Unlocking the potential of immersive technologies: a multi-method approach to evaluating the consumer experience

Quentin Sellier, Ingrid Poncin, Corentin Vande Kerckhove, Jean Vanderdonckt

ABSTRACT

In the battle to provide the best consumer experience, immersive technologies appear as powerful tools. Among these technologies, gestural interaction makes it possible to interact naturally with an interface without using a mouse or any other device. However, the effect of this technology on the consumer experience has been studied very little to date, while it is essential in order to use such interfaces appropriately in the context of e-commerce. In this study, we compare at the experience level gestural interaction to classical user interfaces, using self-reported questionnaires and electroencephalographic data. It emerges from the self-reported scales that using gestural interaction provides a more playful, immersive and stimulating consumer experience. In return for these experiential gains, it is particularly important to consider the loss of perceived control for consumers. These results converge between self-reported subjective data and objective data measured from the brain activity of the participants.

1. Introduction

It is now essential for businesses to provide the best consumer experience to ensure their competitiveness (Pine & Gilmore, 1998; Verhoef et al., 2009; Lemon & Verhoef, 2016). In this goal, technologies are constantly evolving to provide more and more immersive, rich and complete experiences. Some of these technologies originally come from the world of video games and are known for their immersive capacity. Many of them have gone beyond their initial framework to be now used in the context of e-commerce, or even in the future metaverse. One of these trending technologies is the 3D gestural interaction (Daugherty et al., 2015). This consists of the interaction between an individual, producing movements with any part of the body, and a computer system, capable of interpreting those movements.

This topic is particularly interesting as gestural interaction is considered intuitive and natural because it is based on the gestural communication that humans acquire naturally (Baudel & Beaudouin-Lafon, 1993). Yet, while we know the advantages of this type of technology on the user experience (Vanderdonckt & Vatavu, 2018), for example in the context of video games, it has not yet been studied in the context of e-commerce. It is particularly important as we should consider emotional and cognitive aspect of the consumption experience. Understanding this would enable marketers to integrate this type of technology efficiently, with the objective of providing a more complete and richer consumer experience.

In this study, we compare gestural interaction to classical user interfaces at the experience level. In order to do so, we built a digital catalog controllable by mid-air hand gestures. We then conducted a study in which the participants interact with a 3D gestural interface as well as with a classical user interface using a mouse. We compare their experiences with two different methods to triangulate our results and make our findings more robust (Van Camp et al., 2019 ; Zaki & Islam, 2021). The first method consists of self-reported subjective measures via questionnaires. The second is the use of an electroencephalographic headset, making it possible to obtain objective measurements from the brain activity of the participants. The results obtained through machine learning and statistical approaches are then presented in this paper with some recommendations.

2. Metrics and Hypotheses

In this study, we triangulate self-reported scales with EEG data to take advantage of the convergence and complementarity of these methods (Poncin & Derbaix, 2004). Brain activity consists of a set of bursts of electrical signals that travel through cells. This signal can be detected using a neural helmet that records these slight electrical signals. The electrodes of the neural helmets are capable of capturing signals distinguished into different frequencies, each of which corresponds to the state of the subject that emits them (Radüntz, 2018). The headset used in this study is a "Diadem", a dry-EEG (no gel needed) manufactured by the company Bitbrain, used to monitor emotional and cognitive states while remaining comfortable for the user. It directly computes from the brain activity four variables available to the experimenters:

1) Engagement - the degree of connection between the participant and the task
2) Valence - the degree of attraction experienced in response to a stimulus or situation
3) Memorization - the intensity of cognitive processes related to the formation of future memories
4) Workload - the focus or concentration of a participant during the experience

It should be noted that all the values of these variables depend on each participant as everyone has a different physiological response to a situation. In order to measure these responses as metrics, the EEG used in this study starts every session with a calibration phase assessing the low and the high cognitive activity of each participant. The data provided is therefore presented mainly between 0 and 100, with the possibility of going beyond in the event of particularly intense brain activity.

As part of this work, we focus on the cognitive and affective dimensions of the experience. We pay particular attention to the immersive quality of the experience (Novak et al., 2000 ; Nakamura & Csikszentmihalyi, 2014.), the playfulness (Kuts, 2009 ; van Beurden et al., 2012), pleasure, arousal and dominance (Mehrabian, 1995). We are also interested in the engagement as a consequence of the consumer experience, and the valence. Regarding the cognitive dimensions of the experience, we are interested in the perceived ergonomics (ISO 9241-400:2007), ease of use (Davis, 1989), ease of learning (Dix et al., 2003), memorization effort and workload. We therefore formulate the following 12 hypotheses available in Table 1. We expect that, when using gestural interaction instead of a classical user interface:

Table 1: Hypotheses of this study.

Tested with self-reported scales	H1: The playfulness increases H2: The level of immersion is deeper H3: The pleasure increases H4: The level of arousal increases	H5: The dominance decreases H6: The perceived ergonomics decreases H7: The ease-of-use decreases H8: The learnability decreases
Tested with the EEG	H9: The engagement increases H10: The valence increases	H11: The memorization effort increases H12: The workload increases

3. Methodology

3.1 Construction of a gestural interface

We constructed an IKEA digital catalog fully controllable by our mid-air hand gestures, based on the Large User Interface project (Sluÿters et al., 2022). The interface is controlled using the Leap Motion Controller, a small (80×30×13mm) device that connects to a computer via its USB port. The interface that enables to use this sensor is the Large User Interface (LUI), managed by The MIT Media Lab (Parthiban et al., 2022). It is aimed at allowing people to naturally interact by gesture to manipulate multimedia contents on a large display, such as a wall display or a tabletop. Initially, this interface works with a set of pre-recorded gestures. However, the system we used in this work replaced the commands by a gesture recognition algorithm selected following a benchmarking. Thereafter, a gesture elicitation study was then carried out (Wobbrock et al., 2009). The aim of this type of study is to provide a gesture recognition algorithm with enough data to function properly and to implement gestures chosen by end users rather than arbitrarily selected by the researchers. The set of gestures that resulted from this process is available in Figure 1.

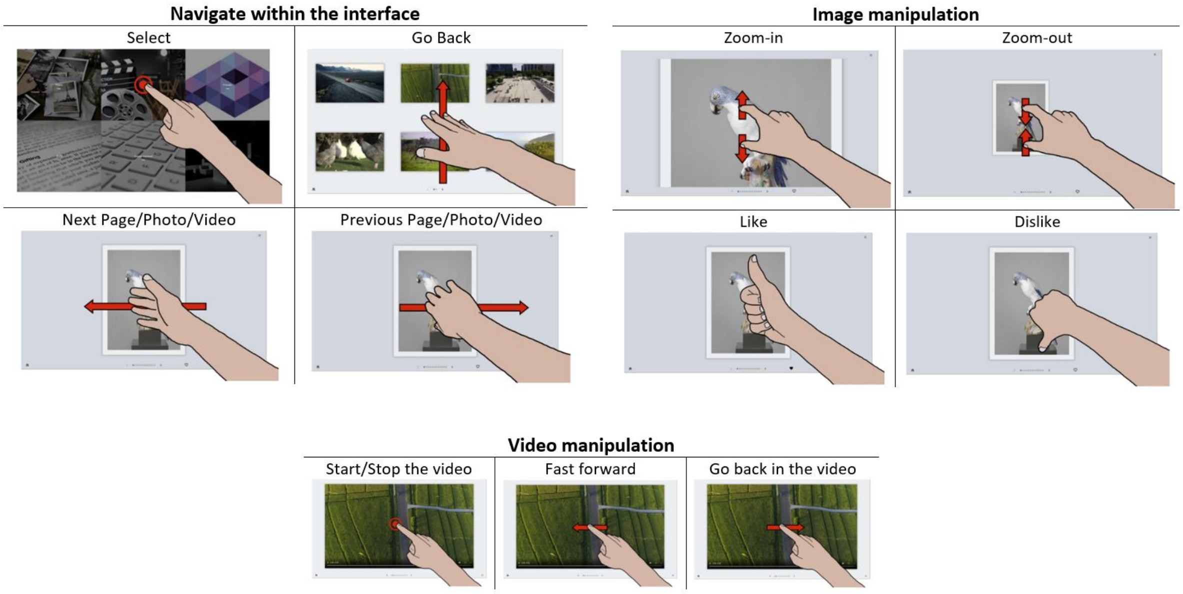

Figure 1: Illustration of each action and its assigned gesture in the interface.

The catalog is controllable by only one hand at a time, and it is possible to change the hand controlling the system in order to adapt to the dominant hand of the user. In addition to being controllable by gesture, it is possible to use a mouse. The interface has also been redesigned to look like an IKEA digital catalog.

3.2 Experimental design

This study was carried out with 39 non-students' participants (M_{age} = 36, 46% female). The experimenters install a computer screen on a desk, as well as the LeapMotion controller and a mouse. During this time, the participants complete the consent form and answer the same questions as in the first study. Once this is done, the participants are installed on a chair, in front of the computer screen, and an electroencephalographic (EEG) headset is installed on their head. The headset used in this study is a "Bitbrain", a dry-EEG (no gel needed) manufactured by the company Bitbrain, used to monitor emotional and cognitive states while remaining comfortable. After installing it on the participant's head, a calibration phase lasting a total of fifteen minutes is carried out. During this process, the participant goes through phases of high cognitive intensity and periods of rest.

Once the installation and calibration of the EEG has been completed, the experimentation can begin. The particularity of this study is that it directly compares the gestural interface with a "classic" interface using a mouse. Even if only one modality (gesture or mouse) is used at a time, each participant tries the interface in both modalities. The order is random, so some participants started with the mouse and ended with gestures, and others the opposite.

The participants begin with a video tutorial explaining how to use the catalog with a mouse or gestures. They then have to perform two tasks: (1) to go to the “videos” application and to choose a video to watch, and (2) to place three items of their choice in their shopping basket. In order to ensure that the participants selected items that they really wanted, they were informed that they had a chance to win one of the articles they had liked. Once these three steps (tutorial, video and shopping) were completed, the participants fillet out a short questionnaire about their immediate experience. The same procedure was then repeated for the second mode of interaction (gesture or mouse). As compensation, all the participants received a voucher worth €40 to redeem in an IKEA store.

4. Analyses and Results

This study uses two methods to triangulate subjective and objective data. A first interest of this approach is the comparison of the results, enabling to observe a possible convergence or complementarity when comparing the results (Poncin & Derbaix, 2004). Secondly, this approach makes it possible to affirm with more robustness the differences in terms of the consumer experience provided by gestural interaction compared to a classic user interface (Van Camp et al., 2019 ; Zaki & Islam, 2021).

4.1 Self-reported scales

In this study we adopt a within subject approach, which means that each individual participates in both conditions. We therefore performed, in SPSS, a paired samples t-test to compare mouse and gestures, with the data of the questionnaires. Since we formulated hypotheses, the results are presented in 1-tailed. Full results are available in Table 2.

Table 2: Comparison of mouse and gestures using the questionnaire (paired samples t-test).

Variable	Interaction	Mean	Std. Deviation	t	df	Sig. (1-tailed)	Decision
Playfulness	Mouse	4.13	1.24	-3.273	38	<.001	H1 accepted
	Gesture	5.11	1.24				
Immersion	Mouse	4.16	1.22	-2.212	38	.017	H2 accepted
	Gesture	4.63	1.36				
Pleasure	Mouse	5.06	1.32	.745	38	.231	H3 rejected
	Gesture	4.84	1.39				
Arousal	Mouse	3.65	1.29	-4.475	38	<.001	H4 accepted
	Gesture	4.94	1.19				
Dominance	Mouse	5.37	1.31	1.999	38	.027	H5 accepted
	Gesture	4.83	1.27				
Ergonomics	Mouse	5.83	1.40	4.152	38	<.001	H6 accepted
	Gesture	4.21	1.76				
Ease-of-use	Mouse	5.96	1.38	4.862	38	<.001	H7 accepted
	Gesture	4.22	1.59				
Learnability	Mouse	6.41	0.99	3.709	38	<.001	H8 accepted
	Gesture	5.32	1.44				

Based on these results, we know that using gestural interaction rather than a classical user interface significantly improves the playfulness, enables a deeper level of immersion, and provides a higher arousal. In return for these gains in experience quality, we have to pay attention to the ergonomics, ease of use, ease of learning and feeling of control, all of which decrease significantly when adopting gestural interaction. It is especially important as consumers desire to exercise control at all stages of the service process (Bateson, 1985; Van Raaij & Pruyn, 1998).

We also observe that there is no significant effect on the pleasure dimension of the emotional state. When analyzing the data, we notice individual disparities in the difference in pleasure between gestural and mouse interaction. This could be explained by the different nature of these two methods of interaction, the gesture being new while the mouse has been known and used by many people for a longer time. There could therefore be an impact of the participant's ease with technologies, itself correlated with the age, on their level of pleasure after having interacted with this new technology. In order to verify this, we performed a correlation analysis using the Pearson coefficient in SPSS. As expected, there is no correlation between the age of the participant and his level of pleasure with the mouse interaction. However, there is a negative correlation between the age of the participant and his level of pleasure with the gestural interaction, $r(37) = -.445$, $p = .002$. Therefore, the older the person interacting with gestures, the lower their level of pleasure. This can be interpreted by a greater difficulty for older people to use this new mode of interaction, or simply the initial expectation that it will be more difficult, which can also impact the level of pleasure (Mehrabian, 1996).

The main risk of this type of technology is therefore to give consumers the feeling of being rather controlled by the system (Coutrix & Mandran, 2012), which tends to provide a poorer experience. This is particularly problematic in the case where pleasure is absent and the arousal high, which is a situation where a feeling of anxiety may appear (Mehrabian, 1996). To avoid this situation, consumers should not be left on their own when discovering such interfaces, and especially older ones. The integration of this type of technology in an e-commerce situation should therefore go hand in hand with an easy and effective learning process.

4.2 Electroencephalographic data

In addition to the questionnaires, we have the electroencephalographic data of the 39 participants, all of whom interacted with the mouse and gestures. For each interaction, the EEG headset provides a measurement every 250ms, for each of the four variables. The main difficulty is to objectively find a method and metrics to analyze these data. It is especially problematic since this EEG headset is a relatively recent product for which few studies have been published. Therefore, we will analyze this data in two different ways. The first will follow a machine learning approach and will allow us to objectively define the aspects of this data to consider for the analyses. The goal is to train a model to predict if an interaction was performed using gestures or mouse, just based on the EEG data of the participant. If it works, it means that there is information in this EEG data that enables to differentiate gestures and mouse. We can then ask the system how it took its decisions and identify the key aspects of the data to consider for the analyses. We will then move on to a second phase following a more traditional statistical approach. It will allow us to interpret these data and compare them to our hypotheses.

Exploration of the data

The first step in such an approach is the data exploration. This was done by first visualizing the data individually, as illustrated in Figure 2. We try to compare gesture and mouse signals for the same variables and individuals. Our goal is to see if it is possible to see any differences in these data with the human eye. We also visualized the aggregate distribution of these data (cf. Figure 3), which allowed us to see the major values and potential outliers.

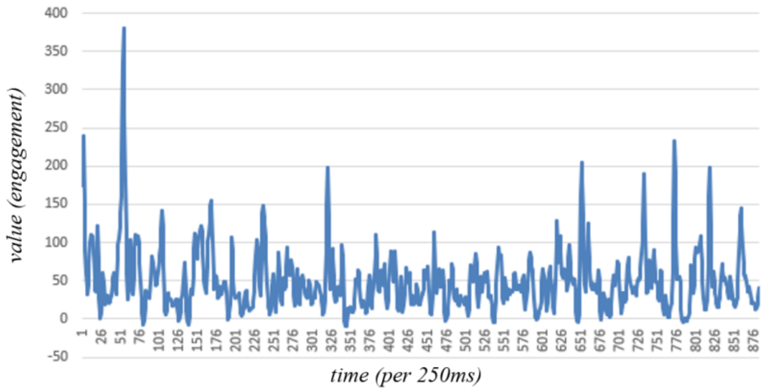

Figure 2: Example of data visualization for one individual and one variable.

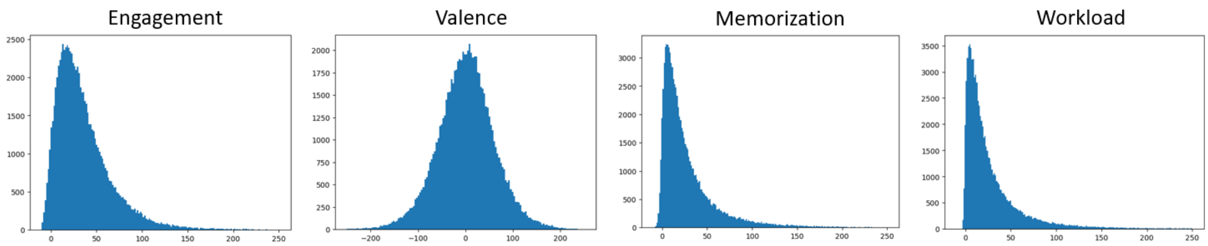

Figure 3: Distribution of the four variables of the EEG.

In addition to this, we computed summary statistics of the dataset (cf. Table 3) as well as specific metrics such as the mean during specific moments of the interaction (cf. Table 4). The objective is again to note if it is possible to see with the human eye potential elements allowing to differentiate the gestural interaction from the classic user interface.

Table 3: Summary statistics of the dataset.

Variable	**Median** of mouse	**Median** of gestures	**Mean** of mouse	**Mean** of gestures	**Std**. of mouse	**Std**. of gestures
Engagement	29.6	28.6	35.4	35.3	6.3	5.4
Memorization	17.7	18.1	24.4	26.7	5.3	4.7
Valence	0.1	2.5	-0.1	2.1	7.5	7.4
Workload	17.4	17.3	24.7	27.5	4.7	4.4

Table 4: Part of participants where the brain activity increases when switching from mouse to gesture.

Variable	**Mean** of the whole interaction	**Max** peak	Mean of the **first second** of interaction	Mean of the **first ten seconds** of interaction	Mean of the **last second** of interaction	Mean of the **last ten seconds** of interaction
Engagement	46%	56%	28%	46%	41%	38%
Memorization	54%	64%	59%	36%	44%	49%
Valence	59%	69%	54%	49%	46%	56%
Workload	46%	62%	62%	41%	38%	41%

Based on the peak-end rule (Do et al., 2008), we know that the most important moments of interaction impacting the experience are the most intense and last moments of interaction. In our experimental design, looking at the first and last moment turned out to be irrelevant as participant take part in both conditions and stay with the interface in front of them. We also found that computing means and medians of the whole interaction smooths the data and gives no result. The most significant observations we made were by looking at the peaks of activity. We therefore decided to focus on the most intense moment of interaction in two different ways: by looking at (1) the highest peak of brain activity and at (2) the proportion of high brain activity, for each participant, interaction session and variable.

A final observation that allowed us to make this exploratory analysis is to notice that we have 61% of the total interaction time with gestures (39% with the mouse). Since no time constraint was given to the participants, they tended to use the interface longer using gestures rather than the mouse. This is an important aspect that will be considered when computing metrics.

Pre-processing of the data

Before analyzing the data, we have to perform some pre-processing. The EEG headset records brain activity throughout the duration of the experiment, even when the tasks have not yet started or when the participants answer the questionnaires. Therefore, we filter the data to keep only the times when the participant interacts with the interface for the shopping task. Then, in order to avoid noise, we apply an “interquartile range” filter technique to detect outliers (Tukey, 1977 ; Vinutha et al., 2018). This technique is particularly relevant when the distribution of the data is skewed on one side (Walfish, 2006). As the data distribution in Figure 3 shows, this is mostly in our case. Since we are interested in the highest peak of activity of individuals, we are particularly sensitive to noise, and we must therefore manage the outliers. In addition, these results will then be used following a classical statistical approach, where it is traditionally required to consider outliers.

The interquartile range technique computes the distance between the 25th percentile (Q1) and 75th percentile (Q3) and defines two boundaries based on this distance. If a value is outside these boundaries, it is considered an outlier:

- InterQuartile Range (IQR) = Q3 – Q1
- Lower Boundary = Q1 – (1.5 * IQR)
- Upper Boundary = Q3 + (1.5 * IQR)

However, if the outliers show extreme deviation in comparison to the boundary, as shown in Table 5, we should replace the inner boundaries by the outer boundaries (Schwertman et al., 2004). To do so, we must replace the "1.5" with "3" in the formulas, and therefore only consider strong outliers (Santhanam & Padmavathi, 2014):

- Lower Boundary = Q1 – (3 * IQR)
- Upper Boundary = Q3 + (3 * IQR)

Table 5: Highest values compared to the inner and outer upper boundaries.

Variable	Mean **highest value** when keeping outliers	Mean inner **boundary** value (with **1.5** in formula)	Mean upper **boundary** value (with **3** in formula)
Engagement	269	101	152
Memorization	288	**77**	118
Valence	215	149	259
Workload	497	**79**	121

This is particularly relevant in our case because by using "1.5" we often establish an upper boundary of less than 100. This leads us to consider perfectly plausible values as outliers, which is not the case when using "3" instead. However, for the sake of making our results more robust, we carried out the same procedure which will follow also using "1.5". This enabled us to see that our results were similar and that the choice of technique does not significantly impact the performance of our model.

Once outliers are detected, they must be managed. For this, we have two main options: delete them, which is not recommended for small datasets like ours, or impute them, i.e. replace them with other values (Toka & Çetin, 2014 ; Santhanam & Padmavathi, 2014). We favored this second method, also called trimming in machine learning literature. Therefore, when a value crosses a boundary, it is replaced by the value of the boundary.

Computation of the metrics

In order to base our choice of features on our intuition, we focus on two main metrics: (1) the highest peak of brain activity and (2) the proportion of high brain activity during an interaction session. Since we have more interaction time with the gestures than with the mouse, we first have to take that into account as it can be problematic. Since we are interested in the maximum peak of brain activity, we assume that interacting longer increases the probability of seeing a high activity peak appear. To account for this, we correct the dataset by randomly removing data points from longer sessions so that the durations are equal for each participant. Once this is done, we store, for each participant and each variable, the value of the peak of activity for the gestures and the mouse. Given the random aspect of this technique, we repeat this process 5,000 times (corresponding to the stabilization of the results) and compute an average of these peaks. This average is the first measure that we will use in the following.

For our second metric, the proportion of high brain activity, it is not necessary to correct the durations of the sessions as we are interested in a proportion of time. The main challenge is to choose an objective threshold defining what constitutes a high activity. We have therefore chosen to consider that any value exceeding 100, which is the threshold defined by Bitbrain for this EEG headset, can be considered as a particularly high brain activity for the individual.

Supervised classification approach

We could ask ourselves why opt for this method rather than a simple discriminant analysis. The main problem is that this type of analysis of classical statistics requires respecting many statistical conditions. In the case of small samples like ours, or with this type of data, it becomes difficult to respect all these conditions to carry out this type of test. We therefore favor the much less restrictive machine learning approach.

Our approach consists in training a model to interpret a series of variables, called "features", in order to classify an observation with a given label. In our case, we have 8 features, namely the 4 variables of the EEG headset all having the computation of their highest peak and the proportion of high values. Our goal here is to know whether an interaction session was performed using gestures or a mouse, which are our two classification labels. Finally, our dataset is composed of 78 observations, namely the 39 participants who all had a session with gestures and another with the mouse.

We prefer to compute a single model including all aspects of our variables rather than 4 separate models including the 2 features per variable. By doing so, we increase our chances of creating a performing model that will allow us to say whether it is relevant to be interested in this aspect of the data. Moreover, it will give us a first indication of the relative importance of these different variables compared to each other.

We first have to choose a classification model, which is the method that will classify an interaction as gesture or mouse. In our case, we are in a supervised machine learning problem, and we want to be able to understand how the model takes its classification decisions (Kotsiantis et al., 2007). We therefore focus on the random forest model (Svetnik et al., 2003), considered as a relevant tradeoff between its performances (Hasan et al., 2014) and the explanatory abilities. This model works with multiple decision trees, each constructed by a different bootstrap sample from the dataset. When a new observation arrives, it cycles through each of the trees and receives a vote that indicates the tree's decision. The model then chooses the decision with the most votes.

Concretely, the system first randomly splits the dataset into 2 parts, one used to train the model and another to test its performances. In our case, we have 80% of the dataset that becomes the training set and the remaining 20% the test set. Once it is trained on the first set, the model reads the values of the features of a new observation coming from the test set and gives a classification prediction. Once all the test set is completed, we compare the classification predictions with the actual values, which allows us to compute model performance metrics. Regarding the hyperparameters of the random forest model, we left the default values without trying to fine-tune them since our dataset is relatively small.

In order to stabilize the results, given the randomness of the sampling method and of the random forest model, we repeat this process 100,000 times. This corresponds to the stabilization of the results, namely that their values did not move more than 0.1% during 5 computations. This allows us to obtain models with an accuracy of 72%, which is the number of correct predictions divided by the total number of predictions. The full confusion matrix is available in Table 6.

Table 6: Confusion matrix of our model.

		Actual values	
		Gesture	**Mouse**
Predicted values	**Gesture**	35.2%	13.2%
	Mouse	14.8%	36.8%

This means that, when the system receives new brain activity data from a person, it can tell while being 72% right whether that person used gestures or the mouse. These results are a significant improve from a random model that would have 50% accuracy. Another baseline model to compare with is the model using the means or medians of the EEG signal, that each have approximately 51% accuracy. This allows us to affirm that the two metrics relating to high brain activity contain relevant information to differentiate gestures and mouse.

Another advantage of the machine learning approach is that we can now ask the random forest model the relative importance of the features it used to make its classification decisions. These are detailed in Table 7.

Table 7: Importance of the features used in the random forest model (with 8 features).

Feature	**Importance**	**Aggregated importance**
Highest engagement peak	9.5%	18.3%
Proportion of high engagement	8.8%	
Highest valence peak	12.3%	23.1%
Proportion of high valence	10.8%	
Highest memorization peak	9.5%	26.1%
Proportion of high memorization	16.6%	
Highest workload peak	9.0%	32.5%
Proportion of high workload	23.5%	

We can see that the most useful variable for the model to make its classification decision is the workload, and more specifically the proportion of high workload. It is followed by the memorization, also placing more importance on the proportion rather than the highest peak. The trend is reversed for the valence and engagement, where the highest peak of activity is considered more important than the proportion of high activity throughout the session. It should also be noted that the engagement only accounts for 18.2% of the weight of the decision, which means that it is either less useful or harder for the model to interpret it. In total, the highest activity peaks represent 40.3% of the weight of the decision against 59.7% for the proportion of high activity.

The next and final step is the simplification of the model. It will make it possible to gain even more in accuracy and to choose among these 8 features which ones to keep for the classic statistical approach. To do this, we remove the least important feature from the model and compute it again. We then systematically repeat the same procedure until only one feature remains. In order to preserve as much as possible the information of the 4 variables, if two less important features have a similar score, we prefer to keep the one allowing to guarantee at least one feature per variable. The performance metrics of all these simplified models are detailed in Table 8.

Table 8: Performance metrics of the simplified models.

Nb. of features	Accuracy	Precision	Recall	F-score
8 features	72.1%	73.9%	72.0%	70,1%
7 features	72.6%	74.2%	72.2%	71,4%
6 features	72.7%	74.5%	71.8%	71,3%
5 features	74.5%	76.1%	73.6%	73,2%
4 features	**77.4%**	**78.5%**	**76.8%**	**76,3%**
3 features	75.7%	76.0%	77.0%	75,1%
2 features	**77.1%**	**78.5%**	**76.1%**	**75,8%**
1 feature	60.8%	62.8%	60.4%	58,7%

The most efficient model is therefore the one with 4 features. The level of importance of these features is detailed in Table 9. In addition to being more precise, this allows us to identify the most important aspect of the 4 variables.

Table 9: Importance of the features used in the random forest model.

Feature	Importance (4-features model)	Importance (2-features model)
Highest engagement peak	17.0%	-
Proportion of high engagement	-	-
Highest valence peak	22.0%	43,7%
Proportion of high valence	-	-
Highest memorization peak	-	-
Proportion of high memorization	24.7%	-
Highest workload peak	-	-
Proportion of high workload	36.3%	56,3%

It should also be noted that the model comprising 2 features also obtains high-performance scores. What is particularly interesting is the fact that it is a combination of a higher peak and a high activity. We might therefore ask ourselves if the most important thing is to combine workload and variance, or if any combination would work equally well. To verify this, we tested the 16 possible combinations of models including 2 features (with a combination of max peak and proportion of high activity). The accuracy of each of these models is detailed in Table 10.

Table 10: Accuracy of the 16 combinations of 2-features models

		Proportion of high activity			
		Engagement	Valence	Memorization	Workload
Max peak	Engagement	48%	51%	66%	70%
	Valence	45%	50%	65%	77%
	Memorization	42%	49%	61%	63%
	Workload	51%	50%	62%	68%

It turns out that the variables play an important role. Thus, using the proportion of high engagement or valence only gives poor accuracy scores. The best score that stands out in particular is the one that had been previously identified, namely the proportion of high workload combined with the highest valence peak.

Classic statistical approach

The metrics we used were exported into SPSS to perform paired samples t-tests to compare gestures and the mouse. As with the questionnaires, since we formulated hypotheses, the results are presented in 1-tailed. Table 11 gives the results using the average highest peaks of brain activity.

Table 11: Comparison of mouse and gestures using the highest brain activity peaks.

Variable	**Interaction**	**Mean**	**Std. Deviation**	**t**	**df**	**Sig. (1-tailed)**	**Decision**
Engagement	Mouse	142.8	23.3	2.969	38	.003	H9 accepted
	Gesture	150.0	21.3				
Valence	Mouse	192.3	31.4	3.192	38	<.001	H10 accepted
	Gesture	211.1	38.2				
Memorization	Mouse	114.1	22.4	3.866	38	.001	H11 accepted
	Gesture	118.4	23.7				
Workload	Mouse	117.7	22.5	3.978	38	<.001	H12 accepted
	Gesture	120.9	23.0				

These results go in the direction of the validation of our last four hypotheses. However, we also have to perform the same analyses using the second metric, the proportion of high brain activity. The results are presented in table 12.

Table 12: Comparison of mouse and gestures using the proportion of high brain activity.

Variable	**Interaction**	**Mean**	**Std. Deviation**	**t**	**df**	**Sig. (1-tailed)**	**Decision**
Engagement	Mouse	.0365	.0247	0.772	38	.222	H9 rejected
	Gesture	.0409	.0268				
Valence	Mouse	.0446	.0168	1.871	38	<.001	H10 accepted
	Gesture	.0514	.0181				
Memorization	Mouse	.0163	.0168	4.686	38	.035	H11 accepted
	Gesture	.0337	.0244				
Workload	Mouse	.0180	.0156	6.695	38	<.001	H12 accepted
	Gesture	.0433	.0280				

In this case, the hypotheses H10, H11 and H12 are still accepted, but H9 is not anymore. The difference in engagement is no longer statistically significant. We can interpret this by the fact that there are individual differences in terms of engagement with the task, especially since our dataset is relatively small. Yet, when we look at the most impactful event causing the peak of engagement, the difference is statistically significant. This is consistent with the results of the random forest model, which placed more importance on the highest peak of engagement rather than the proportion of high engagement throughout the session. According to the peak-end rule (Do et al., 2008) and the results of our machine learning approach, we could put more importance on the peak of activity rather than the proportion throughout the session. In this case, we could conclude that the engagement is significantly higher using gestures rather than the mouse, and therefore validate H9. However, this choice is debatable and above all shows that the effect on the engagement requires further investigation in another study.

Regarding the other three variables, the results converge between the analyses with the two different metrics. When we interact with a gestural interface rather than a classical user interface, our valence increases significantly, meaning that we are more connected with the task. On the other hand, the memorization effort and the cognitive workload each are significantly more intense. This also converges with the results obtained with the self-reported scales. There are therefore affective experiential gains offset by an increase in the cognitive difficulty of controlling this type of new technology.

5. Conclusion and Implications

To understand the effect of gestural interaction technologies on the consumer experience, we set up a study in which participants interact with an IKEA digital catalog fully controllable with gestures. We directly compare this gesture interaction with a classical user interface using questionnaires and an EEG headset. This enables us to triangulate our data and enrich the understanding of what this difference implies on the consumer experience.

We have seen that businesses can use gestural interaction to provide a more playful, immersive and stimulating consumer experience. On the other hand, we must not neglect the loss in user's feeling of control, which will directly influence the consumption and the experience. This is especially the case for older people. We therefore recommend to set up a complete and easy learning of the interface, for example via a tutorial as in this study or via personalized support. Finally, we have shown that the participants' subjective declarations in terms of their experiences correspond to the objective measures observed through the EEG (Van Camp et al., 2018).

It should be kept in mind that this study present limitations. As a first one, we adopt a diachronic point of view. In a future work, it would be particularly interesting to study the experience over time. We could perhaps observe a learning effect, where control becomes easier with time and experience. Another possibility is the gradual disappearance of the novelty effect, possibly influencing the playfulness or other experiential factors. It would also be the opportunity to study the consumer experience through multiple touchpoints of the customer journey. A final limitation is that we performed this study in a controlled lab environment. In order to observe the consumer experience in an ecological context, il would be relevant to test the same setup when completing real online purchases.

This work contributes to the current literature by providing a comparison of different interaction technologies and what it implies in terms of consumer experience. Regarding the methodological contributions, we provide insights into the use of Bitbrain-style EEG data. We also bring the dual approach method starting with machine learning before moving to classical statistics. The first approach allows us to identify the key aspects of the data to study before interpreting them via the second approach. From a managerial point of view, we provide guidelines for the use of this type of new technology in an e-commerce context. More specifically, we study how this type of technology can improve the consumer experience, in an environment where it is crucial to stay one click ahead of the competition.

References

Bateson, J. E. (1985). Self-service consumer: An exploratory study. *Journal of Retailing*, *61*(3), Article 3.

Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: Remote control of objects using free-hand gestures. *Communications of the ACM*, *36*(7), Article 7. https://doi.org/10.1145/159544.159562

Daugherty, P. R., Schybergson, O., & Wilson, H. J. (2015, July 8). Gestures Will Be the Interface for the Internet of Things. *Harvard Business Review*. https://hbr.org/2015/07/gestures-will-be-the-interface-for-the-internet-of-things

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), Article 3. https://doi.org/10.2307/249008

Dix, A., Dix, A. J., Finlay, J., Abowd, G. D., & Beale, R. (2003). *Human-computer Interaction*. Pearson Education.

Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review*, *15*(1), 96–98. https://doi.org/10.3758/PBR.15.1.96

Hasan, M. A. M., Nasser, M., Pal, B., & Ahmad, S. (2014). Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS). *Journal of Intelligent Learning Systems and Applications*, *2014*. https://doi.org/10.4236/jilsa.2014.61005

Holbrook, M. B. (2006). Consumption experience, customer value, and subjective personal introspection: An illustrative photographic essay. *Journal of Business Research*, *59*(6), Article 6. https://doi.org/10.1016/j.jbusres.2006.01.008

Holbrook, M. B., & Hirschman, E. C. (1982). The Experiential Aspects of Consumption Consumer Fantasies, Feelings, and Fun. *Journal of Consumer Research*, *9*(2), Article 2. https://doi.org/10.1086/208906

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), 3-24.

Kuts, E. (2009). *Playful User Interfaces: Literature Review and Model for Analysis*. DiGRA.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), Article 6. https://doi.org/10.1509/jm.15.0420

Mehrabian, A. (1995). Framework for a comprehensive description and measurement of emotional states. *Genetic, Social, and General Psychology Monographs*, *121*(3), Article 3.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology*, *14*(4), Article 4. https://doi.org/10.1007/BF02686918

Pine, B. J., & Gilmore, J. H. (1998, July 1). Welcome to the Experience Economy. *Harvard Business Review*, *July–August 1998*, Article July–August 1998. https://hbr.org/1998/07/welcome-to-the-experience-economy

Poncin, I., & Derbaix, C. (2004). *Post exposure verbal measurement is not so bad: Convergence and complementarity of three methods of affective reactions' measurement*. 33rd EMAC Conference. https://dial.uclouvain.be/pr/boreal/object/boreal:18869

Raaij, W. F. V., & Pruyn, A. T. H. (1998). Customer control and evaluation of service validity and reliability. *Psychology & Marketing*, *15*(8), Article 8. https://doi.org/10.1002/(SICI)1520-6793(199812)15:8<811::AID-MAR6>3.0.CO;2-8

Radüntz, T. (2018). Signal Quality Evaluation of Emerging EEG Devices. *Frontiers in Physiology*, *9*. https://www.frontiersin.org/articles/10.3389/fphys.2018.00098

Santhanam, T., & Padmavathi, M. S. (2014). Comparison of K-Means clustering and statistical outliers in reducing medical datasets. *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, 1–6. https://doi.org/10.1109/ICSEMR.2014.7043602

Schwertman, N. C., Owens, M. A., & Adnan, R. (2004). A simple more general boxplot method for identifying outliers. *Computational Statistics & Data Analysis*, *47*(1), 165–174. https://doi.org/10.1016/j.csda.2003.10.012

Sellier, Q., Poncin, I., & Vanderdonckt, J. (2021). *User, Customer and Consumer Experience: Highlighting the Heterogeneity in the Literature*. 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP. https://dial.uclouvain.be/pr/boreal/object/boreal:239741

Sluÿters, A., Sellier, Q., Vanderdonckt, J., Parthiban, V., & Maes, P. (2022). Consistent, Continuous, and Customizable Mid-Air Gesture Interaction for Browsing Multimedia Objects on Large Displays. *International Journal of Human–Computer Interaction*, 1–32. https://doi.org/10.1080/10447318.2022.2078464

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1947–1958. https://doi.org/10.1021/ci034160g

Toka, O., & Çetin, M. (2016). *Imputation and Deletion Methods Under The Presence of Missing Values and Outliers: A Comparative Study*.

Tukey, J. W. (1977). Exploratory data analysis (Vol. 2, pp. 131-160).

van Beurden, M. H. P. H., Ijsselsteijn, W. A., & de Kort, Y. A. W. (2012). User Experience of Gesture Based Interfaces: A Comparison with Traditional Interaction Methods on Pragmatic and Hedonic Qualities. In E. Efthimiou, G. Kouroupetroglou, & S.-E. Fotinea (Eds.), *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication* (pp. 36–47). Springer. https://doi.org/10.1007/978-3-642-34182-3_4

Van Camp, M., De Boeck, M., Verwulgen, S., & De Bruyne, G. (2019). EEG Technology for UX Evaluation: A Multisensory Perspective. In H. Ayaz & L. Mazur (Eds.), *Advances in Neuroergonomics and Cognitive Engineering* (pp. 337–343). Springer International Publishing. https://doi.org/10.1007/978-3-319-94866-9_34

Vanderdonckt, J., & Vatavu, R.-D. (2018). Designing, Engineering, and Evaluating Gesture User Interfaces. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–4. https://doi.org/10.1145/3170427.3170648

Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, *85*(1), Article 1. Scopus. https://doi.org/10.1016/j.jretai.2008.11.001

Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In S. C. Satapathy, J. M. R. S. Tavares, V. Bhateja, & J. R. Mohanty (Eds.), *Information and Decision Sciences* (pp. 511–518). Springer. https://doi.org/10.1007/978-981-10-7563-6_53

Walfish, S. (2006). A review of statistical outlier methods. Pharmaceutical technology, 30(11), 82.

Zaki, T., & Islam, M. N. (2021). Neurological and physiological measures to evaluate the usability and user-experience (UX) of information systems: A systematic literature review. *Computer Science Review*, *40*, 100375. https://doi.org/10.1016/j.cosrev.2021.100375