consumer's gender differences during the observation of TV commercials. *Comput. Math. Methods Med.* **2014**, *2014*, 912981. [CrossRef]
53. Bonferroni, C.E. Teoria Statistica delle Classi e Calcolo delle Probabilità—Google Libri. Available online: https://books.google.it/books/about/Teoria_statistica_delle_classi_e_calcolo.html?id=3CY-HQAACAAJ&redir_esc=y (accessed on 24 August 2021).
54. Shapiro, S.S.; Wilk, M. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
55. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **1937**, *32*, 675. [CrossRef]
56. Nemenyi, P.B. *Distribution-Free Multiple Comparisons*; ProQuest: Ann Arbor, MI, USA, 1963.
57. Xu, J.; Zhong, B. Review on portable EEG technology in educational research. *Comput. Hum. Behav.* **2018**, *81*, 340–349. [CrossRef]
58. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Routledge: New York, NY, USA, 1988.

59. Mukaka, M.M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **2012**, *24*, 69–71. Available online: www.mmj.medcol.mw (accessed on 1 July 2021). [PubMed]
60. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef]
61. Khodayari-Rostamabad, A.; Reilly, J.P.; Hasey, G.M.; de Bruin, H.; MacCrimmon, D.J. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin. Neurophysiol.* **2013**, *124*, 1975–1985. [CrossRef]
62. Zander, T.O.; Andreessen, L.M.; Berg, A.; Bleuel, M.; Pawlitzki, J.; Zawallich, L.; Krol, L.R.; Gramann, K. Evaluation of a dry EEG system for application of passive brain-computer interfaces in autonomous driving. *Front. Hum. Neurosci.* **2017**, *11*, 78. [CrossRef] [PubMed]
63. Kar, S.; Bhagat, M.; Routray, A. EEG signal analysis for the assessment and quantification of driver's fatigue. *Transp. Res. Part F Traffic Psychol. Behav.* **2010**, *13*, 297–306. [CrossRef]
64. Noshadi, S.; Abootalebi, V.; Sadeghi, M.T.; Shahvazian, M.S. Selection of an efficient feature space for EEG-based mental task discrimination. *Biocybern. Biomed. Eng.* **2014**, *34*, 159–168. [CrossRef]
65. Frederick, J.A. Psychophysics of EEG alpha state discrimination. *Conscious. Cogn.* **2012**, *21*, 1345–1354. [CrossRef] [PubMed]
66. Dietrich, A.; Kanso, R. A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol. Bull.* **2010**, *136*, 822–848. [CrossRef]