A physical learning companion for Mental-Imagery BCI User Training

Léa Pillette, Camille Jeunet, Boris Mansencal, Roger N'Kambou, Bernard N'Kaoua, Fabien Lotte

HAL Id: hal-02434157
https://inria.hal.science/hal-02434157

Submitted on 9 Jan 2020

A physical learning companion for Mental-Imagery BCI User Training

L. Pillette[1, 2], C. Jeunet[3], B. Mansencal[2], R. N'Kambou[4], B. N'Kaoua[5], F. Lotte[1, 2]

[1]Inria Bordeaux Sud-Ouest, France
[2]LaBRI (Univ. Bordeaux, CNRS, Bordeaux-INP), France
[3]CLLE (Univ. Toulouse, CNRS), France
[4]GDAC, UQAM, Quebec, Canada
[5]Handicap, Activity, Cognition, Health, University of Bordeaux, France

E-mail: lea.pillette@inria.fr

Abstract

Mental-Imagery based Brain-Computer Interfaces (MI-BCI) present new opportunities to interact with digital technologies, such as wheelchairs or neuroprostheses, only by performing mental imagery tasks (e.g., imagining an object rotating or imagining hand movements). MI-BCIs can also be used for several applications such as communication or post-stroke rehabilitation. Though, their lack of reliability remains a barrier to a larger scale development of the technology. For example, one task between two is recognized on average 75% of the time. It has been shown that users are more likely to struggle using MI-BCIs if they are non-autonomous or tensed. This might, at least in part, result from a lack of social presence and emotional support, which have yet very little been tested in MI-BCI, despite recommendations from the educational literature. One way to provide such social and emotional context is by using a learning companion. Therefore, we designed, implemented and evaluated the first learning companion dedicated to the improvement of MI-BCI user training. We called this companion PEANUT for Personalized Emotional Agent for Neurotechnology User Training. PEANUT provided social presence and emotional support, depending on the performance and progress of the user, through interventions combining both pronounced sentences and facial expressions. It was designed based on the literature, data analyses and user-studies. We notably conducted various online user surveys to identify the desired characteristics of our learning companion in terms of appearance and supporting speech content. From the results of these surveys we notably deduced which should be the characteristics (personal/non-personal, exclamatory/declarative) of the sentences to be used depending on the performance and progression of a learner. We also found that eyebrows could increase expressiveness of cartoon-like faces. Then, once this companion was implemented, we evaluated it during real online MI-BCI use. We found that non-autonomous people, i.e., who are more inclined to work in a group, that are usually disadvantaged when using MI-BCI were advantaged compared to autonomous people when PEANUT was present with an increase of 3.9% of peak performances. Furthermore, in terms of user experience, PEANUT seems to have improved how people felt about their ability to learn and memorize how to use an MI-BCI by 7.4%, which is a dimension of the user experience we assessed.

Keywords: Mental-Imagery based Brain-Computer Interface, Learning Companion, Social feedback, Emotional feedback, User experience

Introduction

Brain-computer Interface (BCI) enable their users to send commands to digital technologies using their brain-activity alone, often recorded using electroencephalography (EEG) [65]. One of the most commonly used type of BCI is Mental-Imagery based BCI (MI-BCI) which we will focus on in this article. Such BCIs are controlled by their users by performing mental-imagery (MI) tasks, such as imagining objects rotating or performing mental calculation. A famous example of MI-BCI is a smart wheelchair that is controlled by imagining left or right hand movements, e.g., imagining waving at someone, to make the wheelchair turn respectively left or right [6]. MI-BCI applications are broad because they provide new interaction tools. For example, they can also be used to write by controlling a speller [64] or to foster brain plasticity and improve motor rehabilitation for post-stroke patients [3].

All the MI-BCI applications rely, on their ability to send the correct command, i.e., the one selected by the user, to the system. However, the accuracy still has to be improved for the technology to undergo a strong growth outside of research laboratories. For

example, when the system has to decide which task the user is performing between two, e.g., imagining a right versus a left hand movement, on average the system is mistaken once every four guesses [1]. There are several lines of research aiming at improving the efficiency of MI-BCIs. A great deal of them focus on improving the acquisition and processing of the brain activity [42]. However, MI-BCI applications also rely on users themselves. Indeed, on the one hand, the computer has to learn to discriminate the different brain-activity patterns for the tasks performed by a user. But on the other hand, the user has to train and learn how to produce a stable and distinguishable brain-activity pattern for each of the tasks in order for them to be recognized by the computer [42].

When being asked to imagine hand movements, users can adopt a great variety of strategies, e.g., imagining waving at someone or playing the piano. During the training, users have to find their own strategies, i.e., characteristics of mental imagery, which make the system recognize these tasks as correctly as possible. However, the adequacy of the feedback provided during the training has been questioned both by the theoretical literature [40] and experimentally [24]. The inadequacy of the training and more particularly of the feedback are probably part of the reasons why MI-BCIs remain insufficiently reliable [40]. Some users are more likely to struggle when using MI-BCIs [27]. The more "tensed" and "non-autonomous" people are (based on the dimensions of the 16PF5 psychometric questionnaire [7]), and the lower their performances tend to be. "Non-autonomous" people are persons who rather learn in a social context. Yet, while educational and neurophysiological literature show the importance of a social feedback [22, 41], this aspect of feedback as well as emotional support have been neglected during MI-BCI training. Nevertheless, educational literature shows that social presence and emotional support are very important to the learning process [29]. It seems promising to assess their impact on MI-BCI training.

Learning companions, a type of intelligent tutoring system, are computer-simulated, human-like, non-authoritative and social characters meant to foster learning [8]. They have already proven their efficiency in providing social and emotional support in different learning situations [47] but have never been used for MI-BCIs. The work presented in this paper aimed at designing, implementing and testing the first learning companion dedicated to the improvement of user experience and/or user performances during MI-BCI training. We called this learning companion PEANUT for *Personalized Emotional Agent for Neurotechnology User Training* (see Figure 1).

In the following sections, we first introduce the literature related to MI-BCI and learning companions.
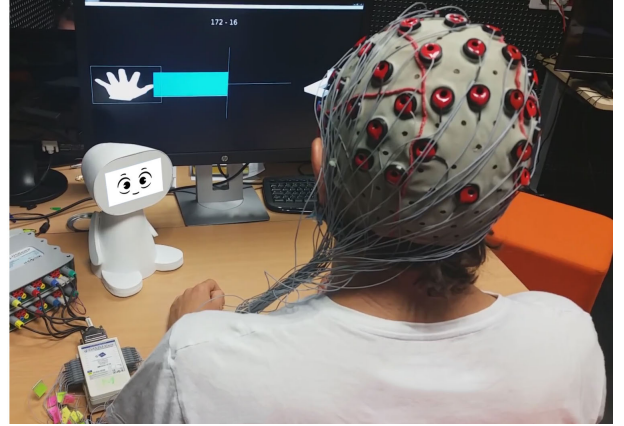

Figure 1: A participant training to use a BCI. He is learning how to perform different MI tasks (imagining a left-hand movement, performing mental calculation tasks and imagining an object rotating) to control the system. Along the training, PEANUT (on the left) provides users with social presence and emotional support, using interventions composed of facial expressions and pronounced sentences adapted to their performance and progression.

Then we describe the different steps which guided our design of the companion, starting with our main contributions regarding: (1) the design of the behavior of PEANUT, (2) the design of the physical appearance of PEANUT and (3) the implementation of PEANUT. Our design approach was carefully motivated and justified based on a review of the literature, the analysis of data from previous experiments and several user-studies. We then present the experiment which enabled us to test the adequacy of PEANUT and its characteristics for improving MI-BCI training to finally discuss these results [1].

1 Related work

1.1 MI-BCI User-Training

As their name suggests, BCIs require an interaction between a human's brain-activity and a machine [26]. Thus, the computer has to be able to understand the mental command sent by the user. In order to facilitate this process, the user must provide the system with stable brain-activity patterns each time the same MI task is performed. Brain-activity patterns from the different MI-tasks must also be distinct from one another and be consistent with the training set [1]. Like any other skill, the acquisition

---

[1]Preliminary results regarding the impact of PEANUT on MI-BCI performances and user experience were previously published in a short conference paper presented at the *7th International BCI Conference* [52]. Here we present additional and more complete results on the BCI experiments, and present for the first time the experiments and results on the design of PEANUT.

of this ability requires an adapted training process [46]. To enable the learning process, once the system has decoded the mental command sent by the user, a feedback is provided. In standard MI-BCI training protocols [50], this feedback is provided as an extending bar on the screen which indicates the user's performance i.e., which mental command is recognized and how confident the system is in its recognition (see Figure 6). The direction of the bar indicates the task which has been recognized and the length of the bar represents the confidence of the system in its recognition (i.e., the longer the bar is and the more confident the system is in its recognition).

1.1.1 Why current standard feedback is inappropriate for training MI-BCI users

While instructional design literature recommends the feedback to be, among others, explanatory/non-evaluative, multimodal and supportive [60], the standard MI-BCI feedback is evaluative/non-explanatory, unimodal and non-supportive. Thus, the fact that 15-30% of users cannot control an MI-BCI is most likely partly due to the fact that current feedback does not comply with recommendations from the literature [40] and thus does not support enough users in acquiring BCI-related skills.

1.1.2 How could the feedback be improved

MI-BCI users should be provided with an *informative feedback*. Such feedback would provide an explanation to the users about how they should change their strategies, e.g., imagining a right hand waving or playing the piano, for the system to recognize them as well as possible. There is still a lack of theoretical cognitive model providing information about the traits (e.g., computer anxiety) and states (e.g., motivation) of MI-BCI users which influence their performances, and how these characteristics interact [25, 34]. Neither do models providing an explanation on why a given mental-imagery task performed by a user is recognized or not exist. Therefore, providing an informative feedback remains a challenge.

Regarding feedback *multimodality*, it has been shown that providing feedback on other channels in addition, such as auditory feedback [16], or in comparison to the visual one, such as tactile feedback [28], seems to improve MI-BCI performance. Such results could be caused by a competition between the visual feedback monitoring task and the motor imagery task for the limited visual-related cognitive resources. Also, standard visual feedback requires too many resources to be processed [24]. The visual channel being often overtaxed in interactive situations, this feedback overloads users' cognitive resources, leading to a decrease in performance. Providing the feedback on another channel may enable this overload to be prevented, thus helping to achieve better performances

[28].

Finally, to our knowledge, the *supportive* dimension, that includes social presence and emotional support, has been very little formally investigated in the context of MI-BCI user-training. Bonnet, Lotte, Lécuyer showed that when playing a MI-BCI video game, a 2 interacting players condition improved the user experience, in particular fun and motivation, compared to a single-user condition. It could even improve the performance of the best-performing participants [5]. This reinforces the idea that a social presence is useful in MI-BCI. Two studies used simple smiley faces as feedback [36, 37] to maintain motivation along the MI-BCI training. While associated with good performance and user experience, neither of these studies offered a formal comparison with the standard feedback to prove their efficiency. However, in a neurofeedback study, i.e., people learning how to self-regulate a brain-activity, Mathiak et al. have shown that using an avatar with a variating smile's width was more efficient than a typical moving bar to control a specific brain activity, i.e., the activation of the dorsal anterior cingulate cortex, monitored using functional Magnetic Resonance Imaging [41]. Providing emotional support and social presence seems to be a very promising approach for improving MI-BCI training both in terms of performance and user experience. Indeed, MI-BCI users perform their training alone, in front of a computer for often an hour or so. They lack any form of support, which is essential for maintaining motivation and acquiring skills [45].

1.2 Using Learning Companions to Provide Learners with Social Presence & Emotional Support

Emotions have a significant impact on learning [45]. For instance, positive emotions, induced by emotional support, can result in increased creativity and flexibility during a problem solving task [21]. Furthermore, a social reward (i.e., positive appreciation) can be considered just as much a reward as a monetary one [22]. Some distance learning systems propose the use of *Learning Companions* to address the lack of social presence and emotional support [57]. For instance, *DragonBot* is a learning companion which has been used to teach children about nutrition [59]. Learning companions are always on an equal footing with the learner and they differentiate from the other types of educational agent by their non authoritative attitude [8]. We chose such type of educational agent because we still lack a cognitive model of the task (see section *1.1.2 How could the feedback be improved*). The knowledge of an agent could not be significantly higher than the one of the user. Thus, the user and the agent had to be on an equal footing. Given Nass's paradigm, learning

companions can be seen as social actors which are just as capable of influencing users as any other social actor [55, 63]. Learning companions can have a positive effect on motivation [38], interest in the task and efficiency while performing the task [33]. They can also induce emotions that favor learning, such as self-confidence [2]. While being potentially beneficial when well conceived, inappropriately designed companions can also have a detrimental impact on performance [32, 63]. For instance, discrepancies between users' expectations towards the companion and its real capacities would lead to a bad perception of the companion [48]. For example, such a situation is likely to occur when the design of the companion suggests a high level of functionalities (e.g., highly realistic companion) whereas the implemented functionalities are basic ones (e.g., no possible interaction with the learner). As a consequence, the design process of such a companion must be undertaken cautiously [63]. In the following section we will present the results of the review of the literature as well as the different user-studies we led in order to create a learning companion which would be consistent in terms of physical appearance and behavior.

2 Designing the behavior of PEANUT

As stated herein-above, theoretical knowledge is still lacking to provide informative feedback to users with an explanatory feedback. Moreover, during the training, the users are asked not to move in order to limit motor related artifacts that could create noise in the recorded brain activity. Therefore, a complex interaction between the user and the learning companion was hardly feasible. The behavior of the companion as well as its physical appearance had to be consistent. They had to reflect the limited amount of information that the learning companion would be able to provide and focus on the emotional and social feedback that we aimed at providing. As a result, PEANUT provided the user with interventions composed of both a pronounced sentence and a facial expression expressing one or two consecutive of the following emotions: Serenity, Joy, Ecstasy, Acceptance, Trust, Admiration, Distraction, Surprise, Amazement, Sadness. All of them belong to the wheel of emotions of Plutchik [53]. We mostly chose positive emotions but also selected a few negative ones. The use of negative emotions, enabled us to display two consecutive emotions with a negative one followed by a positive one to create a contrast and increase the perceived intensity of the second emotion displayed. Their use also aimed at improving the empathy towards users and improve the social feedback by reflecting the emotional state users were likely to feel in the given learning phase [43]. For ex-

ample, when the performance, and/or progress, was decreasing users might have felt sad to be failing. In such situation, the companion could exhibit sadness and then trust in order to maintain their motivation. The interventions were solely selected with respect to the MI-BCI performance and progression, which are objective measures reflecting the MI-BCI skills of users. The performance is a measure of how confident the system is in its recognition of the mental task that the user is performing. The progress corresponds to the evolution of the performance over time. In the following paragraph we consider the context as both the current performance and progress. In order to design a relevant behavior for PEANUT for a given context, different aspects had to be considered:

- Support content - Which intervention (sentence & facial expression) should the participant be provided with according to the context (performance & progression)?

- Intervention style - How should the intervention be expressed with respect to the context? When expressing an opinion, the interpretation remains subjective to the contexts and the participants [31]. For example, when hearing the sentence "You're doing good", someone could perceive it as a supportive sentence in case of improvement but in the context of a failure, it could be perceived as ironic and this interpretation is personal. Karamibekr, Ghorbani[31] have hypothesized that it could depend on the type of the sentence (e.g., exclamatory or declarative). In line with their results, we also hypothesized that the subject pronoun (e.g., second or third) used in the sentence could influence its perception. The second person would be more explicit, e.g., "You're doing good", whereas the third would be more implicit, e.g., "Results are improving". Therefore, we asked ourselves if a sentence should be exclamatory or declarative; personal (second person) or non-personal (third person) to be perceived as motivational.

- Performance and progression thresholds - To deal with the continuum of performances and progress specific to each user we chose to separate them into three levels i.e., poor/average/good for the performances and negative/neutral/positive for the progression. Therefore we needed to define thresholds to define to which category a performance or progress would belong to relatively to each participant. The relevance of the interventions depends on these thresholds.

2.1 Support Content

The support content was elaborated after a review of both the educational and the intelligent tutoring sys-

tem literature. The intervention style was selected based on a user-study. Hereafter is a list of the possible intervention categories of PEANUT, the context for which they were created, their goal and the literature justifying their use. An intervention corresponds to the association of a sentence and a facial expression (see also Figure 3 for an exhaustive description of the intervention selection rules).

- Temporal interventions are related to the temporal dimension of the experiment. They are divided into 2 categories, *Temporal-Start* and *Temporal-End*, the goal of which is to greet and say goodbye to the users, e.g., "I am happy to meet you". Both these intervention types were associated with a facial expression of *Joy* for PEANUT. They aim at providing the companion with a polite behavior which is primordial for social interactions [63].

- Effort-related intervention categories i.e., *General-Effort* and *Support-Effort*, contain sentences like "Your efforts will be rewarded". They value the efforts that are made by the participant throughout training [11]. These sentences focus on the fact that learning is the goal, and are intended to minimize the importance of current performance while promoting long-term learning [66]. More specifically, *General-Effort* and *Support-Effort* interventions are respectively adapted to negative or neutral progression and positive progression. Therefore, *General-Effort* and *Support-Effort* interventions were respectively associated with *Trust* and *Joy*.

- The category expressing empathy, i.e. *General-Empathy*, aims at letting users know that the companion understands that they are facing a difficult training process by using sentences such as "Don't let difficulties discourage you". Learning has been suggested to correlate with the amount of empathy and support received [18]. This type of intervention was preferably provided for negative or neutral progression, especially when combined with bad performance. These interventions were associated with an animation ranging from *Sadness* to *Trust*.

- Categories associated with performance/results and progression, i.e. *Results-Good*, *Results-VeryGood* and *Progress-Good*, only target positive performance and progression, e.g., "You are doing a good job!". Positive intervention regarding the performances or progress, should respectively induce positive intrinsic motivation (i.e., performing an action for its own sake) or positive extrinsic motivation (i.e., performing an action for its outcomes, e.g., grades or praise)

[49]. The sentences in this category were designed to motivate users by focusing on the positive performances and progress and therefore on the abilities users had already acquired [23]. *Results-Good* and *Results-VeryGood* were respectively associated to *Joy* and *Admiration*. *Progress-Good* was associated to an animation going from *Surprise* to *Trust*.

- The last category consisted in strategy-related interventions, i.e., *Strategy-Change* and *Strategy-Keep*, with sentences such as "You seem to have found an efficient strategy". These interventions aimed at encouraging people to keep the same strategy when progression was positive or to change strategy when it was negative/neutral. *Strategy-Keep* and *Strategy-Change* were respectively associated with *Joy* and an animation going from *Pensiveness* to *Joy*.

2.2 Style of the Interventions

Each intervention could have been provided in different styles, e.g., exclamatory and personal "You're doing good!" or declarative and non-personal "This is good.". We hypothesized that depending on the context, the users' perception of these different styles could vary. Therefore, we led a user-study to determine the style in which the intervention should be provided, depending on the context. This user-study consisted in an online questionnaire simulating a MI-BCI user-training process.

2.2.1 Materials & Methods

We created 3 online questionnaires, each of them simulating an MI-BCI training process in a different context of progress. The results of the participants predefined for each of the questionnaires and their evolution was either negative, neutral or positive depending on the questionnaire. Each questionnaire included 8 situations, with two possible interventions for each situation (which resulted in 16 intervention sentences per questionnaire). Each situation corresponded to an MI-BCI task that the participant was asked to perform (left-hand motor imagery, mental subtraction or mental rotation - as explained in Figure 1), followed by a feedback indicating an alleged success of the task (see Figure 2). This feedback was fixed in each of the questionnaires. It did not correspond to anything that the participant was doing and the participants were informed of that. After the situation was introduced, two different sentences were displayed on screen. Participants were asked to rate each sentence (on a Likert scale ranging from 1 to 5) based on five criteria: appropriate, clear, evaluative, funny, motivating.

Figure 2: Commented example of a part of questionnaire written in french as it was provided to participants. The feedback bar indicates an alleged slightly good success to the participant. Then the participants are presented with a potential sentence that the companion could say in this particular context "Try even harder, we're on the right track" and have to evaluate the sentence based on five criteria: appropriate, clear, evaluative, funny and motivating.

The object of this questionnaire was to determine the impact of the *Context* (negative, neutral or positive progression), of the *Type* (exclamatory or declarative) and of the *Mode* (personal or non-personal) on the five dimensions introduced above. Thus, four kinds of sentences were presented in each context: exclamatory/personal, e.g., "You're doing good!", exclamatory/non-personal, e.g., "This is good!", declarative/personal, e.g., "You're doing good.", declarative/non-personal, e.g., "This is good.". 104 people answered the online questionnaires. Each of them was randomly allocated to one questionnaire, which makes around 34 participants per *Context*. We led five 3-way ANOVAs for repeated measures, one per dimension, to assess the impact of the *Context* (C_3 - independent measures), *Type* (T_2 - repeated measures) and *Mode* (M_2 - repeated measures) on each dimension.

2.2.2 Results

For the 5 dimensions, the ANOVAs showed *Context*Type*Mode* interactions: appropriate [F(2,101)=5.861 ; p≤0.005, η^2=0.104], clear [F(2,101)= 21.596 ; p≤0.001, η^2=0.300], evaluative [F(2,101)=11.461 ; p≤0.001, η^2=0.185], funny [F(2,101)=4.114 ; p≤0.05, η^2=0.075], motivating [D(2,101)= 7.854; p≤0.001, η^2=0.135].
These results (see Table 1) seem to confirm that the *Type* and *Mode* of each intervention should be adapted to the *Context*:
Negative progression - In this context, people definitely prefer declarative and personal sentences that they find more appropriate, clear, funny, motivating and less evaluative.
Neutral progression - Here, people prefer personal sentences, but appreciate as much the declarative and exclamatory sentences for all the dimensions.
Positive progression - In this context, declarative and non-personal sentences are perceived as more clear, appropriate and less evaluative. Exclamatory and personal sentences are perceived as more funny and motivating.

2.2.3 Discussion

Our aim was to design a learning companion whose interventions were adapted to the performance and progress of the user. Based on our results, we chose to provide users facing a negative progression only with declarative personal interventions and those facing a neutral progression with randomly chosen declarative or exclamatory personal interventions. Finally, depending on the intervention goal, we chose to provide participants showing a positive progression with declarative non-personal sentences (when the goal was to give clear information about the task) or exclamatory personal sentences (when the goal was to increase motivation) (see Figure 3).

One should note that when an exclamatory sentence was used for the intervention, the emotion displayed through the facial expressions of PEANUT was made more intense than for an equivalent declarative sentence (see more details about the facial expressions in Section *3 Physical Appearance of PEANUT*.).
These results are rather general and thus may prove useful for other training applications involving a learning companion, or more generally involving support during a training process. For instance, exclamatory sentences can be perceived as more aggressive than declarative sentences, and should therefore be avoided in situations of failure. Also, in case of failure, emotional support is very important. Thus, personal sentences should be favored to make the user feel that the companion is really caring for them. On the contrary, good performers seem to consider that they do not really require this support and thus prefer general, non-personal interventions.

2.3 Performance and Progression Thresholds

For PEANUT to provide interventions based on the user's performances and progression, we had to determine thresholds of performance/progression delimiting intervals within which specific interventions should be provided. We decided to define 2 performance thresholds delimiting 3 intervals: bad, average and good performance. These thresholds were labeled the "low performance threshold" and the "high performance threshold". Similarly, we determined a "negative progression threshold" and a "positive progression threshold", separating negative from neutral, and neutral from positive progression, respectively. We estimated those thresholds and ensured that these estimations could reliably predict performance and progression thresholds in subsequent uses of the BCI by the user. To do so, we re-analyzed the data of 18 participants from a previous study reported in [27]. In this experiment, the participants had to learn to perform the same three mental tasks as in the present study, over the course of 6 sessions, using the same training protocol (without the companion) as in the present paper. A session comprised 5 sequences called runs. A run was divided into 40 trials. Participants were asked to perform a specific mental task during each of these trials. Run 1 of session 1 was used to calibrate the system, i.e., for it to be able to deduce which task the user is performing by analyzing the differences in brain activity patterns when the user performs each of the tasks. We used the classification accuracy, i.e., the percentage of EEG time windows that were correctly classified as the mental task the user was asked to do for this trial, as metric of performance for each trial (see Section *5.1.3 EEG Recordings & Signal Processing* for details). In order to estimate the different thresh-

		Positive Progress		Negative Progress		Neutral Progress	
		Exclamatory	*Declarative*	*Exclamatory*	*Declarative*	*Exclamatory*	*Declarative*
		Avg ± Std	Avg ± Std	Avg ± Std	Avg ± Std	Avg ± Std	Avg ± Std
Appropriate	*Personal*	3,68 ± 0,13	3,73 ± 0,12	3,73 ± 0,12	4,02 ± 0,12	3,54 ± 0,13	3,64 ± 0,12
	Not personal	3,68 ± 0,13	3,92 ± 0,14	3,8 ± 0,12	3,59 ± 0,13	3,68 ± 0,13	3,09 ± 0,14
Clear	*Personal*	4,33 ± 0,13	4,06 ± 0,13	4,12 ± 0,12	4,51 ± 0,12	4,08 ± 0,12	4,07 ± 0,12
	Not personal	4,09 ± 0,14	4,4 ± 0,16	4,28 ± 0,13	4,11 ± 0,15	4,14 ± 0,13	3,61 ± 0,16
Evaluative	*Personal*	3,12 ± 0,14	3,2 ± 0,14	3,23 ± 0,14	2,86 ± 0,13	2,75 ± 0,14	2,78 ± 0,13
	Not personal	2,88 ± 0,15	2,84 ± 0,16	2,61 ± 0,14	3,1 ± 0,15	2,53 ± 0,15	2,64 ± 0,16
Funny	*Personal*	2,57 ± 0,15	2,13 ± 0,15	2,07 ± 0,14	2,14 ± 0,14	2,46 ± 0,15	2,29 ± 0,15
	Not personal	2,33 ± 016	2,14 ± 0,15	2,05 ± 0,15	1,9 ± 0,14	2,31 ± 0,15	1,82 ± 0,14
Motivating	*Personal*	3,87 ± 0,13	3,76 ± 0,12	3,61 ± 0,12	3,66 ± 0,12	3,51 ± 0,13	3,62 ± 0,12
	Not personal	3,53 ± 0,13	3,59 ± 0,14	3,36 ± 0,13	3,21 ± 0,13	3,42 ± 0,13	2,74 ± 0,14

Table 1: Mean rate given to the sentences depending on their Mode (Personal, Not personal), Type (Exclamatory, Declarative) and on the Progress (Positive, Negative, Neutral). We highlighted in yellow all the values that deviated by more than one standard deviation from the group mean score. For the Appropriate, Clear, Funny and Motivating dimensions, the highlighted values were higher than the average score plus one standard deviation while for the Evaluative dimension, the highlighted values were lower than the average score minus one standard deviation.

olds, the data was analyzed offline with Matlab.

2.3.1 Estimating the performance thresholds

We constructed the distribution of performance values over trials and defined the bad and good performance thresholds as the 25th and the 75th percentiles of that distribution, respectively. Thus, the bottom 25% of each participant's performances were considered bad performances, the top 25% good performances, and the remaining performances in-between were considered neutral. The question was to assess the feasibility of predicting future performance (and thus thresholds) based on the data collected at the beginning of the training (first run of the first session). Indeed, the sooner we are able to determine the performance thresholds, the sooner we can provide users with interventions adapted to their performance, thus maximizing the relevance of these interventions.

First, we checked whether we could estimate those thresholds on the first run with BCI use, i.e., on run 2 of session 1 (run 1 being the calibration run). We thus estimated the performance thresholds independently on run 2, and on runs 3, 4 and 5 of session 1 together. We then computed their correlations over participants, to find whether thresholds estimated on run 2 could be used to predict thresholds estimated on run 3, 4, 5. We obtained significant correlations of $r = 0.6422$ ($p < 0.01$) for bad performance thresholds, and of $r = 0.5482$ ($p < 0.05$) for good performance thresholds. Thus, in order to select the appropriate behavior for PEANUT, we used the thresholds estimated on run 2 to compute the thresholds for runs 3, 4 and 5 of session 1 using the same ratio as the ones found in these control data. However, thresholds estimated on the data from a single run

are bound to be less reliable than thresholds based on several runs. We thus studied whether thresholds estimated on runs 2 to 5 of the first session, could be used to predict the thresholds of the runs of subsequent sessions. They appear to be correlated with $r = 0.6628$ ($p < 0.01$) and 0.4438 ($p = 0.07$ - not significant but a trend) for bad and good performance thresholds respectively. Thus, to determine the behavior of PEANUT for subsequent sessions, we computed the thresholds using the runs 2 to 5 of session 1 still using the same ratio as the ones found in these control data.

2.3.2 Estimating the progression thresholds

To estimate progression thresholds, we used the performances from N successive trials, and computed the slope of a linear regression relating time (here trial indexes) with performance. A positive/negative slope indicated a positive/negative progression, respectively. We then constructed the distribution of these regression slopes over trials, and determined the negative progression threshold as the 25th percentile of this distribution, and the positive progression threshold as the 75th percentile of this distribution. Similarly as for the performance thresholds, we studied whether we could predict the future progression thresholds from their estimation on the first runs. Nonetheless, progression estimation requires more trials than performance estimation (N versus 1). As such there are fewer progression measures in a single run, which in practice made it impossible to reliably predict the progression thresholds of runs 3, 4 and 5 by using run 2 alone for threshold-estimation. However, it appeared to be possible to predict progression thresholds for all the runs of sessions 2 to 6, from the threshold-estimated based on runs 2 to 5 of session 1. In particular, the posi-

tive progression threshold of the runs of the session 1 appeared to be significantly correlated with both the positive ($r = 0.4843$, $p < 0.05$) and negative ($r = -0.5476$, $p < 0.05$) progression thresholds from the runs of the subsequent sessions. Note that these correlations were obtained for $N = 6$. Indeed, we studied N between 2 and 10, and selected the best N as the one maximizing the correlations, to obtain the most reliable thresholds. Therefore, the progression thresholds from sessions 2 to 6 were estimated by computing the positive progression threshold from runs 2 to 5 of session 1 using the same ratio as the ones found in these control data. The companion thus provided progression related interventions only from session 2 onward.

These analyses also guided the choice of the frequency of the interventions of PEANUT. Since progression was measured over N=6 trials, we informally tested different intervention frequencies of about one every 6 trials. These informal tests with pilot testers revealed that interventions every 6 ± 2 trials seemed appropriate, as they were neither annoying nor too rare. PEANUT thus intervened at that frequency, the exact trial of intervention being randomly selected in the 6 ± 2 trials following the previous intervention.

2.4 Rule tree

Once all the parameters governing the behavior of PEANUT had been determined, we were able to build the rule tree that enables the system to select one specific intervention (i.e., sentence & sentence style & facial expression) with respect to the context. Figure 3 is a schematic representation of this rule tree: based on a specific performance and progression, it executes a set of rules to select the appropriate intervention. For example, if the user had a good performance and a neutral progress then the rule tree would select an appropriate sentence which would either advice him to try a new strategy in a declarative sentence if it had been some time that the progress did not change, e.g., "Maybe you could try a new strategy.", or an exclamatory or declarative sentence of encouragement, e.g., "You're doing good!".

3 Physical Appearance of PEANUT

Designing the appearance of PEANUT consisted in two steps: designing its body, and designing its face and facial expressions. The decisions concerning the face have been made based on a user-study. Those concerning the body were based on a review of the literature.

3.1 Body of PEANUT

To increase social presence we decided to make a physical companion instead of a virtual one [20, 58] and used anthropomorphic features to facilitate social interactions [10].The combination of physical characteristics, personality/abilities, functionalities and learning function had to be consistent [48]. We were inspired by TEEGI [14] and TOBE [17], two avatars providing users with tools to explore their inner state (EEG and physiological data, among others). Since their functions are simple and they are unable to interact with the user, their designers chose to propose cartoon-like characters with anthropomorphic child-like body shapes, which can induce positive emotions through design [62]. The functionalities of our companion being basic as well, we also decided to design a cartoon and child-like companion rather than a realistic one. We used the voice of a child to record the interventions of PEANUT, which also enabled us not to associate PEANUT with a gender. We also took into account our own constraints deriving from the size of the smartphone we used to display the face of PEANUT and the learning environment. Finally, concerning the size of the companion, since PEANUT was on the desk right next to the computer screen on which the feedback was displayed, its proportions had to be suitable: not too small so that the body was proportional to its face, and not too large so that it could always be within a user's field of view without concealing the screen. This process resulted in a 30 cm high companion, see Figure 1.

3.2 Facial Expressions of PEANUT

Based on the results of PEANUT behavior design, we wanted the companion to be able to express eight emotions: Trust, Joy, Surprise, Admiration, Boredom, Sadness, Anger and a Neutral expressions. We asked a designer to create three styles of faces (see Figure 4) [2]. We wanted the faces to be cartoon-like, so that they fitted the body and complied with the recommendations from the literature [10, 48, 62]. The object of the user-study introduced hereafter was to find the best style (among three) for PEANUT with respect to 5 dimensions: expressiveness, sympathy, appeal, childlike, consistent (with the expression it was supposed to convey).

---

[2]To learn more about Marie Ecarlat's work - http://marieecarlat.tumblr.com/

PERFORMANCE = Bad

PROGRESS = -   PROGRESS = =   PROGRESS = +

PERFORMANCE = Average

PROGRESS = -   PROGRESS = =   PROGRESS = +

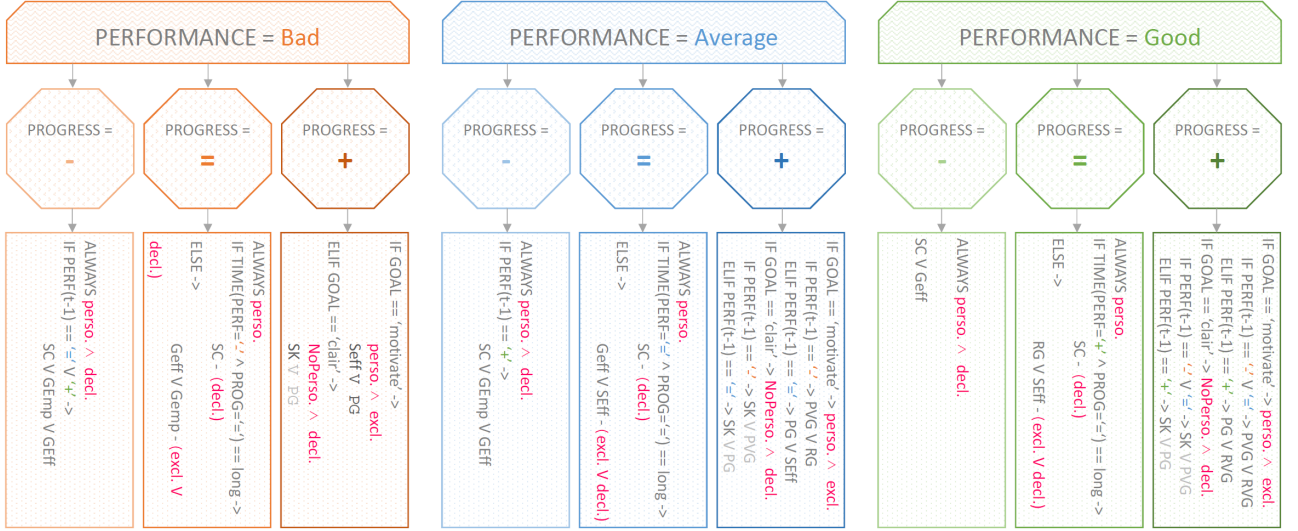PERFORMANCE = Good

PROGRESS = -   PROGRESS = =   PROGRESS = +

Figure 3: The rule tree corresponds to a set of rules that selects the interventions of PEANUT (i.e., type and mode of sentence) depending on users' performance and progression ("-"=negative, "="=neutral, "+"=positive). Type of sentences: "perso." for personal, "NoPerso." for non-personal ; Mode of the sentence: "decl." for declarative, "excl." for exclamatory. Interventions: "GEff" for general effort, "SEff" for support effort, "GEmp" for general empathy, "SK" for strategy keep, "SC" for strategy change, "RG" for results good, "RVG" for results very good, "PG" for progress good, "PVG" for progress very good. Moreover, the "∧" sign represents the logical operator "and" and the "∨" sign represents the logical operator "or".

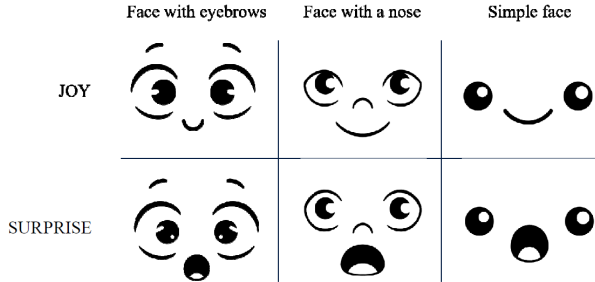Face with eyebrows   Face with a nose   Simple face

JOY

SURPRISE

Figure 4: Three face styles, with the example of 2 emotions: Joy and Surprise. Participants of the dedicated user-study selected the face with eyebrows for PEANUT. See annex file for the rest of the facial expressions.

3.2.1 Materials & Methods

We created an online questionnaire which was divided into different items, with each item corresponding to one emotion. These items were presented in a random order. For each item, the three face styles were presented (in a counterbalanced order), side by side. Participants were asked to chose which of the three styles corresponded the most to each of the following dimensions: expressive, sympathetic, appealing, childlike and consistent. They were also asked to rate each style on a 5-point Likert scale, 1 corresponding to "I don't like it at all" and 5 to "I like it a lot". Ninety-seven participants answered the online questionnaire. We first led a 1-way ANOVA to determine if the rates associated with each style were different. Then, we led a 3-way ANOVA for repeated measures, to assess the impact of the face style (F_3 - repeated measures), the type of emotion (E_8 - repeated measures) and the dimension (D_5 - repeated measures) on the allocated score.

3.2.2 Results

On a 5-point Likert scale, the face with eyebrows was rated 3.58 ± 1.26, the face with a nose 2.96 ± 1.37 and the simple face 3.86 ± 1.10. The 1-way ANOVA for repeated measures revealed a main effect of the style [$F(1,93)=8.442$; $p \leq 0.005$, $\eta^2=0.083$]. The simple face and the face with eyebrows were significantly better rated than the face with a nose. However, there was no difference of rating between the simple face and that with eyebrows. Thus, we then performed a 3-way ANOVA for repeated measures to evaluate the effect of the face, of the emotion and of the dimension on the rating. Results suggested a main effect of the style of face [$F(1,93)=17.543$; $p \leq 0.001$, $\eta^2=0.159$], of the emotion [$F(1,93)=11.307$; $p \leq 0.001$, $\eta^2=0.108$] and of the dimension [$F(1,93)=12.184$; $p \leq 0.001$, $\eta^2=0.116$]. Moreover, face*dimension [$F(1,93)=58.531$; $p \leq 0.001$, $\eta^2=0.386$], face*emotion [$F(1,93)$ $=11.307$; $p \leq 0.001$, $\eta^2=0.108$] and dimension*emotion [$F(1,93)=17.543$; $p \leq 0.001$, $\eta^2=0.159$] interaction effects were revealed. The face with the

10

eyebrows was significantly preferred to the others, which was strengthened by participants' comments indicating that eyebrows increased expressiveness. However, this face was not preferred for the Ecstatic (i.e., high intensity of Joy) and Admiration items. An analysis of the comments helped us improve those expressions. Several people felt like the shape of the eyes gave the impression the companion was about to cry and that it was squinting.

3.2.3 Discussion

Based on our results, we selected the face with eyebrows (see Figure 4) for PEANUT. We asked the designer to improve the expressions of Ecstatic (i.e., high intensity of Joy) and Admiration with respect to participants' comments. In a second instance, the designer animated each of the expressions. The animations enabled a transfer from a neutral expression to a high intensity of each of the selected emotions. For example, the *Joy* emotion had three possible levels of intensity, i.e., serenity, joy and ecstatic. Once, the behavior and appearance of PEANUT developed, they had to be implemented in one whole system related to the BCI protocol which will be presented in the following section.

4 System Architecture

Implementing the whole BCI system as well as PEANUT required to design, assemble and connect multiple pieces of hardware and software. Users' EEG signals were first measured using EEG hardware (g.tec gUSBAmp, g.tec, Austria) and then collected and processed online using the software OpenViBE [56]. OpenViBE provided users with a visual feedback about the estimated mental task, and computed users' performances which were then transmitted to a home-made software, the "Rule Engine" using the Lab Streaming Layer (LSL) protocol [35]. The rule engine processed performance measures received from OpenViBE to compute progression measures and browsed the Rule Tree described in Figure 3 in order to select an appropriate intervention (sentence and facial expression) for PEANUT with respect to the context. The selected intervention was then transmitted to an Android smartphone application, using WebSocket, which enunciated the sentence and animated the facial expression of PEANUT. This whole architecture is summarized in Figure 5 and described in more details in the following sections.

4.1 Rule Engine

The Rule Engine software receives from OpenViBE the markers indicating the start and end of trials, as well as performance measures at the end of each trial. It first computes a progression measure (see section *2.3.2 Estimating the progression thresholds*) and then browses the rule tree in order to select the intervention type to be triggered. Each intervention type contained between 1 and 17 sentences. One of them was selected randomly, taking care not to take a sentence that had already been chosen in the same run (thanks to a small cache of already triggered sentences kept for each intervention type) in order to avoid repetition. Finally, the Rule Engine sent intervention identifiers to the smartphone application.

4.2 Smartphone - Sentence Enunciation, Facial Expression Animation

To display the facial animations and enunciate the sentences, we used a smartphone. We designed an application that displays the face of the companion, plays animations and sounds when required. By default a neutral facial expression is shown, with eye-blinks occurring from time to time. When intervention identifiers were received from the Rule Engine, the application animated the facial expressions and enunciated the sentences. We used Praat software [4] offline in order to realize phonetic alignment with the companion's mouth movements for each sentence.

5 Evaluation of the efficiency to improve BCI user-training of PEANUT

Once the companion's behavior and appearance had been designed and implemented, the next step consisted in testing its efficiency to improve MI-BCI user-training both in terms of MI-BCI performance and user experience. Below we present the study performed to test the efficiency of PEANUT.

5.1 Materials & Methods

5.1.1 Participants

Twenty-eight MI-BCI-naive participants (14 women ; aged 21.21±1.6 year-old) took part in this study, which was conducted in accordance with the relevant guidelines for ethical research according to the Declaration of Helsinki. This study was also approved by the legal and ethical authorities of Inria Bordeaux Sud-Ouest (the COERLE, approval number: 2016-02) as it satisfied the ethical rules and principles of the institute. All the participants signed an informed consent form at the beginning of the experiment and received a compensation of 50 euros. The *experimental group* (N=10 ; 5 women ; aged 20.7±2.11 year-old), received emotional and social support adapted to their MI-BCI performance & progression throughout the MI-BCI training sessions. For the *control*
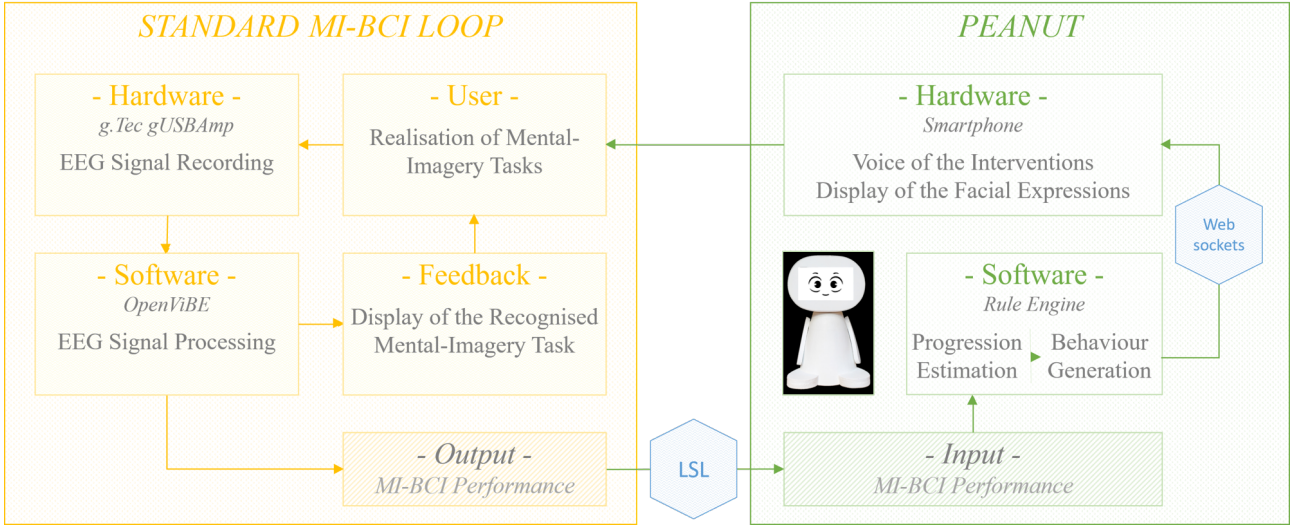
Figure 5: Software and hardware architecture of PEANUT.

group (N=18 ; 9 women ; aged 21.5±1.2), data from the 3 first sessions (out of 6) from a previous experiment [27] were used. Participants from this study followed the same training protocol without the learning companion. The same data was used to define the equations to compute the thresholds (see Section *2.3. Performance and Progression Thresholds*).

5.1.2 Experimental Protocol

Before the first session, participants were asked to complete a validated psychometric questionnaire, the 16PF5 [7], that enabled us to compute their "autonomy" and "tension" scores. Each participant took part in 3 sessions, on 3 different days. Each session lasted around 2 hours and was organized as follows: completion of questionnaires, installation of the EEG cap, five runs during which participants had to learn to perform three MI-tasks (around 60 min, including breaks between the runs), uninstallation of the EEG cap, completion of questionnaires, and debriefing. The MI-tasks, i.e., left-hand motor imagery, mental rotation and mental subtraction, were chosen according to Friedrich et al. [15]. "Left-hand motor imagery" (*L-HAND*) refers to the kinaesthetic continuous imagination of a left-hand movement, chosen by the participant, without any actual movement [15]. "Mental rotation" (*ROTATION*) and "mental subtraction" (*SUBTRACTION*) correspond respectively to the mental visualization of a 3 Dimensional shape rotating in a 3 Dimensional space [15] and to successive subtractions of a 2-digit number (ranging between 11 and 19) from a 3-digit number, both being randomly generated and displayed on a screen [15].

During each run, participants had to perform 45 trials (15 trials per task, presented in a random order), each trial lasting 8s (see Figure 6). At t=0s, an arrow was displayed with a left hand pictogram on the left (*L-HAND* task), the subtraction to be performed at the top (*SUBTRACTION* task) and a 3D shape on the right (*ROTATION* task). At t=2s, a "beep" announced the coming instruction and one second later, at t=3s, a red arrow was displayed for 1.250s. The direction of the arrow informed the participant which task to perform, e.g., an arrow pointing to the left meant the user had to perform a *L-HAND* task. In order to stress this information, the pictogram representing the task to be performed was also framed with a white square until the end of the trial. Finally, at t=4.250s, a visual feedback was provided in the shape of a blue bar, the length of which varied according to the classifier output. Only positive feedback was displayed, i.e., the feedback was provided only when there was a match between the instruction and the recognized task. Participants were instructed to find strategies that would maximize the length of the blue bar. The feedback lasted 4s and was updated at 16Hz, using a 1s sliding window. During the first run of the first session (i.e., the calibration run, see next Section), as the classifier was not yet trained to recognize the mental tasks being performed by the user, it could not provide a consistent feedback. In order to limit biases with the other runs, e.g., EEG changes due to different visual processing between runs, the user was provided with an equivalent sham feedback, i.e., a blue bar randomly appearing and varying in length, and not updated according to the classifier output, as in [15]. A gap lasting between 3.500s and 4.500s separated each trial.
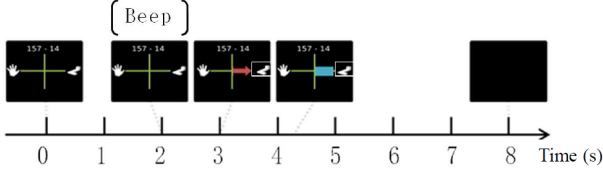
Figure 6: Timing of a trial.

The participants from the *experimental group* were accompanied by PEANUT during their training, from the second run of session 1 (after the calibration run). The interventions of PEANUT were adapted to each participants' performance during the first session, and to each of their performance and progression during the second and third sessions. Finally, after the last session we asked participants from both groups to assess the usability of the MI-BCI system using a questionnaire focusing on the 4 following dimensions: Learnability/Memorability (LM), efficiency/effectiveness (EE), safety (Saf.) and satisfaction (Sat.). Each dimension was associated with different sentences which the participants had to give their opinion about on a Likert scale ranging from 1 (i.e., do not agree at all) to 5 (i.e., totally agree). For example, the satisfaction was in part evaluated though the sentence "Overall, I am satisfied with the system". Participants trained with PEANUT also had a questionnaire assessing the adequacy of the latter regarding its appearance, the content and the frequency of its interventions and its general appreciation. Once again, each evaluated dimension was associated with different sentences which the participants had to give their opinion about on a Likert scale ranging from 1 (i.e., do not agree at all) to 7 (i.e., totally agree). For example, the content of the intervention was in part evaluated though the sentence "I think that the interventions of the companion were relevant".

5.1.3 EEG Recordings & Signal Processing

The EEG signals were recorded from a g.USBamp amplifier, using 30 scalp electrodes (F3, Fz, F4, FT7,FC5, FC3, FCz, FC4, FC6, FT8, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz, CP4, P5, P3, P1, Pz, P2, P4, P6, PO7, PO8, 10-20 system) [15], referenced to the left ear and grounded to AFz. EEG data were sampled at 256 Hz. In order to classify the 3 mental imagery tasks on which our BCI is based, the following EEG signal processing pipeline was used. First, EEG signals were band-pass filtered in 8-30Hz, using a Butterworth filter of order 4. Then EEG signals were spatially filtered using 3 sets of Common Spatial Pattern (CSP) filters [54]. The CSP algorithm aims at finding spatial filters whose resulting EEG band power is maximally different between two classes. Each set of CSP filters was optimised on each

user's calibration run (i.e., the first run of the first session) to discriminate EEG signals for a given class from those for the other two classes. We optimized 2 pairs of spatial filters for each class, corresponding to the 2 largest and lowest eigen values of the CSP optimization problem for that class, thus leading to 12 CSP filters. The band power of the spatially filtered EEG signals was then computed by squaring the signals, averaging them over the last 1 second time window (with 15/16s overlap between consecutive time windows) and log-transformed. These resulted in 12 band-power features that were fed to a multi-class shrinkage Linear Discriminant Analysis (sLDA) [39], built by combining three sLDA in a one-versus-the-rest scheme. As for the CSP filters, the sLDA were optimized on the EEG signals collected during the calibration run, i.e., during the first run of the first session. The resulting classifier was then used online to distinguish between the 3 MI-tasks during the 3 sessions. The sLDA classifier output (i.e., the distance of the feature vector from the LDA separating hyperplane) for the mental imagery task to be performed was used as feedback provided to the user. In particular, if the required mental task was performed correctly (i.e., correctly classified), a blue bar with a length proportional to the LDA output and extending towards the required task picture was displayed on screen and updated at 16Hz. This processing pipeline led to a total of 64 classification outputs per trial (16 per second for 4 seconds). OpenViBE thus computed the user's performance for this trial as the rate of correct classification outputs among these 64 outputs, and sent it to the rule engine, which in turn computed progression measures.

5.1.4 Variables & Factors

We used both the mean and the peak classification accuracy as a measure of performance. These measures are traditionally used by the community. The mean accuracy represents the percentage of time windows from the feedback periods that were correctly classified. The peak classification was computed by averaging the performances obtained during the time window of the feedback period for which the classification accuracy over all trials is maximal (see Section *5.1.3 EEG Recordings & Signal Processing* for more details on the classifier). We studied the impact of the group (no companion, PEANUT) on participants' MI-BCI performance, with respect to the session and participant's profile ("autonomy" and "tension" scores according to the 16PF5 questionnaire [7]). We also evaluated the impact of the group on MI-BCI usability and on the perception of the companion, with respect to MI-BCI performance.

5.2 Results

We checked the normality of the variables that we obtained using Lilliefors corrected Kolmogorov-Smirnov tests. If the variables were Gaussian, we performed t-tests to compare the two groups. In the opposite case we performed Mann-Whitney U tests. Mean and peak performances from each session had a normal distribution ($p \geq 0.1$). We also verified that there was no confounding factor between our two groups. Participants from the two groups were statistically similar before the training. There were no significant differences of age [Mann-Whitney U test, U=50, p=0.06], initial performances computed using a 5-fold LDA classification on CSP characteristics from the first run of the first session where PEANUT was not present for either group [t-test, t(26)=0.85 ; p=0.4], tension [Mann-Whitney U test, U=75.5, p=0.49] or autonomy [Mann-Whitney U test, U=60.5, p=0.16].

5.2.1 Assessment of the influence of PEANUT on MI-BCI Performances

Then, we compared the group's MI-BCI performance in terms of mean and peak classification accuracy. We performed a 2-way repeated measures mixed ANOVA with *"Group*Session"* as independent variables and the repeated measures of mean or peak performance over the session as dependent variable. When using the mean performances as dependent variable, results revealed no significant effect of the *"Group"* [F(1,26)=0.63; p=0.43, η^2=0.02], *"Session"* [F(2,52)=0.03; p=0.97, η^2=0], nor *"Session*Group"* [F(2,52)=0.79; p=0.46, η^2=0.03], i.e., the evolution of the performances over the sessions. Similar results were obtained with the peak performances. They revealed no significant effect of the *"Group"* [F(1,26)=0.87; p=0.36, η^2=0.03], *"Session"* [F(2,52)=0; p=1, η^2=0], nor *"Session*Group"* [F(2,52)=0.46; p=0.64, η^2=0.02], i.e., the evolution of the performances over the sessions. Averaged over all runs and sessions, the group with no companion (N=18) and the group with PEANUT (N=10) respectively obtained peak performances scores of 65.73% $\pm$ 6.21 and 63.14% $\pm$ 8.4 and mean performances scores of 52.76% $\pm$ 5.62 and 50.74% $\pm$ 7.77 (see Figure 7).

Nevertheless, we performed analyses to assess the impact of users' profile on performance, depending on the group. The influence of the "autonomy" of participants training without PEANUT on their MI-BCI performances previously found in [27] when taking into account the 6 sessions of the participants' training could still be found when taking into account only the first 3 sessions to compare the results of both groups. We observed a positive correlation of the mean and peak performances with the autonomy of the participants who had a clas-

sical training without PEANUT [Spearman correlation ; mean: r=0.54, p=0.02 ; peak: r=0.5, p=0.03] which means that participants who like to work in group tend to be disadvantaged. Interestingly, an opposite significant negative correlation between the measure of autonomy and the mean and peak performances over the sessions for the participants trained with PEANUT [Spearman correlation, mean: r=-0.78, p=0.01, peak: r=-0.75, p=0.01] which means that participants who are prone to work in a group tend to perform better than those who rather work alone when PEANUT is part of the training. To further investigate the influence of PEANUT and the autonomy of the participants on their BCI performances, we separated the participants into two groups depending on their autonomy. The threshold between high and low autonomy was defined using the median autonomy score (i.e., score of 5, 10 being the maximum). We then led 2-way ANOVAs to determine the influence of *Group* (PEANUT or no PEANUT) and the *Autonomy* (Autonomous or non Autonomous) on MI-BCI performances. Results indicate a *Group*Autonomy* interaction for both mean performances [F(1,24)=6.35 ; p=0.02, η^2=0.21] and peak performances [F(1,24)=7.23 ; p=0.01, η^2=0.23] (see Figure 8). Overall these results confirm the importance of this personality trait for BCI training as was suggested in [27]. They also indicate a possible differential influence of a learning companion on MI-BCI performances depending on the personality trait.

However, the previous influence of tension on MI-BCI performances found on the participants trained without PEANUT in [27] when taking into account the 6 sessions of the participants' training could not be found when taking into account only the first 3 sessions [Spearman correlation ; mean: r=-0.25, p=0.33 ; peak: r=-0.21, p=0.4]. It could neither be found on the results of the participants trained with PEANUT [Spearman correlation ; mean: r=-0.14, p=0.7 ; peak: r=-0.13, p=0.72]. This aspect of psychological profile influence on MI-BCI performances might require further investigations with longer term experiments.

We also observed a strong negative correlation between the performances and the measure of sensibility (based on the dimension of the 16PF5 psychometric questionnaire) of the participants trained with PEANUT [Spearman correlation ; mean: r=-0.89, p=10^{-3} ; peak: r=-0.91, p$\leq 10^{-3}$]. The more sensitive people were, the less likely to have good MI-BCI performances they were. This correlation is not found for the participants trained without PEANUT [Spearman correlation ; mean: r=-0.04, p=0.88 ; peak: r=-0.06, p=0.82].
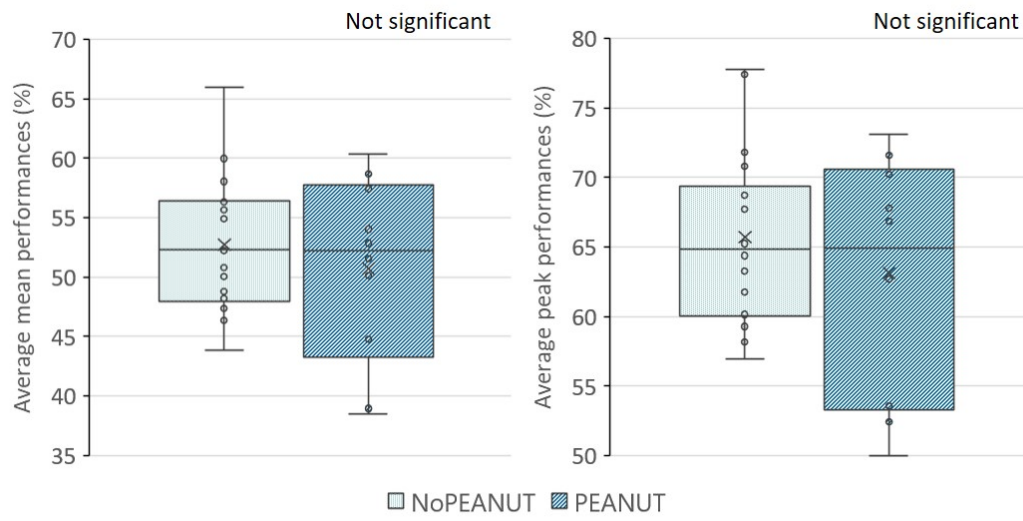
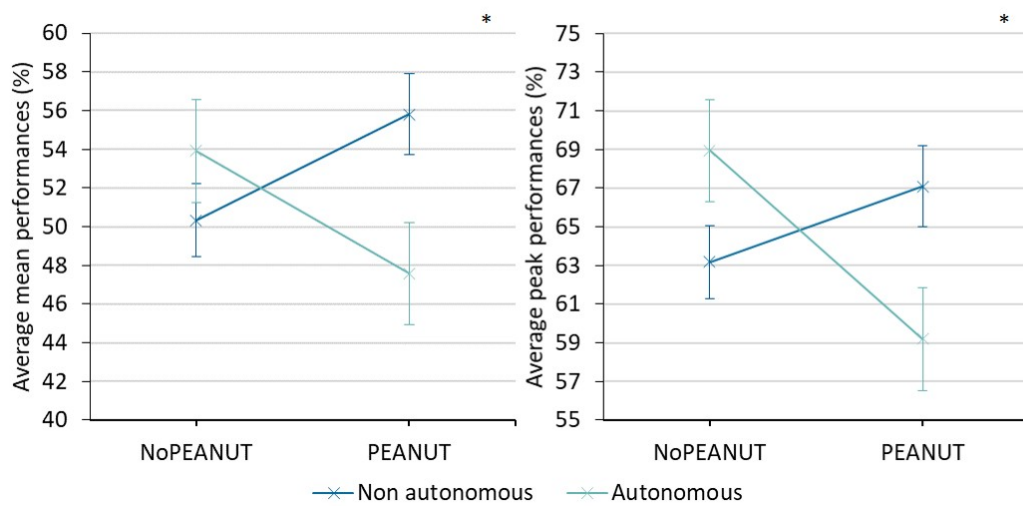Figure 7: Average mean and peak performances for both the experimental and the control group.

Figure 8: Average mean and peak performances of the participants depending on there *Autonomy* and the *Group* they belonged to.

5.2.2 Assessment of the influence of PEANUT on the user experience

Then, we analysed the influence of *Autonomy* and *Group* on usability scores, which were divided into 4 dimensions: learnability/memorability (LM), efficiency/effectiveness (EE), safety (Saf), satisfaction (Sat) [19]. We performed four 2-way ANCOVAs (one per dimension) with the *Autonomy* and *Group* as factor, the usability score for the target dimension as dependent variable and the mean or peak classification accuracy as co-variable for the LM, EE and Saf dimensions to remove the influence of performances on their evaluation (Spearman correlation; mean: LM [r=0.58, p=10^{-3}], EE [r=0.54, p$\leq$$10^{-2}$], Saf [r=0.59, p=$10^{-3}$], Sat [r=0.07, p=0.73] ; peak: LM [r=0.56, p$\leq$$10^{-2}$], EE [r=0.56, p$\leq$$10^{-2}$], Saf [r=0.548, p$\leq$$10^{-2}$], Sat [r=0.03, p=0.89]) (see Figures 9 and 10).

Results reveal a close to significant effect of the group on the LM dimension [mean: D(1,28)=3.68, p=0.07, η^2=0.14 ; peak: D(1,28)=3.99, p=0.06, η^2=0.15]. In average, participants who were provided with PEANUT consider the system's learnability/memorability to be higher by 7.4% than those without PEANUT. A *Group*Autonomy* interaction [mean: D(1,28)=3.2, p=0.09, η^2=0.12 ; peak: D(1,28)=4.05, p=0.06, η^2=0.15] on the EE dimension also tends to be significant when using the peak performance as covariate. Autonomous participants reported feeling that they were more Efficient/Effective by 13.4% when PEANUT was present. To the contrary, non autonomous participants reported feeling that they were less Efficient/Effective by 1.8%.

5.2.3 Assessment of the characteristics of PEANUT

Finally, we analyzed the results of the open questionnaire that participants in the experimental group answered about the characteristics of PEANUT, i.e., appearance, content and frequency of intervention, general appreciation. We summed the responses to the Likert scales for each characteristic and divided them in relation to the maximum score that could have been given to these questions to obtain the following percentages. The higher the percentage is and the better the participants rated the characteristic of PEANUT. On average, the users rated the different characteristics as follow: appearance [M=82,14%, SD=13.07%], content [M=56.9%, SD=16.92%] and frequency of intervention [M=80.36%, SD=13.1%], general appreciation [M=67.14%, SD=19.22%] (see Figure 11).

The appearance of PEANUT and the frequency of its intervention seem to have been appreciated. Though, improvements should probably be made regarding the content of its interventions and the general ap-

preciation of PEANUT. The comments from the participants provide further information. Two participants reported not understanding its role and expected a more informative feedback. This is in line with recommendations from the literature but providing an informative feedback still remains a challenge (see Section *1 Related work*)). Two also reported that the sentences did not always seem in agreement with the visual feedback they received. This could be because the rule tree took into account the last performance of the user when choosing a sentence regarding the progression of the user but could still lead to PEANUT congratulating participants when their last performance was not promising. For example, PEANUT could still tell participants that they were improving when their last performance was considered as poor. Lastly, a positive correlation was found between the "tension" of the participants and the responses they gave regarding the content of intervention [r=0.671, p=0.034] and the general appreciation [r=0.703, p=0.023] of PEANUT. This indicates that the more tensed participants tended be and the more they appreciated PEANUT in general and its content of intervention.

5.3 Discussion

First of all, we found that non-autonomous users, who had lower MI-BCI performances than the others when using a classical feedback, seem to have better performances by 3.9% than the others when using PEANUT. Second, using PEANUT seems to have improved the usability of the MI-BCI. Participants who trained with PEANUT gave in average 7.4% higher learning/memorability scores than the members of the control group. Furthermore, autonomous participants trained with PEANUT found that they were more efficient than the ones trained without PEANUT by 13.4% in average. However, PEANUT had a negative impact on the performances of sensitive and autonomous participants. This could be related to the margin of improvement reported by the participants regarding the content of the interventions of PEANUT who expected a more informative feedback. Even though PEANUT was providing feedback in-between trials, some participants may also have been distracted by it and not have benefited from the feedback as much as the others [32]. Finally, the influence of a learning companion depends on the task and the user's personality [61]. Therefore, the impact of PEANUT could be limited by the fact that it does not adapt to the user's personality and because it does not reduce the complexity of the task.

Through the feedback provided by PEANUT, participants in the experimental group were informed of the evolution of their performances and advised to keep or change strategies. These meta-information
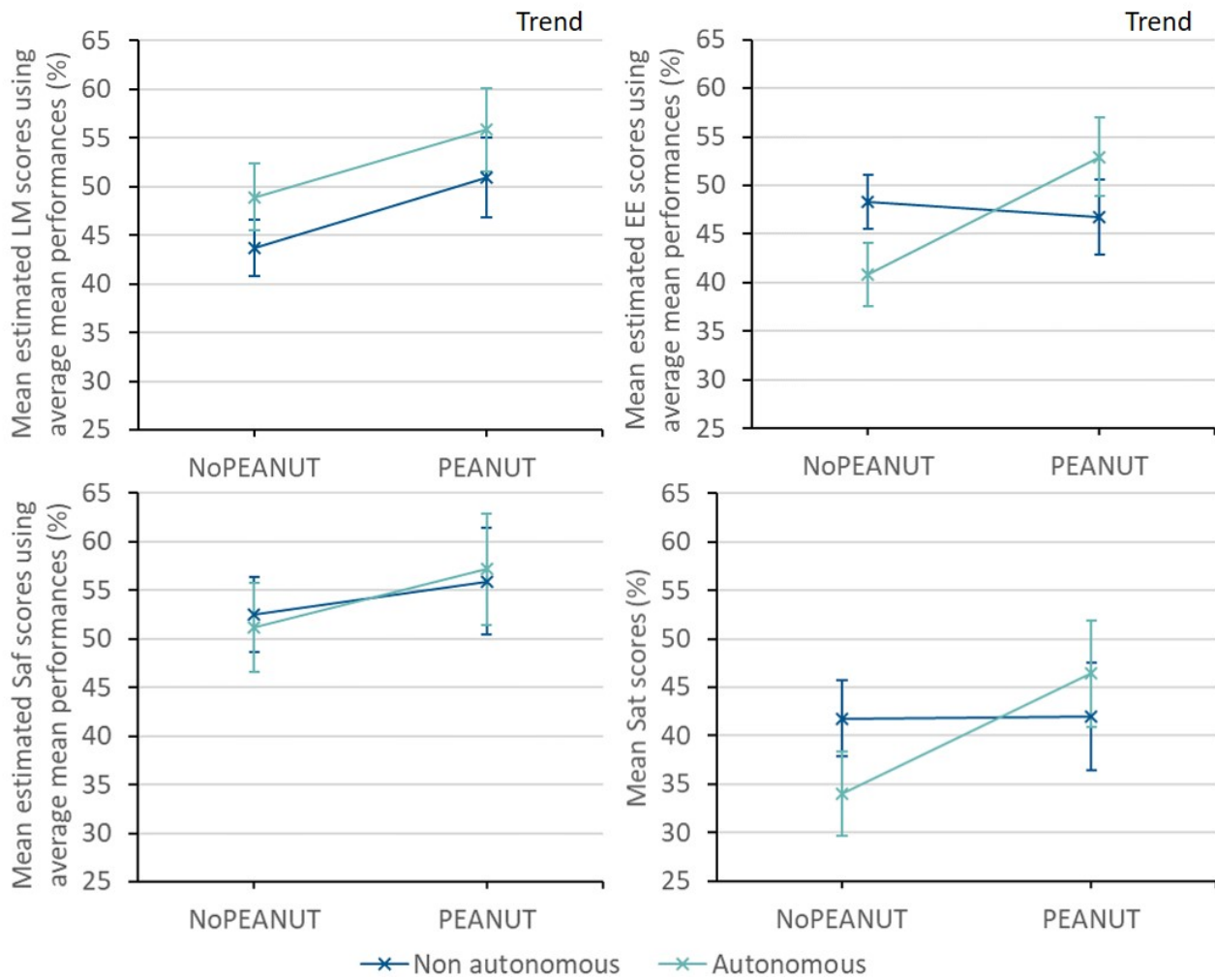
Figure 9: Usability scores, with respect to users' group and autonomy, corrected using the average mean performance if needed.
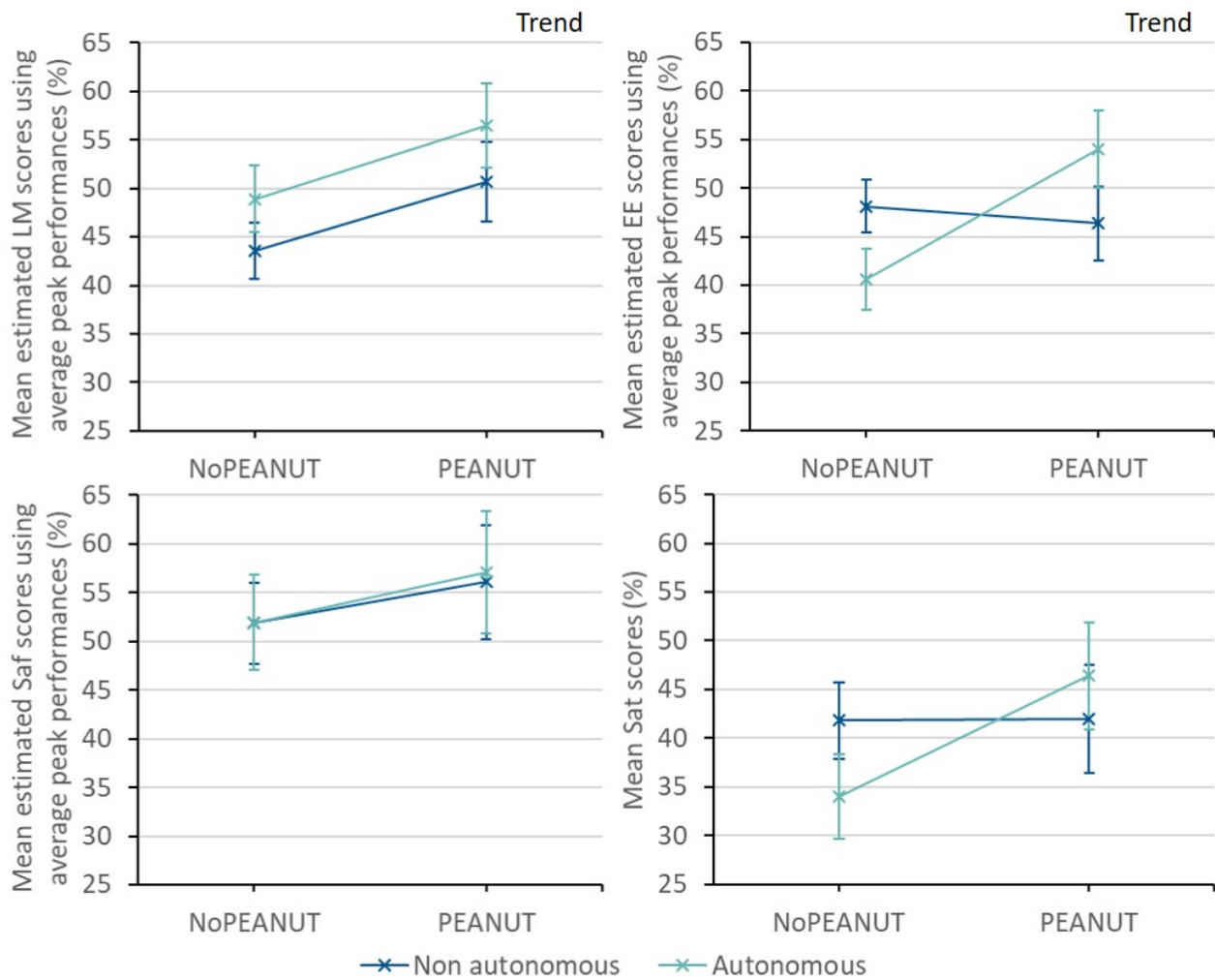
Figure 10: Usability scores, with respect to users' group and autonomy, corrected using the average peak performance if needed.
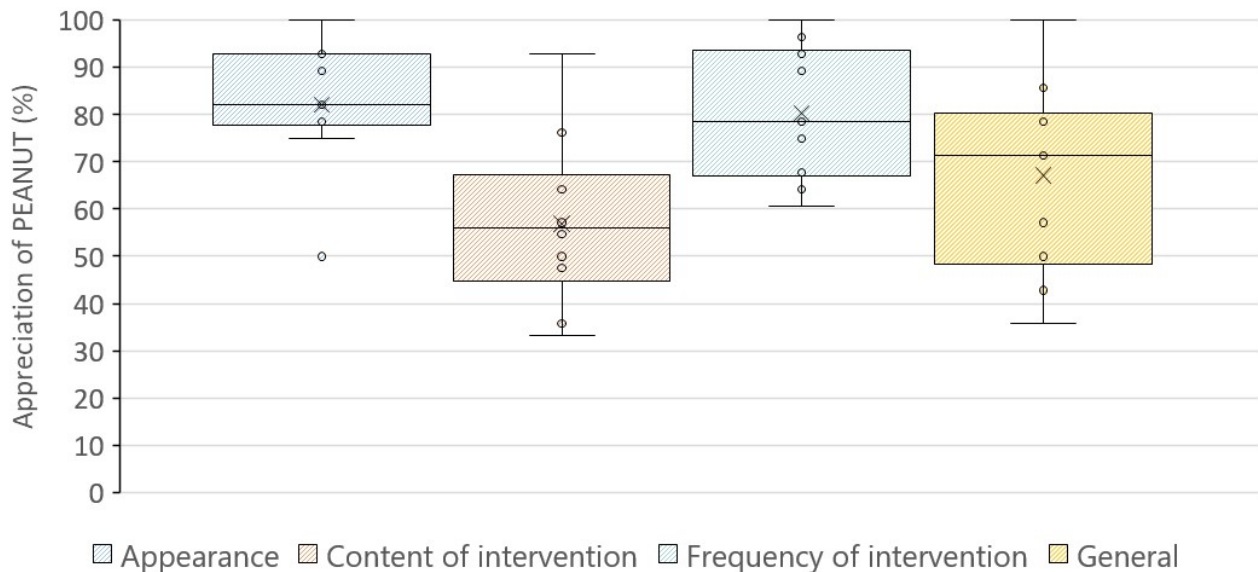
Figure 11: Percentage of appreciation of PEANUT regarding its appearance, content and frequency of intervention and general appreciation.

regarding the performance, which were not present for the control group, might also explain the observed differences. However, as the improvement was only present for the non-autonomous participants we believe that the social presence and the emotional feedback were the main factors underlying the improvement of the performances. Despite the promising results, our study suffers from the low number of participants included in it. This limitation needs to be overcome in future experiments to be able to generalize the results.

Conclusion

In this paper, we introduced the design, implementation and evaluation of the first learning companion dedicated to MI-BCI user-training: PEANUT. The strength of this experimental protocol is the design of the companion: a combination of recommendations from the literature, the analysis of data from previous experiments and user-studies. PEANUT was evaluated in an MI-BCI study (10 participants trained with PEANUT, 18 control participants, 3 sessions per participant). This study revealed that using PEANUT had an impact on performances depending on the autonomy of the users. Indeed, there seems to be a beneficial influence of PEANUT on non-autonomous persons, who were shown to have lower performances than the others in previous studies [27]. Furthermore, PEANUT tends to have a beneficial impact on the user experience. Both autonomous and non autonomous users found it easier to learn and memorize how to use the MI-BCI sys-

tem. While the specific target application explored here was MI-BCI control, many of the results could benefit other applications. First, our user studies provided useful insights about the kind of interventions, and more particularly concerning the style (exclamatory/declarative, personal/non-personal) that users prefer depending on their performance and progression. Second, our user studies suggested that the use of eyebrows favor expressiveness in cartoon-like companions, independently of BCI use, which is in line with the work of Ekman who highlighted the major influence of eyebrows for expressing numerous emotions such as happiness, surprise or anger [12].

PEANUT could potentially be used to help users train to control other applications. Since PEANUT provides interventions based only on performance and progression, it could possibly be used in other application training procedures in which these two metrics are relevant, e.g., biofeedback and physiological computing [13] or even computer-assisted motor and sports training [30], in which a social and emotional feedback should also be carefully considered [44]. To this end, we designed and implemented PEANUT for a low cost, using only open-source and free software. Ultimately, the emotional feedback and social presence could be improved by adapting it to the psychological profile of the users. For example, autonomous participants do not seem to benefit from the presence of PEANUT so it would be worth specifically studying their expectations. Emotional and social feedback could also be improved by using emotion estimation algorithms. For instance by using passive BCIs [67], which enable the extrapolation

of some mental states of the users from their brain activity, physiological computing, or emotion facial expressions from video [9, 51].

In the future, other benefits than adding social and emotional compounds to the feedback could arise from the use of learning companions in MI-BCIs [52]. Indeed, learning companions could also be used to provide a task related feedback. For example, one or several learning companion(s) could show and explain how the brain activity is modified when the user performs one task. They might also be able to provide an informative feedback. For example, an example-based learning companion might be able to provide cognitive feedback based on the previous strategies that the users tried. This could also contribute to the creation of a cognitive model, i.e., a model providing information about how the mental imagery tasks should be performed, through out the analysis of the examples' effectiveness to improve performances of the participants depending on their profile. To conclude, this experiment tested just one of the many advantages that learning companions could bring to MI-BCIs. Many more remain to be tested.

Acknowledgements

References

[1] Allison B. Z., Neuper C. Could anyone use a BCI? In: Brain-computer interfaces, 2010, 35–54.

[2] Arroyo I., Woolf B. P., Royer J. M, Tai M. Affective Gendered Learning Companions. In: AIED. Vol. 200. 2009, 41–48.

[3] Biasiucci A. et al. Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. Nature communications. 2018;9(1):2421.

[4] Boersma Paul et al. Praat, a system for doing phonetics by computer. Glot international. 2002;5.

[5] Bonnet Laurent, Lotte Fabien, Lécuyer Anatole. Two brains, one game: design and evaluation of a multiuser BCI video game based on motor imagery. IEEE Transactions on Computational Intelligence and AI in games. 2013;5(2):185–198.

[6] Carlson T., Millan Jose del R. Brain-controlled wheelchairs: a robotic architecture. IEEE Robotics & Automation Magazine. 2013;20(1):65–73.

[7] Cattell R. B., Cattell H. EP. Personality structure and the new fifth edition of the 16PF. Educational and Psychological Measurement. 1995;55:926–37.

[8] Chou C.-Y., Chan T.-W., Lin C.-J. Redefining the learning companion: the past, present, and future of educational agents. Computers & Education. 2003;40(3):255–269.

[9] D'Mello Sidney, Olney Andrew, Williams Claire, Hays Patrick. Gaze tutor: A gaze-reactive intelligent tutoring system. International Journal of human-computer studies. 2012;70(5):377–398.

[10] Duffy B. R. Anthropomorphism and the social robot. Robotics and autonomous systems. 2003;42(3):177–190.

[11] Dweck Carol S. Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In: Improving academic achievement, 2002, 37–60.

[12] Ekman Paul. Facial expression and emotion. American psychologist. 1993;48(4):384.

[13] Fairclough S.H. Fundamentals of physiological-computing. Interacting with Computers. 2009;21(1-2):133–145.

[14] Frey J., Gervais R., Fleck S., Lotte F., Hachet M. Teegi: tangible EEG interface. In: Proceedings of the 27th annual ACM symposium on User interface software and technology. ACM. 2014, 301–308.

[15] Friedrich E. VC., Neuper C., Scherer R. Whatever Works: A Systematic User-Centered Training Protocol to Optimize Brain-Computer Interfacing Individually. PloS one. 2013;8(9):e76214.

[16] Gargiulo Gaetano D et al. Investigating the role of combined acoustic-visual feedback in one-dimensional synchronous brain computer interfaces, a preliminary study. Medical devices (Auckland, NZ). 2012;5:81.

[17] Gervais R., Frey J., Gay A., Lotte F., Hachet M. TOBE: Tangible Out-of-Body Experience. In: Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction. ACM. 2016, 227–235.

[18] Graham S., Weiner B. Theories and principles of motivation. Handbook of educational psychology. 1996;4:63–84.

[19] Heutte Jean, Fenouillet Fabien, Kaplan Jonathan, Martin-Krumm Charles, Bachelet Rémi. The EduFlow model: A contribution toward the study of optimal learning environments. In: Flow Experience, 2016, 127–143.

[20] Hornecker E. The role of physicality in tangible and embodied interactions. interactions. 2011;18(2):19–23.

[21] Isen A. M., Daubman K. A., Nowicki G. P. Positive affect facilitates creative problem solv-

ing. Journal of personality and social psychology. 1987;52(6):1122.

[22] Izuma K., Saito D. N, Sadato N. Processing of social and monetary rewards in the human striatum. Neuron. 2008;58(2):284–294.

[23] Jaques P. A., Vicari R. M., Pesty S., Bonneville J.-F. Applying affective tactics for a better learning. In: ECAI. Vol. 16. 2004, 109.

[24] Jeunet C., Jahanpour E., Lotte F. Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study. Journal of neural engineering. 2016;13(3):036024.

[25] Jeunet C., N'Kaoua B., Lotte F. Towards a cognitive model of mi-bci user training. In: 7th International BCI Conference. 2017.

[26] Jeunet Camille, Lotte Fabien, n'Kaoua Bernard. Human learning for brain–computer interfaces. Brain–Computer Interfaces 1: Foundations and Methods. 2016;233–250.

[27] Jeunet Camille, N'Kaoua Bernard, Subramanian Sriram, Hachet Martin, Lotte Fabien. Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns. PloS one. 2015;10(12):e0143962.

[28] Jeunet Camille, Vi Chi, Spelmezan Daniel, N'Kaoua Bernard, Lotte Fabien, Subramanian Sriram. Continuous tactile feedback for motor-imagery based brain-computer interaction in a multitasking context. In: Human-Computer Interaction. Springer. 2015, 488–505.

[29] Johnson D. W., Johnson R. T. An educational psychology success story: Social interdependence theory and cooperative learning. Educational researcher. 2009;38(5):365–379.

[30] Jovanov E., Milenkovic A., Otto C., De Groen P.C. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. Journal of NeuroEngineering and rehabilitation. 2005;2(1):1.

[31] Karamibekr Mostafa, Ghorbani Ali A. Sentence subjectivity analysis in social domains. In: Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 01. IEEE Computer Society. 2013, 268–275.

[32] Kennedy J., Baxter P., Belpaeme T. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. ACM. 2015, 67–74.

[33] Kim Y., Baylor A. L., Group PALS, et al. Pedagogical agents as learning companions: The role of agent competency and type of interaction. Educational Technology Research and Development. 2006;54(3):223–243.

[34] Kleih Sonja C, Kübler Andrea. Psychological factors influencing brain-computer interface (BCI)

performance. In: Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on. IEEE. 2015, 3192–3196.

[35] Kothe C. Lab streaming layer (LSL). https://github. com/sccn/labstreaminglayer. Accessed on October. 2014;26:2015.

[36] Kübler A., Kotchoubey B., Kaiser J., Wolpaw J. R., Birbaumer N. Brain-computer communication: unlocking the locked in. Psychology Bulletin. 2001;127(3):358–375.

[37] Leeb R., Lee F., Keinrath C., Scherer R., Bischof H., Pfurtscheller G. Brain-Computer Communication: Motivation, aim and impact of exploring a virtual apartment. IEEE, Transactions on Neural Systems & Rehabilitation Engineering. 2007;15(4):473 – 482.

[38] Lester J. C., Converse S. A., Kahler S. E., Barlow S. T., Stone B. A., Bhogal R. S. The persona effect: affective impact of animated pedagogical agents. In: Proceedings of the ACM SIGCHI Conference on Human factors in computing systems. ACM. 1997, 359–366.

[39] Lotte F., Guan C.T. Learning from other Subjects Helps Reducing Brain-Computer Interface Calibration Time. In: International Conference on Audio, Speech and Signal Processing (ICASSP'2010). 2010, 614–617.

[40] Lotte F., Larrue F., Mühl C. Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. Frontiers in Human Neuroscience. 2013;7(568).

[41] Mathiak K. A et al. Social reward improves the voluntary control over localized brain activity in fMRI-based neurofeedback training. Frontiers in behavioral neuroscience. 2015;9:136.

[42] McFarland Dennis J, Wolpaw Jon R. Brain–computer interface use is a skill that user and system acquire together. PLoS biology. 2018;16(7):e2006719.

[43] McQuiggan Scott W, Lester James C. Modeling and evaluating empathy in embodied companion agents. International Journal of Human-Computer Studies. 2007;65(4):348–360.

[44] Mencarini E., De Angeli A., Zancanaro M. Emotions in climbing: a design opportunity for haptic communication. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM. 2016, 867–871.

[45] Meyer D. K., Turner J. C. Discovering emotion in classroom motivation research. Educational psychologist. 2002;37(2):107–114.

[46] Neuper C., Pfurtscheller G. Neurofeedback training for BCI control. In: Brain-computer interfaces, 2009, 65–78.

[47] Nkambou R., Bourdeau J., Mizoguchi R. Advances in intelligent tutoring systems. Vol. 308. Springer, 2010.

[48] Norman D. A. How might people interact with agents. Communications of the ACM. 1994;37(7):68–71.

[49] Pekrun Reinhard. The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. Applied Psychology. 1992;41(4):359–376.

[50] Pfurtscheller G., Neuper C. Motor Imagery and Direct Brain-Computer Communication. proceedings of the IEEE. 2001;89(7):1123–1134.

[51] Picard Rosalind W. Affective computing: challenges. International Journal of Human-Computer Studies. 2003;59(1-2):55–64.

[52] Pillette L., Jeunet C., N'Kambou R., N'Kaoua B., Lotte F. Towards Artificial Learning Companions for Mental Imagery-based Brain-Computer Interfaces. In: Workshop sur les "Affects, Compagnons Artificiels et Interactions"(ACAI). 2018.

[53] Plutchik Robert. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American scientist. 2001;89(4):344–350.

[54] Ramoser H., Muller-Gerking J., Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering. 2000;8(4):441–446.

[55] Reeves B., Nass C. How people treat computers, television, and new media like real people and places. CSLI Publications and Cambridge university press Cambridge, UK, 1996.

[56] Renard Y. et al. OpenViBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments. Presence: teleoperators and virtual environments. 2010;19(1):35–53.

[57] Robison J., McQuiggan S., Lester J. Evaluating the consequences of affective feedback in intelligent tutoring systems. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. IEEE. 2009, 1–6.

[58] Schmitz Michael. Tangible interaction with anthropomorphic smart objects in instrumented environments. PhD thesis. Citeseer, 2010.

[59] Short E. et al. How to train your dragonbot: Socially assistive robots for teaching children about nutrition through play. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. IEEE. 2014, 924–929.

[60] Shute V.J. Focus on Formative Feedback. Review of Educational Research. 2008;78:153–189.

[61] Silverman Barry G et al. Modeling emotion and behavior in animated personas to facilitate human behavior change: the case of the HEART-SENSE game. Health care management science. 2001;4(3):213–228.

[62] Um E., Plass J. L., Hayward E. O., Homer B. D., et al. Emotional design in multimedia learning. Journal of Educational Psychology. 2012;104(2):485.

[63] Wang Ning, Johnson W Lewis, Mayer Richard E, Rizzo Paola, Shaw Erin, Collins Heather. The politeness effect: Pedagogical agents and learning outcomes. International journal of human-computer studies. 2008;66(2):98–112.

[64] Williamson John, Murray-Smith Roderick, Blankertz Benjamin, Krauledat Matthias, Müller K-R. Designing for uncertain, asymmetric control: Interaction design for brain–computer interfaces. International Journal of Human-Computer Studies. 2009;67(10):827–841.

[65] Wolpaw J, Wolpaw E.W. Brain-computer interfaces: principles and practice. Oxford University Press, 2012.

[66] Woolf B. P., Arroyo I., Cooper D., Burleson W., Muldner K. Affective tutors: Automatic detection of and response to student emotion. In: Advances in Intelligent Tutoring Systems, 2010, 207–227.

[67] Zander T.O., Jatzev S. Detecting affective covert user states with passive brain-computer interfaces. In: Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. IEEE. 2009, 1–9.