Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality

LISA-MARIE VORTMANN and FELIX PUTZE, Cognitive Systems Lab, University of Bremen, Germany

Adding attention-awareness to an Augmented Reality setting by using a Brain-Computer Interface promises many interesting new applications and improved usability. The possibly complicated setup and relatively long training period of EEG-based BCIs however, reduce this positive effect immensely. In this study, we aim at finding solutions for person-independent, training-free BCI integration into AR to classify internally and externally directed attention. We assessed several different classifier settings on a dataset of 14 participants consisting of simultaneously recorded EEG and eye tracking data. For this, we compared the classification accuracies of a linear algorithm, a non-linear algorithm, and a neural net that were trained on a specifically generated feature set, as well as a shallow neural net for raw EEG data. With a real-time system in mind, we also tested different window lengths of the data aiming at the best payoff between short window length and high classification accuracy. Our results showed that the shallow neural net based on 4-second raw EEG data windows was best suited for real-time person-independent classification. The accuracy for the binary classification of internal and external attention periods reached up to 88% accuracy with a model that was trained on a set of selected participants. On average, the person-independent classification rate reached 60%. Overall, the high individual differences could be seen in the results. In the future, further datasets are necessary to compare these results before optimizing a real-time person-independent attention classifier for AR.

CCS Concepts: • **Human-centered computing → Mixed / augmented reality**; **Interactive systems and tools**; Ubiquitous and mobile computing systems and tools; • **Computing methodologies → Online learning settings**; *Modeling methodologies.*

Additional Key Words and Phrases: EEG, attention, person-independence, eye tracking, augmented reality

ACM Reference Format:
Lisa-Marie Vortmann and Felix Putze. 2021. Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 80 (June 2021), 27 pages. https://doi.org/10.1145/3463507

1 INTRODUCTION

Augmented reality (AR) devices enable the embedding of visual content in a real-world surrounding. Their promising future can be seen in the rise of scientific and public interest. Searching for "Augmented Reality" on Google Scholar gives 88.200 results before the year 2000. This increased to 236.000 publications between 2000 and 2009 and reached 506.000 registered publications between 2010 and 2019. With the advance of computational power and mobile technologies, AR becomes increasingly pervasive and allows for new interesting applications on different devices.

Currently, most end users probably experience AR mainly through their smartphone camera. It is a very conscious decision to interact with the device and often requires holding up the smartphone and pointing the camera

Authors' address: Lisa-Marie Vortmann, vortmann@uni-bremen.de; Felix Putze, felix.putze@uni-bremen.de, Cognitive Systems Lab, University of Bremen, P.O. Box 1212, Bremen, Germany, 28359.

2474-9567/2021/6-ART80 $15.00
https://doi.org/10.1145/3463507

towards points of interest. In contrast, AR glasses can make use of a head-mounted display that eliminates the necessity of holding anything in hand. However, continuous visibility brings advantages and disadvantages for usability. While it makes interaction more effortless, it introduces a higher risk of distraction. In times of reminiscence, memory recall, mental arithmetic, or other forms of thoughts, the user's attention is directed internally (internal attention) and many changes of the augmented reality content could disrupt these processes. For some applications, it could be important to detect times of external attention (observation, visual search, general attentiveness to one's surroundings) and adapt the user interface accordingly.[1]
Depending on the AR application and current scenario, internally directed attention can be desired or undesired. If it is required to think about the presented information or situation, new visual or auditory input would possibly interrupt the thought process. The central purpose of attentional mechanisms is the focus on relevant information while supposedly irrelevant information is suppressed. The difficulty of the suppression of the sensory input increases with increasing perceptual saliency [19], and while the appearance of possible distractors in the real surroundings can not be manipulated through an Augmented Reality device, the virtual content that is displayed should be sensitive to the attentional state of the user and the goal of the application. Contrastingly, if the application is aware that the presented content is relevant and important but the user's attention is directed internally (i.e. mind-wandering, task-irrelevant thoughts), an AR application should react and actively regain the user's attention by increasing object saliency, or wait with relevant information until the user's attention is directed externally again. An example application is three-dimensional learning content visualized and animated in Augmented Reality that pauses its presentation when internally directed attention is detected. Through this attention-awareness, the possible task-relevant thoughts are not interrupted and important external information is not missed.

The mentioned example application illustrates the importance of the "real-time" aspect of the classification. Switching from one attentional state into another happens fast and can happen often. In a ten-second interval, the attentional focus can change several times. Therefore, the application should be able to react quickly and base its decision on short time intervals that most likely do not contain multiple states.

In this work, we explored different possibilities for a machine learning-based classification of internal and external attention. When questioned about their attentional state, test participants may not remember their state (as it was subconscious or forgotten), or they may simply be unable to verbalize it. Through the examination and visualization of recorded user data, the causes for behavior can be found without relying on the fallible human memory. The attentional state is subject to fast and frequent changes and the conscious reports of such would themselves required attention shifts and disrupt fluent task solving. Thus, since explicit labeling of attentional states by the user or the experimenter is unreasonable, we rely on decodable biosignals in real-time. For this purpose, we used electroencephalography (EEG) and eye tracking (ET) data. Several studies claim that EEG and ET features complement each other well and their combination can improve classification results for mental states ([7, 8, 18, 23, 31, 35])

EEG data reflects electrical brain activity with a high temporal resolution, which makes it a favored measure for Brain-Computer Interfaces despite its lower spatial resolution and higher setup time compared to methods that measure oxy- and deoxyhemoglobin concentration (i.e fMRI and fNIRS). The spatial resolution of EEG recordings can be increased to a certain level by adding more electrodes to the head. However, this increases the setup time and decreases the usability for BCIs in many situations because of the decreasing comfort. Thus, neurophysiological knowledge about attentional states is needed to improve the electrode positions for a suitable

[1]For a more detailed taxonomy of internal and external attention see Chun et al. [11]

setup with few electrodes. A further advantage of EEG recordings for this purpose is the possible mobility of the system and the fact that the system is secured on the head which allows for moderate movements without signal loss.

Eye tracking, on the other hand, measures gaze behavior which is usually quantified by fixations, saccades, blinks, pupil dilation, and gaze speed. It has been shown in the past, that eye gaze behavior changes with changing perceptual and cognitive demands. These findings led to the interpretation of eye behavior as distinctive for different mental states [9]. The effect of the attentional state, specifically internally and externally directed attention, on gaze behavior can also be used for classification. Eye tracking devices are comparatively cheap and easy to set up. In the case of head-mounted AR displays, the tracking cameras can be integrated into the headset or easily attached to it without decreasing the comfort of the user. The obvious advantages are contrasted by high task-specificity, especially if a certain viewing behavior is required by the task, and more difficulties drawing fine-grained conclusions (which could be necessary for successful classification) that generalize over participants because the viewing strategies across participants might differ.

Overall, the topic of generalizability over participants is of high interest for the combination of BCIs and Augmented Reality. If a high classification correctness and real-time monitoring of the attentional state are possible, the two main issues that stand in the way of this technology being widely used are a high discomfort of the setup and long calibration and training times before the actual usage. For many BCIs, the classifier has to be trained on person-dependent data for the accuracy to be high enough to allow for a helpful application. Thus, models are usually very individual for each user and not generalizable over participants. The data for the classifier training has to be collected in long training sessions where the recorded data can be labeled explicitly because the classification algorithm needs to learn something from the data. Person-independence, however, would exclude the need for previous training data collection and allow using the system "out-of-the-box" or with a shorter calibration interval (i.e. for eye tracking, calibration is usually necessary and EEG often requires impedance improvements). If the data were person-independent, the classifier could be pre-trained on data from several other users, increasing the available amount of training data and hence, increasing the variance in the data and reducing the bias or potential overfitting on only a few training runs.

In a previous work, we described an AR paradigm for the collection of data during times of internally and externally directed attention [46]. We were able to show that the data collected during the task was classifiable into the two attentional states using a person-dependently trained Linear Discriminant Analysis (LDA) for power spectral density based EEG features generated over a time interval of 13 seconds. The experimental paradigm was a spatial alignment task in AR with one condition requiring internally directed attention and one condition requiring externally directed attention. For this work, we shifted our focus from proving the initial validity of the paradigm to a detailed analysis of modalities, data segmentation, classifier choice, and most importantly: systematically exploring the potential for person-independent classification that could be applied to a real-time setup in the future. Accordingly, we added the modality of eye tracking to our analysis to test whether it is better suited for a person-independent approach or whether a multimodal approach would increase the classification accuracy. As another way to improve the classification accuracy, we compared the previously used LDA to a non-linear algorithm (Random Forest) and a simple Neural Net that were trained on the same feature set, as well as a shallow convolutional neural net that was trained on the raw EEG data. For the aspect of real-time analysis, we assessed using shorter data intervals for the classification. The generalizability over participants was analyzed using a pooled dataset of all participants before running a person-independent classification analysis with either a leave-one participant out training or training on a set of selected participants that had the highest person-dependent classification accuracies.

1.1 Related and Preliminary Work

Within the wide field of attention, this study focuses on the detection of internally and externally directed attention. A detailed taxonomy, including classic debates on the topic and new issues, was published by Chun et al. [11]. They acknowledge the complexity and ubiquity of the field and provide an organized framework that characterizes the terms. To the best of our knowledge, there is no previous study that reports results of a person-independent BCI for internal and external attention differentiation but several studies have reported neurophysiological mechanisms of these two states.

In 1985, Ray and Cole [34] already reported that alpha power in the EEG (8-14 Hz) decreases in sensory-intake tasks and suggested that a high alpha activity reflects strong internal attentional focus. The same effect was reported by Cooper et al. [12]. The authors showed that increased alpha-band power is recorded during times of internally directed attention tasks and conclude that this shows that external stimuli are actively suppressed. In a study from 2014, Benedek et al. [3] found differences in the frequency power spectrum in the right parietal brain region for internal and external attention. During times of high internal attention demands, the alpha power in the right parietal cortex increases which, according to the authors, might reflect the activity of the ventral attention network. Harmony et al. [21] performed a narrow filter band analysis based on data recorded during a mental arithmetic task and the Sternberg paradigm and supported their hypotheses that increased delta band activity (1-4 Hz) is an indicator of internal attention processes. Using lateralized power spectra for a spatial attention task, Van der Lubbe et al. [45] found that both internal and external attention induce early posterior increased contralateral theta power (4-8 Hz) and late posterior increased ipsilateral alpha power, whereas only external spatial attention relates to posterior contralateral negativity. Arguably, the classification of these attentional states should be possible based on EEG data. Indeed, Putze et al. [32] recorded EEG data in a computer-based task and reliably classified internal and external attention on a single-trial basis. This classification was performed in a person-dependent fashion. As reported before, a similar person-dependent classification based on EEG data was successful in Vortmann et al. [46]) for an AR paradigm.

The second modality that we will analyze in the context of person-independent BCIs for internal and external attention awareness in Augmented Reality is eye tracking. As mentioned before, eye gaze behavior is influenced by the cognitive processes of the user. Faber et al. [15] and Bixler et al. [5] detected phases of mind-wandering during reading tasks based on fixations, saccades, blinks, and pupil dilation compiled over 12-second windows. Hutt et al. [22] focused their study on the classification of mind wandering phases during lecture viewing. Their Bayesian Network classifier outperformed a chance-classifier if it was trained on global gaze features, such as number and duration of fixations. A detailed review of these and other features for internally and externally directed attention was given in Annerer-Walcher et al. [1]. The authors described previous findings for pupil diameter, pupil diameter variations, fixation disparity, fixation disparity variation, blinks, saccades, and microsaccades dependent on the attentional state, and tested the features for a newly collected dataset of internal and external attention. The same team also performed an EEG and eye tracking co-registration study for internally and externally directed attention, published in Ceh et al. [10]. They concluded from the results that EEG activity and eye tracking are well suited for internal focus detection and that EEG alpha power is correlated with pupil dilation, suggesting that they are involved in a neurophysiological gating mechanism that serves for shielding internal cognition from irrelevant sensory information. That eye tracking data alone can be used to differentiate internal and external attention reliably for person-dependent analyses in a computer-based setting was shown in Benedek et al. [4].

A first real-time classifier for internal and external attention was developed in Vortmann et al. [49]. However, the classifier was person-dependent and thus, required the previous recording of a large amount of training data. Additionally, the classification accuracies were mediocre for the real-time system. We are not aware of a person-independent EEG classifier for internal and external attention. However, a few approaches for person-independent classification of other cognitive phenomena related to attention have been made. Fazli et al. tested a classifier for motor imagery based on spatio-temporal filters on two different datasets (83 recordings in [16] and 45 participants in [17]). Their results in both studies show that the classification is possible. Zhang et al. [51] worked with four different movement intentions and were able to classify them better than baseline using a convolutional recurrent attention model. In Pandey et al. [29], a multilayer perceptron was used for person-independent emotion recognition, and in Pandey et al. [28], the authors used wavelet transform features for a deep neural network to classify emotion based on valence and arousal independent of the subject.

There have been several approaches for applications that used attention-related processes to adapt an AR interface without using a BCI. For example, Lu et al. [26] reported the facilitation of visual search in AR when effective subtle cueing methods were used. Bonanni et al. [6] used a layered interface to reduce the cognitive load of the user in AR. The interface was designed following cueing and search principles of attention theory. While these studies only used knowledge about attention, the introduction of neural information about the attentional state would have even more benefits.
Si-Mohammed et al. [40] reported the current state-of-the-art for combining BCIs and AR and assessed the general feasibility and design of their combination with 4 studies [41]. They conclude that this interaction will be the future of AR.

To test the idea that a differentiated interface reaction depending on internal or external attention would improve the usability of the system, we also performed a preliminary experiment to compare an attention-aware system with an attention-unaware system. As proof of the overall concept, we ran this pilot experiment that had participants rate the distraction and usability of the two systems. The full study was reported in Vortmann et al. [47]). The basic system was a Steady-State Visually Evoked Potential (SSVEP)-based Smart-Home System in AR that reacted independently of the attentional state (introduced in [33]). This system was compared to the same system where we added real-time classification of internal and external attention (introduced in [48]. In the attention-aware system, the display of objects in AR was limited to times of external attention. The goal was to show that the inclusion of such classification-based adaptations improves the usability of the system. The average System Usability Score (SUS) of the attention-aware system was significantly higher than the SUS of the attention-unaware system. The same held for the distraction rating, where the attention-aware system was rated significantly less distracting.
These conclusions motivate us to improve our attention classifier. Its improvement in the usability of AR applications might help AR devices on their path to ubiquity. One limitation of the system is the fact that the classifier required person-dependent calibration, did not make use of the ET data, and its parameters were not optimized with regards to classifier selection, window size, and others.
The exploration of person-independent EEG-BCIs for attentional processes is still a very young field but worth a deeper dive.

1.2 Hypotheses and Research Questions

The overall goal is to find an optimal, robust setup and analysis for a real-time classification that does not require previous classifier training.
We hypothesize that (a) multiple modalities improve classification accuracy, (b) short time intervals of data

suffice for reliable classification results, that (c) brain activity and eye gaze behavior follow the same patterns for all participants during the two attentional states, and thus, (d) classifier trained to detect internal and external attention should be able to reliably classify new datasets person-independently. The motivation for all performed analysis steps and comparisons will be explained in detail in the following part by posing our research questions.

Based on the preliminary study, the related work done on this topic, and our hypotheses and goals, we decided on several research questions we want to answer and the steps we want to follow to achieve this.
As a first follow-up analysis of the paper by Vortmann et al. [46], we analyzed the collected eye tracking data with a Linear Discriminant Analysis (LDA). Eye tracking devices are easier to set up and calibrate than EEG systems and would, therefore, improve the users' comfort. First, we checked whether the EEG and ET classification results correlated, or whether single features correlated with each other. This could suggest that ET artifacts in the EEG data influence the EEG classification. One of our main research questions was whether classification purely based on ET data would lead to similar or better results in the classification of internal and external attention compared to EEG data and whether a combination of the two modalities would increase the achievable accuracy of a single modality.

As a next topic, we addressed the question of which time window length was best suited for a high accuracy while remaining close to real-time analysis. The user can switch back and forth between internal and external attention quickly, which means that long time windows would contain periods of both states. Shorter epoch lengths would mean a better estimation of the current state but could cause a decrease in accuracy due to fewer data samples. This goes hand-in-hand with the question, whether a higher number of time windows (and thus feature sets) for the training of the classifier increases the accuracy of the predicted state or whether the overall length of the training data in seconds is more important. The answer to this question would determine whether cutting a short training period in several small time windows would have the same effect as collecting longer time windows in a longer training session.

In their 10-year update paper of "Classification Algorithms for EEG-based Brain-Computer Interfaces", Lotte et al. [25] suggested that smaller amounts of data (as only available in this study) are classified best using an LDA or a Random Forest classifier. Thus, we decided to compare the results that were achieved using the LDA with the classification accuracy of a Random Forest classifier. Additionally, we wanted to know whether a Neural Net approach would lead to more accurate results than our linear models or the Random Forest. Appriou et al. [2] suggested that Neural Networks work better for EEG data during workload classification, outperforming linear approaches. For better comparability, we first choose to train a vanilla neural network with the same feature set as the shrinkage LDA and the Random Forest to see if it has any advantages if supplied with the exact same features. On top of that, we used a convolutional neural network that was specifically designed for the classification of raw EEG data and compared it to the other classifiers. Given that they are trained on different input data (a generated feature set that lost time resolution vs. the raw time series of each electrode) the comparison is less meaningful. However, since we are mainly interested in finding the best performing classifier setup, the absolute classification accuracies will be a good comparison metric. For the convolutional neural net, we decided to use the shallow Filter Bank Common Spatial Patterns (FBCSP) Neural Network from Schirrmeister et al. [39].

After answering all these questions on a person-dependent level, we wanted to know how the effects play out in datasets that pool multiple participants (in contrast to person-dependent classifiers). This would answer the question of how generalizable the collected data is. In a "pseudo"-person-independent analysis of the pooled datasets, we evaluated possible advantages of the processing, the datasets, and model fitting for a possible calibration-free classification.
Ultimately, all previous results were considered to attempt person-independent classification of internal and

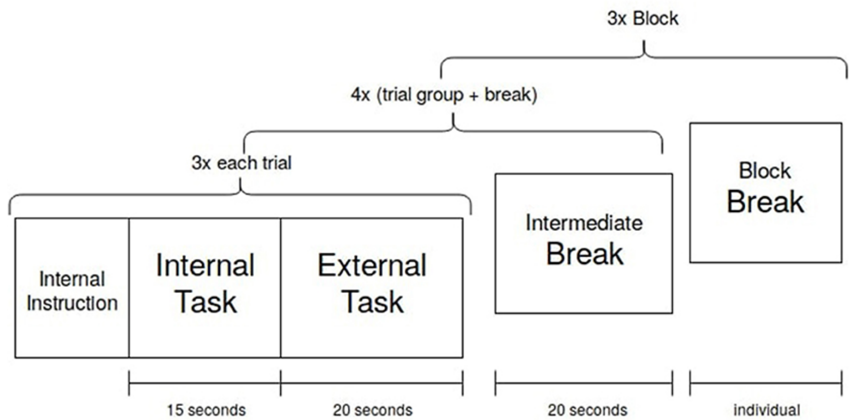


Fig. 1. Experimental procedure taken from Vortmann et al. [46].

external attention in the given AR task. We trained several different versions of the models that were based on different subsets of participants or all but one participant (leave-1-out approach).

2 THE DATA SET

The data set analyzed in this paper was previously recorded and published in the study by Vortmann et al. [46]. We refer to their paper for a detailed description of the task implementation and data collection.

The recordings were executed in an uncontrolled office setting to simulate realistic usage. While performing the task, the participants wore an EEG cap and the Microsoft HoloLens with an attached eye tracker. The paradigm induced phases of internal and external attention in the AR setting.

2.1 Experimental Task

A spatial perspective alignment task was implemented for the HoloLens as a paradigm for the data collection. In the task, the participant had to keep a ball and a tube visually aligned by slightly moving their upper body. The participant was instructed to keep the moving virtual ball "inside" the fixed virtual tube in front of them. Both objects were visible during the external trials. Thus, externally-directed attention was required for good task performance. During internal trials, neither the ball nor the tube were moving but still displayed on the side of the visual field. However, the participant was asked to imagine the tube in the same position as before. The movement of an imagined ball was described to the participant by a sequence of numbers. Each number represented a position on the screen and an endpoint of the ball movement. The participant was instructed to move their upper body following the imagined spatial alignment. To achieve a good performance, internally-directed attention on the imagined objects was necessary.

In total, one session contained 36 internal and 36 external trials in an alternating order with breaks in between. The internal trials lasted 15 seconds and the external trials 20 seconds (see Figure 1). Before the recording of the real trials, every participant got accustomed to the task in a tutorial that was individually paced. Therefore, the recorded session should only contain trials that were performed correctly.

2.2 Participants

Fifteen healthy participants (mean age 27.4 ± 10.4; three females) were recruited for the experiment. All had normal or corrected to normal vision, were right-handed and all but three participants had previously used an Augmented or Virtual Reality Device. Technical problems during three of the sessions led to the exclusion of one participant and two sessions with a reduced trial number (54 and 48 instead of 72; marked * in subsequent tables).

2.3 Data Collection

Every participant filled out a demographic and a mind-wandering-related questionnaire (MWQ) before the experiment. The MWQ was presented and validated in Mrazek et al. [27]. It contains questions concerning the probability of a person falling into a state of wind-wandering in different settings and situations. The results of the questionnaire were collected to test if any participant had a suspiciously high score and if their data during the experiment was influenced by this.

During the task, interaction data, movement data, EEG, and eye tracking data were recorded.

The EEG data recording was performed at 500 Hz with a 16-channel wet electrode cap from g.tec nautilus (electrode positions: Cz, FP2, F3, Fz, F4, FT7, C3, FP1, C4, FT8, P3, PZ, P4, PO7, PO8, Oz of the 10-20 system). Impedances were kept below 20 kΩ with CZ as a recording reference.

A binocular, wearable Pupil labs eye tracker was used to record eye gaze during the task. The gaze points and detection confidence were sampled at 120 Hz, using the provided recording software. In the first step, the two eye tracking cameras that were fixed to the HoloLens were adjusted in depth and angle individually for each participant. For the calibration, a fixed gaze task with 9 sequentially shown points was displayed on the HoloLens[2].

After the sessions, the participants filled out a short questionnaire about their workload perception during the task.

3 PREPROCESSING

The eye tracking and EEG data were recorded simultaneously but preprocessed individually. The simultaneous recording was managed through LabStreamingLayer (LSL) [38], which led to synchronized timestamps on both modalities and an additional marker stream. The marker stream contained all necessary information about the current events in the experimental setup (i.e. internal trials start/end, tutorial end, break). The preprocessing includes all steps from the raw data to feature vectors for the specified time windows. The whole process can be seen in Figure 2.

3.1 EEG

After extracting the EEG data from the LSL output, the raw signal was processed using the MNE Toolbox [20]. A notch filter was applied at the power frequency of 50 Hz. The filter length was chosen automatically, as suggested by MNE, based on the size of the transition regions (3.3 x reciprocal of the shortest transition band). Additionally, the signal was band-pass filtered between 1 and 50 Hz with a FIR filter design using the window method with the same filter length as the Notch filter. Broken channels were excluded after a visual examination and interpolated before computing and applying the common average reference (CAR).

The information about internal and external trial onset was taken from the markers and used to cut the EEG data into different labeled windows. The first second after the trial onset was never used to avoid interference of unwanted artifacts, related to the instruction, audio signal, or belated reaction. To evaluate the effect of the length of an EEG window, we generated several non-overlapping windows (epochs) from each trial with different interval lengths. The chosen epoch lengths were 13 seconds, 7 seconds, 4 seconds, 2 seconds, and 1 second. Each epoch was baseline-corrected based on the first 0.5 seconds after the epoch onset. Label information was saved

[2]Pupil Labs provides a library for Unity – https://github.com/pupil-labs/hmd-eyes

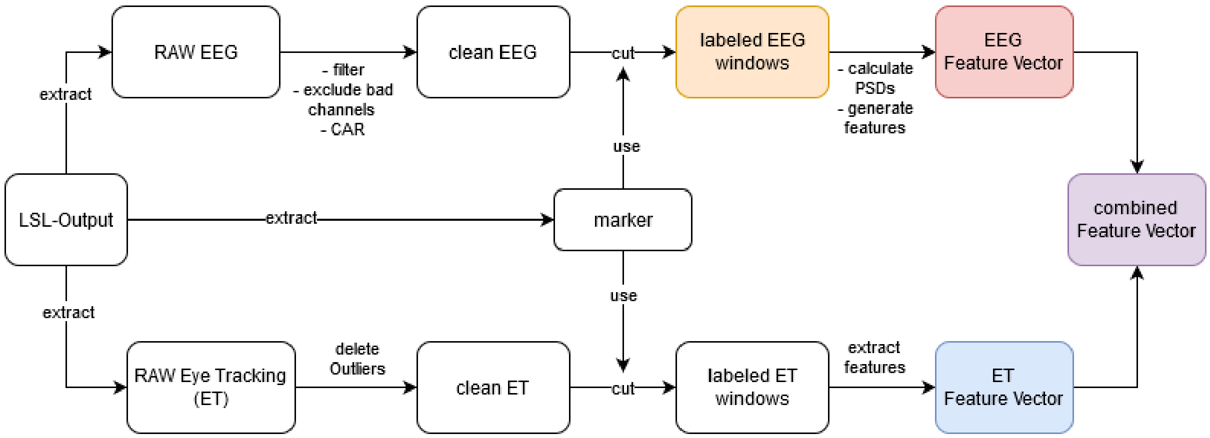


Fig. 2. Preprocessing pipeline for EEG and eye tracking (ET) data from experiment output to feature vectors. Yellow= Output used for sFBCSP-NN; Red = Output used for only EEG related analyses; Blue = Output used for only ET related analyses; Purple = Output used for combined modality analyses

with each epoch.

The features for the epochs were generated based on the Power Spectral Densities (PSDs) of each EEG channel in the epoch. To calculate the PSDs, we used the multitaper method. We computed the average and maximum density for the Theta- (θ: 4-8 Hz), Alpha- (α: 8-14 Hz), Beta- (β: 14-30 Hz), and lower Gamma-band (γ: 30 - 45 Hz) and used them as features. The total number of features for one epoch is *number of channels x number of frequency bands x 2 (average/max)*, which results in 16 x 5 x 2 = 160 features.

3.2 Eye Tracking

During the experiment, the Pupil Labs eye tracker recorded the confidence for the measurement as well as the x- and y-coordinate of the currently generated gaze point. All points whose coordinates exceeded the range 0-1 were deleted and instead set to 0 or 1. This range was specified by the manufacturer and all points outside these thresholds should be considered as outliers and filtered from the data, as recommended. We assumed that the proportion of filtered/unfiltered data could contain important information about the attentional state and thus decided, to include it as a feature in our feature set. The gaze point is calculated based on the binocular data.

After this cleaning, the ET data was cut into smaller labeled windows in the same manner that the EEG data was windowed.

For each epoch, the following 14 features were generated from the ET data: Proportion of filtered/unfiltered data, total distance covered by the eyes, average gaze speed (overall gaze distance/window length), the variance on the x-axis, variance on the y-axis, mean x-location, mean y-location, the average length of fixations following Salvucci et al. [37], average number, speed and length of saccades (Engbert and Kliegl algorithm [14]), average number and length of fixations following Smeets et al. [42], average confidence of the eye tracker. All features were used for the combined feature vector.

4 THE CLASSIFICATION

One of the research questions guiding this paper was what effect the classifier choice would have on the accuracy of the classification. A Linear Discriminant Analysis was chosen in Vortmann et al. [46] based on the argumentation of Wang et al. [50] that it works sufficiently well for binary classification tasks. For the comparison, we chose to

also include a non-linear state-of-the-art classifier: A Random Forest (suggested by Lotte et al [25]). A study by Appriou et al. [2] suggests, however, that a neural network approach leads to better results in EEG classification. We implemented a vanilla neural network that classifies the same generated feature set as the LDA and the Random Forest, as well as a more complex shallow convolutional neural net that is based on a Filter Bank Common Spatial Pattern analysis and works with raw EEG data. Thus, we chose to compare four classification approaches by performing our initial analyses with all of them.
The evaluation of the results was always based on the calculated accuracy of how well the classifier predicted the label of the training set. The two classes (internal and external) were equally frequent in the data set and test and training splits were stratified to keep this distribution. The chance level to guess the correct class was 50%, which was used as the baseline.

4.1 Linear Discriminant Analysis

For the implementation of the Linear Discriminant Analysis, the scikit-learn toolbox was used [30]. The size of training and testing sets varied, depending on the current research question. The general approach was using 50% of the data for training in a shuffled but stratified fashion. However, we also examined the effect of the training data amount (see Section 5.3.1). The training data was z-normalized. The LDA was computed with a least-squares solver and an optimized shrinkage factor. This hyperparameter optimization of the shrinkage factor was performed on the training data in a shuffled, stratified 10-fold cross-validation for each participant and the best performing configuration was used for the classifier training afterward. The same configuration of the LDA was used to classify both EEG and ET data.

4.2 Random Forest

The Random Forest was also implemented with the Scikit-Learn toolbox [30]. All training and testing settings were the same as in the LDA approach but no hyperparameter optimization had to be performed. During the classification pipeline, we forewent normalizing the training data because Random Forests work better on not-normalized data. The feature vector that was generated on the power spectral densities of each electrode was used for the training and testing and the maximum tree depth was set to 40 while the number of estimators was set to 100. These parameters were systematically tested on a set of participants and then fixed for all participants.

4.3 Vanilla Neural Net

The simple Vanilla Neural Net was implemented in Tensorflow and Keras. It consists of two hidden layers with a sigmoid activation function. The Adam optimizer uses a fixed learning rate of 0.0015 and the loss is computed using the binary cross-entropy. Again, the same training and testing split strategies as for the LDA were used. The model was trained for at most 1000 epochs, with an early stopping callback in case the training accuracy did not go up for at least 30 epochs. The batch size was set to 1500.

4.4 Shallow Filter Bank Common Spatial Patterns Neural Network

The shallow FBCSP Neural Network (sFBCSP-NN) was implemented with the braindecode toolbox [39]. It is a convolutional neural network that is designed to use the spectral power as features, which is inspired by a pipeline to extract FBCSP. After a temporal convolution layer and a spatial filtering layer, the network performs a mean pooling on the sixths layer, has a drop-out layer with a drop probability of 0.5, and applies a logarithmic activation function to achieve a linear classification result (see Table 1). The suggested cropped training strategy by Schirrmeister et al. [39] was applied. The learning rate was set to 0.015 and the weight decay between the layers was set to 0. This parameter optimization was preliminarily performed on selected data and the parameters were kept the same across all analysis runs. Each model was trained for 100 epochs. The sFBCSP-NN was only

Table 1. Architecture description of the Shallow FBCSP Neural Net based on an exemplary EEG input length of 2000 samples. Implemented in PyTorch.

Layer	Name	Type	Output Shape	Num. Parameters	Parameters in %
1	dimshuffle	Expression	[-1, 1, 2000, 16]	0	0
2	conv_time	Conv2d	[-1, 40, 1977, 16]	1040	2,82
3	conv_spat	Conv2d	[-1, 40, 1977, 1]	25600	69,41
4	bnorm	BatchNorm2d	[-1, 40, 1977, 1]	80	0,22
5	conv_nonlin	Expression	[-1, 40, 1977, 1]	0	0
6	pool	AvgPool2d	[-1, 40, 1903, 1]	0	0
7	pool_nonlin	Expression	[-1, 40, 1903, 1]	0	0
8	drop	Dropout	[-1, 40, 1903, 1]	0	0
9	conv_classifier	Conv2d	[-1, 2, 13, 1]	10162	27,55
10	softmax	LogSoftMax	[-1, 2, 13, 1]	0	0
11	squeeze	Expression	[-1, 2, 13]	0	0
Total				36882	

used to classify EEG data, not for ET data because it is built especially for EEG data. Finding a suitable Neural Net to classify the recorded eye tracking data is still subject of ongoing research and will this not be part of this work.

5 RESULTS

The main goal of this work is to explore the effects of different choices during the recording, the processing, and the classification. To evaluate these effects, we perform several comparisons of settings. For more clarity, the methods used for comparison will be described immediately before we report the results, and the implications drawn from the results will be mentioned immediately afterward. The order of the report reflects the iterative process that was followed to answer our research questions. The first analyses were performed in person-dependent fashion, followed by similar analyses on a dataset that was pooled over all participants, and finally, true person-independent analyses with settings that proved most successful in the previous comparisons. When we average our results over participants, we do not only report the average classification accuracy but a 95%-confidence interval of the estimated accuracy. This way, we get a more realistic and reliable statement from a comparatively small dataset.

Participant 14 was left out for some of the analyses because the recorded dataset contained fewer trials and was not usable for some comparisons. A significance level of $\alpha = 0.05$ is assumed. For significance testing, paired two-tailed t-tests were applied.

5.1 Effect of Modality Choice

The first comparison explores the effect of modality choice on classification accuracy. In Vortmann et al. [46], the EEG data was already compared to a classification based on movement data recorded during the experiment. The movement data was recorded by the HoloLens and entailed the information about the current head position and rotation in relation to the virtual objects. In this work, we compare the EEG results to results obtained from classifying eye tracking data and from a combined feature vector that includes the features used for EEG analysis and the features used for ET analysis. For this multi-modal approach, we considered an early fusion of the features. In contrast to late fusion, an early fusion can exploit the relations between the tightly synchronized modalities. The features for EEG and ET were generated as described in Section 3.

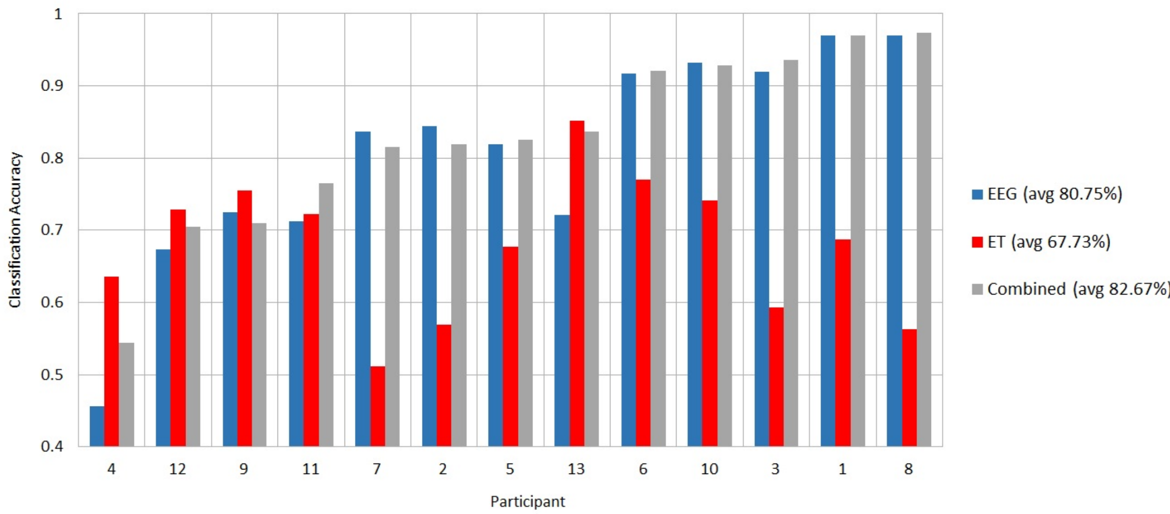


Fig. 3. The average classification results of 100 runs per participant. Performed with only the EEG feature vector, only the ET feature vector or a combined feature vector. Sorted increasingly by best result for a combined feature vector. Classifier: LDA, Test-Training-Split: 1:1, Epoch length: 13 seconds

The features were generated on 13-second windows of the cleaned data and split into 50% training and 50% testing data in a random but stratified fashion. A new LDA was trained and tested for 100 different splits per participant. The average classification accuracy per participant can be seen in Figure 3 for both modalities and their combination. Based on these individual results, the 95%-confidence interval for the mean EEG classification accuracy was [0.73, 0.88], for ET [0.63, 0.73] and for a combined feature set [0.76, 0.89]. The accuracy of the predictions based on ET data outperformed the EEG-based classification for 5 of the 13 participants. The combined feature set improved the classification accuracy compared to a single modality feature set in 6 of the participants. The combination is significantly better than only ET features with $t(12)= 3.2801, p<.001$. However, the improvement is not significant compared to the EEG feature-based classification ($t(12)=1.6497, p=.1249$).

5.1.1 Correlation between Modalities. We want to examine, whether the information that is classified in the EEG data might be a neural response caused by eye movements or eye movement artifacts. If the neural correlates of the eye movements were a reliable help in the classification process, the results for the two modalities should be statistically dependent in a linear way. Therefore, we compared the ET and EEG classification accuracies pairwise for each participant (Figure 4). A correlation analysis yielded that there is only a weak linear correlation between the results (*Pearson's r = -0.2*). On top of that, we compared the predictions for each test trial of each participant after splitting the EEG and Eye Tracking data into the same training and test sets. For all participants, this analysis showed neither a strong positive nor a strong negative correlation between the predictions for each trial. We conclude that eye movements do not play a mayor role in the classification process of the EEG data.

5.1.2 Correlation between Features. For a more detailed insight into the role of eye movements and their artifacts in the classification process of EEG data, we analyzed the correlations between the individual features for every participant. The correlation matrices were examined visually for abnormalities or unexpected correlation patterns.

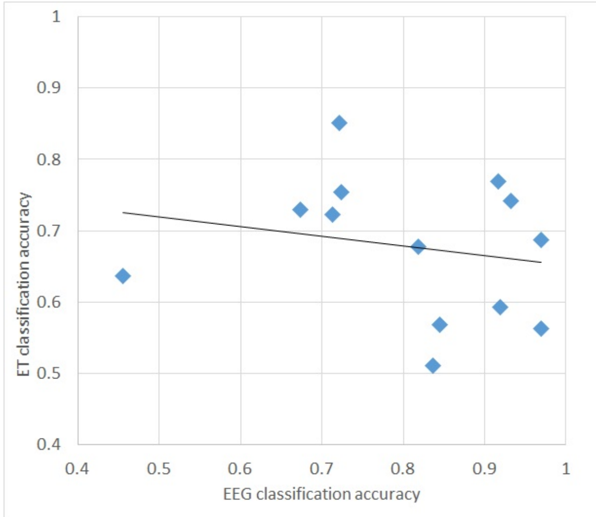


Fig. 4. ET and EEG results for each participant (represented by a square). The Pearson's correlation of r = -0.2 is visualized by a linear approximation of the data points to show the trend in the data.

For further insight into feature importance and selection in the classification process, refer to the previous paper by Vortmann et al. [46].

The correlation matrix of participant 1 in Figure 5 shows high correlations within ET and within EEG features but low correlations between the two subsets. This behavior is as expected. Participant 1 was chosen because it has one of the best classification accuracies for ET and for EEG data.
The correlation matrix of participant 9 instead, shows strong negative correlations between some of the ET features and the EEG features (see Figure 5). Additionally, there seems to be no or only a weak correlation between some related EEG features (adjacent features represent the same frequency band for adjacent electrodes). Participant 9 had among the lowest classification accuracies. These two participants had the most striking correlation matrices. Most participants' matrices are somewhere between these extremes.
We conclude that eye movement artifacts or other recording problems could weaken the classification performance and that the EEG classifier does not improve by learning eye artifacts.

Overall, the substitution with eye tracking features, or the addition thereof, did not significantly improve person-dependent classification accuracies for internal and external attention in the AR paradigm for person-dependent classification. Whether there is a common pattern among participants for each modality is crucial for the success of a person-independent classifier. If the inter-participant differences are high, the predictions for a new participant, based on a model that was trained on other participants, will vary greatly from the true attentional state. In consistency with the person-dependent analysis, we classified the pooled dataset for the 13-second epochs (see Section 5.1).
The 95%-confidence interval of the resulting classification accuracies based on the EEG feature vector yielded the range of [0.688, 0.726]. The mean accuracy based on the ET features was only 52.75% (confidence interval: [0.508, 0.547]) and the combined feature vector had slightly worse results than the pure EEG-based classification with a range of [0.683, 0.722]. The difference between the EEG and ET based classification is significant with $t(12)=2.4313$, and $p=.0317$. So is the difference between the ET and the combined feature set with $t(12)=3.2801$, and

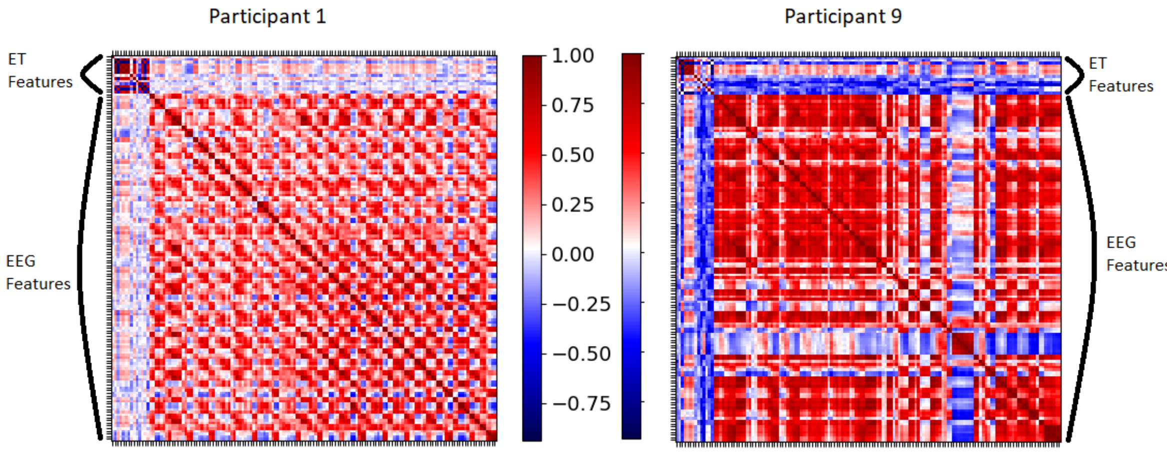


Fig. 5. Feature correlation matrices for participants 1 and 9. Exact feature names were omitted for clarity reasons. The first 14 features are ET features and the following features are EEG features.

p=.0066. The difference between the classification accuracies achieved with the EEG features and the accuracies achieved with the combined feature set is not significant (*t(12)=t = 1.6497, p=.1249*).
The chosen ET features seem to generalize poorly over participants for the differentiation between internal and external attention in augmented reality. For a better understanding of this effect, the feature values for each participant were compared in the next step.

5.1.3 Eye Tracking Feature Exploration. For a detailed examination of the features, each feature was visualized as a boxplot for both conditions across all participants. The participants were ordered from the highest individual ET classification accuracy to the lowest. As an example, Figure 6 shows the boxplot for the number of saccades. Each individual feature was inspected for similarities between participants or common patterns in the means, variations, or outlier range that is represented in the boxplots but none was found. As it can be seen in the example plot the feature "number of saccades" is represented very differently between participants: sometimes, the mean number of saccades is higher for internal attention and sometimes for external attention. The range varies greatly and there is no common effect among participants with a higher accuracy (plotted towards the right) compared to participants with a lower individual accuracy (plotted towards the left) when it comes to the difference between the means. We examined all eye tracking features but from the collected data, there seems to be no characteristic eye gaze behavior that generalizes over participants. A bigger or different feature set with other features might result in better accuracies. We chose the features in this approach based on currents standards from the literature. Finding eye tracking features that generalize better over participants for Augmented Reality tasks is out of the scope of this work and will be the topic of future research. For all further analyses, the ET data will be left out because for now, it does not help to improve person-independent real-time attention classification in Augmented Reality.

5.2 Effect of Classifier Choice

Following the research question of whether a different classifier would improve the classification results for the EEG data in a person-dependent analysis, we tested all four classifiers that were described in Section 4. Again, for each participant, the classifiers were trained and tested 100 times with random train-test-splits with a 1:1 ratio.

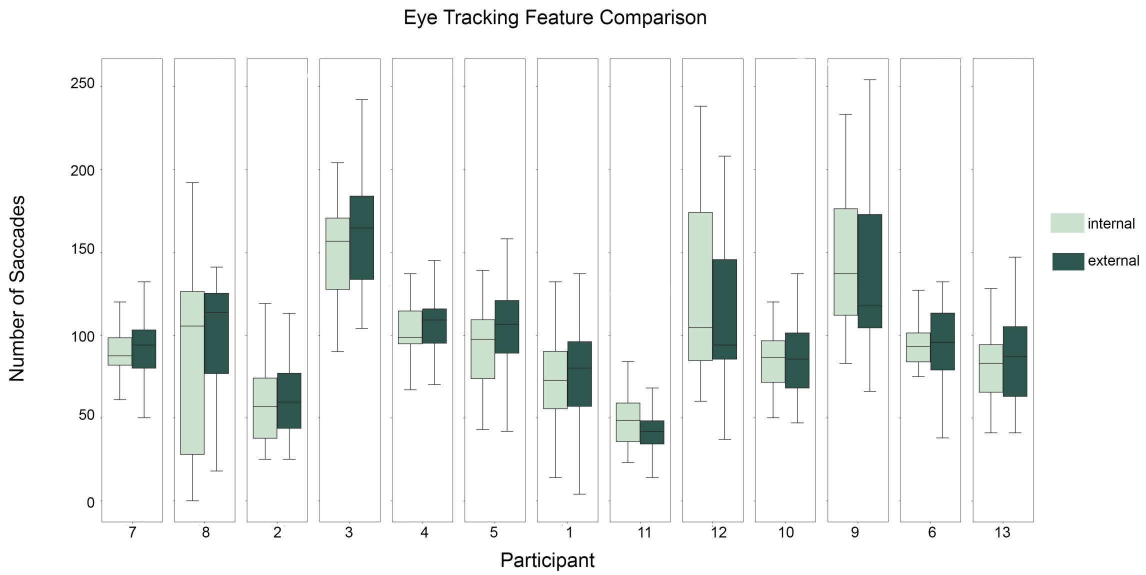


Fig. 6. Participant-dependent boxplot for the number of saccades split into internal and external attention trials. Participants are sorted from the highest classification accuracy on the left to the lowest on the right.

The 95%-confidence interval for the LDA resulted in [0.768, 0.847], for the Random Forest in [0.702, 0.838], for the Vanilla Neural Net in [0.653, 0.696] and for the sFBCSP-NN in [0.754, 0.839]. The results of the Vanilla Neural Net are significantly worse that for the LDA (*t(12)= 3.6130, p=.0036*), the Random Forest (*t(12)= 2.9046, p=.0132*), and the sFBCSP-NN (*t(12)= 3.1516, p=.0084*). The difference in classification accuracy is not significant between the sFBCSP-NN and the LDA (*t(12)=0.6814, p=0.5085*), the sFBCSP-NN and the Random Forest (*t(12)= 0.6687, p=0.5164*), or the Random Forest and the LDA (*t(12)= 0.9701, p=0.3512*). See Table 2 for the exact results per participant.

For the 13-second window in this analysis, the LDA is the best performing classifier on average, however, the difference is only significant compared to the Vanilla Neural Net. For shorter window lengths, the amount of available training and testing data will change and the raw input for the sFBCSP-NN will shorten. We decided to perform the analysis of shorter windowing intervals for the person-dependent analysis using all classifiers in case one profits more than the other from the changes in the datasets.

5.3 Effect of EEG Epoch Length

For an attempt at a real-time classification system, epoch lengths of 13 seconds are not ideal. Basing a label prediction on a shorter window would be a more timely representation of the current attentional state. In this comparison, we want to analyze which epoch length has sufficient data to classify internal and external attention reliably.

As described in Section 3, we cut the data in five different ways to achieve five different window lengths: 13 seconds, 7 seconds, 4 seconds, 2 seconds, and 1 second. All classifiers were trained for each epoch length 100 times and the average per participant was calculated. As an example, the participant-wise results for the LDA can be seen in Figure 7. The averaged results over all participants for each classifier can be seen in Figure 8.

Table 2. Classification accuracies for every participant calculated on 13-second epochs with 50% training data. Mean calculated over 100 random train-test-splits.

Participant	LDA	RF	VNN	sFBCSP-NN
1	0.970	0.944	0.781	0.979
2	0.845	0.917	0.810	0.880
3	0.919	0.833	0.705	0.921
4	0.456	0.639	0.619	0.449
5	0.819	0.639	0.576	0.682
6	0.916	0.694	0.829	0.926
7	0.837	0.778	0.595	0.837
8	0.970	0.972	0.710	0.978
9	0.724	0.611	0.600	0.629
10	0.933	0.972	0.610	0.941
11	0.713	0.639	0.610	0.770
12	0.673	0.75	0.607	0.717
13	0.721	0.722	0.724	0.648
Mean	0.807	0.778	0.645	0.797
Std.	0.141	0.131	0.086	0.157

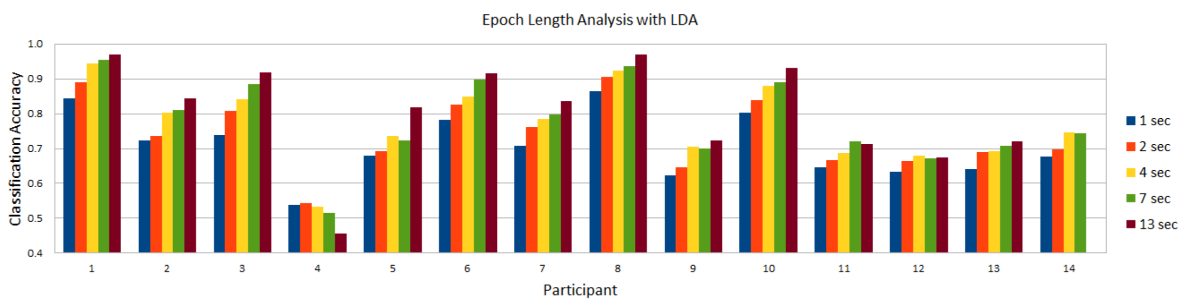


Fig. 7. Mean classification accuracy computed with the LDA on 50% of the data for every participant and different epoch lengths. Means were computed over 100 random train-test-splits.

Participant 14 was again left out of the analysis for 13-second epochs because of the smaller amount of available trials.

Overall, the Random Forest and the Vanilla Neural Net performed worse for all epoch lengths and thus we decided, to not consider them further in this work for the person-independent analysis that follows.

For the LDA, the best performance was achieved using 13-second epochs. The performance decreased with shorter epoch lengths. The calculated confidence intervals for each epoch length are: 13 seconds = [0.768, 0.847], 7 seconds = [0.752, 0.813], 4 seconds = [0.745, 0.799], 2 seconds = [0.715, 0.766], 1 second = [0.685, 0.73]. The difference between 13 and 7 seconds ($t(12)= 2.2494$, $p=.0440$) is statistically significant. Between 7 and 4 seconds the mean differences were not significant with $t(12)= 2.0528$, $p=.0626$. The difference between 4 and 2 seconds ($t(12)= 4.7243$, $p<.001$) and 2 and 1 seconds ($t(12)=6.7575$, $p<.001$) were highly significant.

The neural net performs better than the LDA for 2 seconds, 4 seconds, and 7 seconds. The best result was achieved for the 7 second epoch length with a confidence interval of [0.765, 0.836], followed by 13 seconds

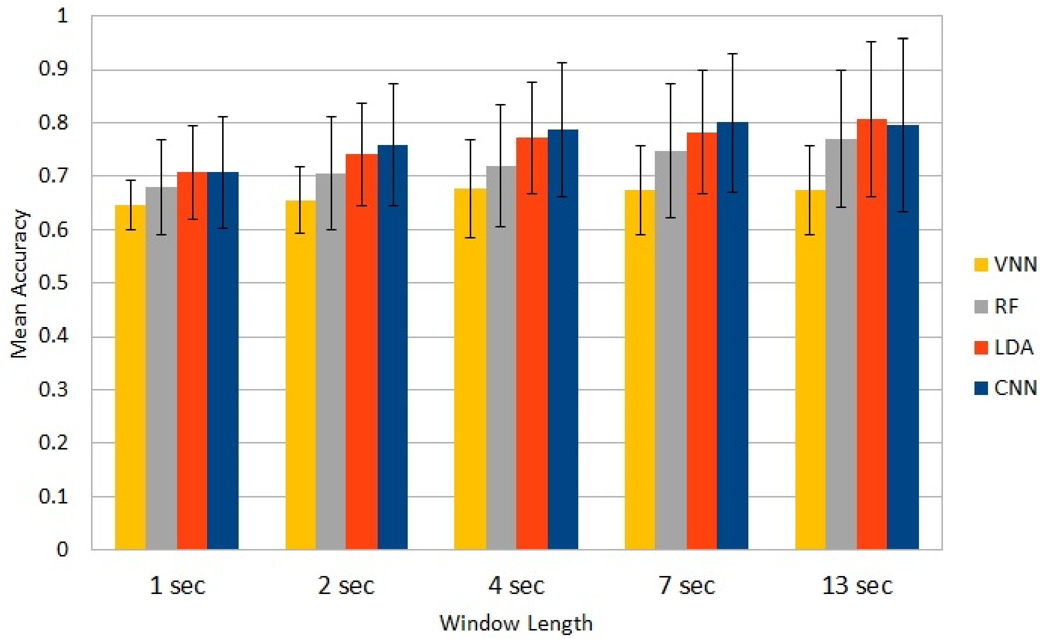


Fig. 8. Mean classification accuracies and Standard Deviation for all classifiers and epoch lengths. VNN = Vanilla Neural Net, LDA = Linear Discriminant Analysis, RF = Random Forest, CNN = sFBCSP-NN

([0.754, 0.893]) and 4 seconds ([0.754, 0.82]). Even so, the difference between 13 and 7 seconds was not significant (*t(13)=1.2547,p=.2317*). The confidence interval for the 2-second epochs is [0.728,0.788]. The sFBCSP-NN and the LDA performed almost equally on the 1-second epochs (NN: [0.68, 0.734]). With *t(13)= 2.3417, p=.0358* the differences between the means for 7 and 4 seconds were significant and the differences between 4 and 2 seconds (*t(13)=4.8948, p<.001*) and 2 and 1 second (*t(13)=5.5806, p<.001*) were highly significant.

The overall best performance was delivered by the LDA for 13-second epochs. However, the accuracy decrease for 7 and 4-second epochs, especially using the NN, is small, compared to the gain the shorter windows would have in a real-time classification system. The payoff for the slightly lower classification accuracy is worth it, considering that attention switches between internal and external attention can happen fast and often. Thus, a 4-second interval will probably result in better accuracy for a less controlled setting and environment because it reacts faster to attentional changes than a classifier, that relies on 13 seconds of data. Therefore, we will primarily consider shorter window lengths for the investigation of real-time classifiers.

By splitting the cleaned EEG data into smaller windows, the number of epochs that are used for training and testing increased. For 7 seconds, the amount of epochs is twice the amount of 13 seconds, while still maintaining the same number of samples. Reducing the window size to 4 seconds increases the number of epochs by a factor of 3 compared to 13-second epochs, for 2-second windows by a factor of 6 and for 1-second windows by a factor of 13. While the duration in seconds that the features for the classifiers are based on does not change, the higher amount of epochs that can be used to train the model might affect the results. For the LDA, the amount of features

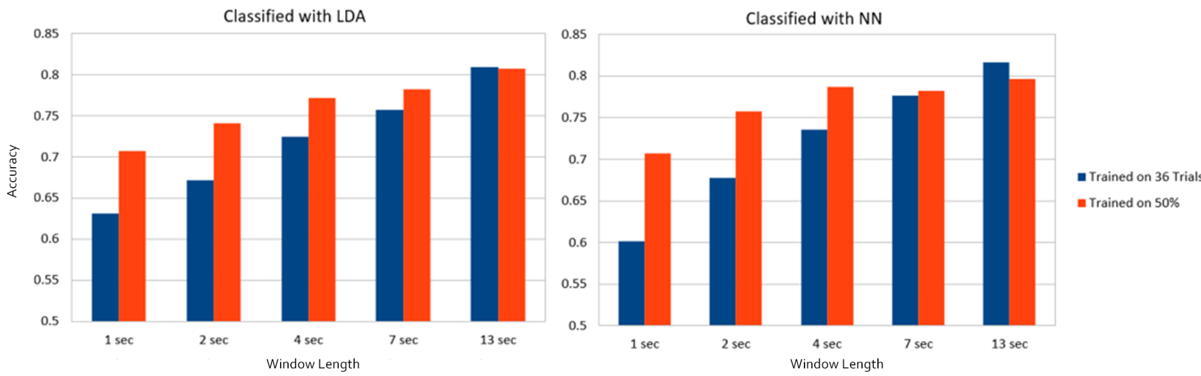


Fig. 9. Comparison of the mean classification accuracies depending on the amount of training data for a person-dependent analysis. The sFBCSP-NN was used as the Neural Net (NN).

in each feature vector stays unaffected. Only the time interval on which the features are generated changes. For the sFBCSP-NN however, the actual input data is less, because the Neural Net is fed with the raw EEG data. For better comparability, we repeated the analysis with the same number of feature vectors for training instead of the same length of the training time. Accordingly, the next step will be to train the classifiers on the same number of epochs, independent of the epoch length.

5.3.1 Effect of EEG Training Epochs Amount. To compensate for the higher amount of feature vectors used for the training with shorter epoch lengths, we repeated the previous analyses with a different train-test split strategy and compared them to the previous results. We limited the number of training epochs to 36 for all epoch lengths (50% of 72 Trials in the 13-second set). Thus, the 13-second epoch analysis did not differ for the two approaches in all subjects but subject 14 because of the smaller overall amount of data. As expected, the mean accuracy decreased for all other epoch lengths, compared to 50% training data. This effect can be seen in Figure 9. The differences between the means are significant for all lengths with both classifiers.
We conclude that the same number of training epochs will lead to worse classification results if the calculated features are based on shorter time windows of EEG data. A longer period to record training data is necessary for more reliable results. As expected, more training data leads to better results.

5.4 Pooled Data Analyses

The pooled data analysis aims at gaining further insight into how well the results generalize over multiple participants. All results in this section are based on pooled datasets and independent of the participants. The EEG data of all participants were treated as coming from one participant and then split into training and test data - independent of information about the participants. Again, each analysis was run 100 times to correct for random effects that are based on the random splitting.
With this method, data of the same participant can be in the test, and in the training data, and therefore, the analysis is not truly person-independent. Nevertheless, the results will help us understand the trade-off between increased training data size and increased variance in the data as well as give us a better idea about the generalizability of the data. If this approach works significantly better than a person-independent classification, we would conclude that a transfer-learning approach would be a good idea.

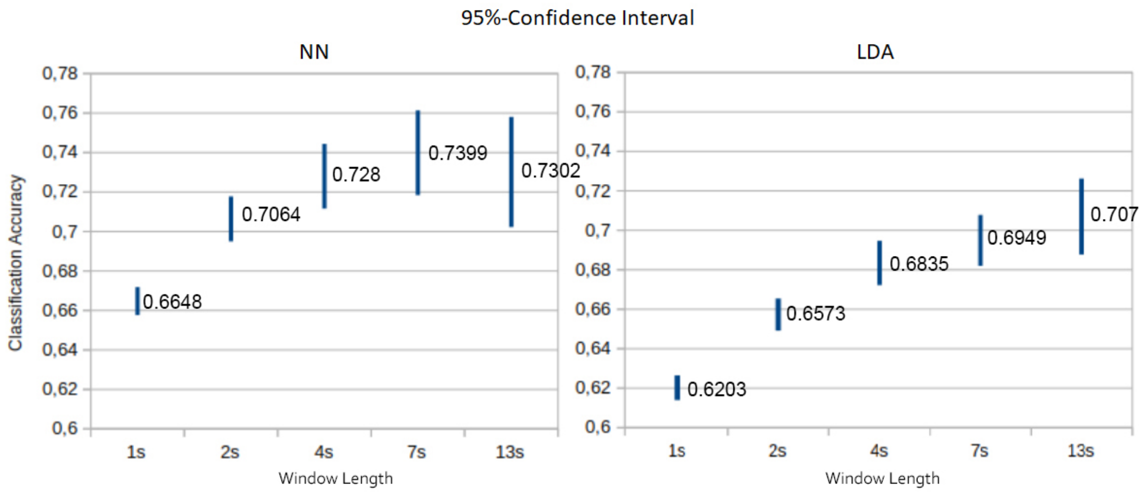


Fig. 10. Analysis of the classification accuracies for the pooled dataset. Each bar visualizes the 95%-confidence interval calculated on 100 runs with random train-test-splits. The number reports the average classification accuracy. The sFBCSP-NN was used as the Neural Net (NN).

5.4.1 Classifier and Window Length. In order to optimize the classification accuracy of the EEG data, we compared all possible combinations of window lengths with the sFBCSP-NN and the LDA. We followed the same procedure as before, training and testing the classifier 100 times to calculate the confidence interval (see Figure 10).

The overall pattern for each classifier was the same for the pooled dataset as for the person-dependent data: the classification accuracy decreased for shorter windows. For the sFBCSP-NN, the 7 seconds epoch length was classified with the highest accuracy of 73.99% and a 95%-confidence interval of between 71.85% and 76.14%. All differences to the other epoch lengths were statistically highly significant, with $p < .001$. Only the difference between the 4-second and the 13-second epochs was not significant with $t(99) = 1.7396, p = .0851$. For the LDA, the 13-second epochs worked best with an average classification accuracy of 70.7% and a 95%-confidence interval between 72.63% and 68.78%. The differences between all epoch lengths for the LDA were significant. Strikingly, the sFBCSP-NN outperformed the LDA for all epoch lengths on the pooled dataset and for pairwise comparison of each epoch length, the difference between the sFBCSP-NN and the LDA were always highly significant. The mean classification accuracies for both classifiers and each epoch length are reported in Figure 10. The bar visualizes the 95%-confidence interval.

Based on these results and the goal to come close to real-time attentional state classification, we decided to perform the real person-independent analysis on 4-second epochs with the neural network even though the 7-second epoch had a better classification accuracy. The trade-off between a slightly smaller classification accuracy but shorter windowing seems appropriate in this context. As argued before, the classification accuracy in a less controlled setting can be expected to be higher for a 4-second interval because it can capture the changes in the attentional state faster and thus classify the real-time attentional state more accurately, than if the classification were based on a 7-second interval.

Table 3. Mean and lower and upper boundaries of the 95%-confidence interval of the person-independent classification accuracy for each participant.

Participant	Lower Border	Upper Border	Mean
1	0.640	0.670	0.655
2	0.755	0.767	0.761
3	0.651	0.671	0.661
4	0.520	0.529	0.525
5	0.501	0.508	0.505
6	0.747	0.763	0.755
7	0.513	0.517	0.515
8	0.603	0.633	0.618
9	0.594	0.611	0.602
10	0.746	0.763	0.754
11	0.413	0.442	0.428
12	0.526	0.556	0.541
13	0.504	0.509	0.506
14	0.500	0.527	0.513

5.5 Person-Independent Analysis

We followed two different approaches to test how well EEG data can be classified person-independently for internal and external attention in AR. In the first approach, we trained the model in a leave-1-out fashion and tested the model on the participant that was left out. In the second approach, we trained the model on a set of selected participants, based on a high classification accuracy in the first approach to check whether "good training data" improves classification accuracies for datasets with a lower accuracy. As mentioned, the classification was performed using the sFBCSP-NN with raw 4-second EEG data windows as input.

5.5.1 Leave-1-out. The leave-1-out procedure was repeated 100 times for each participant. The borders of the individual 95%-confidence intervals are reported in Table 3. The accuracies vary greatly between participants. On average, the classification accuracy was 59.56% with a standard deviation of 10.42%. The overall 95%-confidence interval is [0.541, 0.65] for the classification accuracy if a classifier was never trained on data from the specific person it was tested on. (Reminder: Chance level = 0.5).

5.5.2 Selected Training Set. The seven participants with the highest classification scores from the first approach (all above 60% accuracy) were chosen for the model training in the second approach. The goal was to leave out "bad data" in the training that does not generalize well over the other participants. The datasets from participants 1,2,3,6,8, 9, and 10 were used for training in the leave-1-out approach on their subset. Each of these seven versions of the model was used to classify the participants that were not within the subset. The data of the participants in the subset was classified with the one version of the model that was not trained on their data. The accuracies can be seen in Figure 11. The confidence interval over all participants is [0.538, 0.672]. One participant achieved an accuracy of 88%. There is no general pattern on whether or not the selected classifier training improved the classification accuracy in a person-independent analysis. Also, there is no version with selective training that worked significantly better than the other versions for all participants.

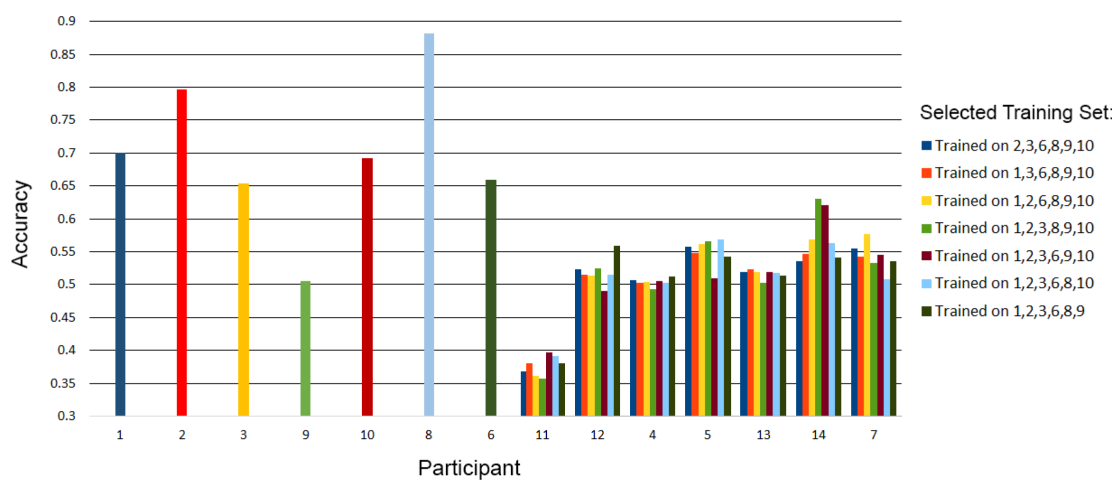


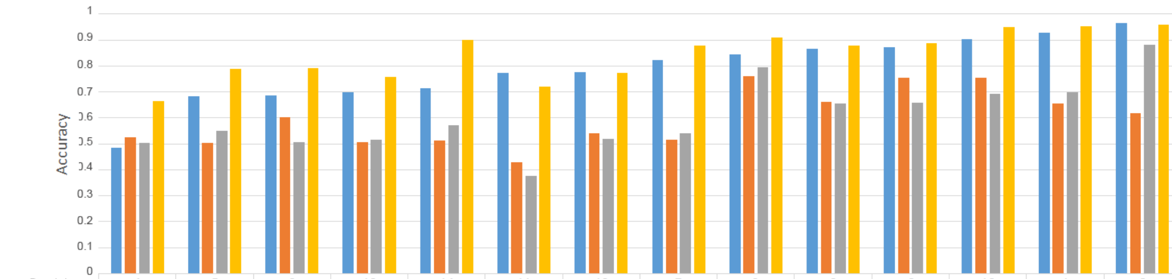
Fig. 11. Classification accuracies of the person-independent testing with a selected subset for the training of the sFBCSP-NN.

5.6 Summary

All in all, the previous results suggest that a real-time person-independent classifier of internal and external attention in AR is possible with a good accuracy. However, there seem to be many dependencies and participant-based differences. The main results of this systematic study are the following:

- Person-dependent:
  - Our eye tracking features lead to worse results than the EEG features
  - EEG and ET classification accuracies are only weakly correlated
  - A combined feature vector for ET and EEG data slightly improved the single modality results but the improvement was not significant compared to only EEG data
  - The sFBCSP-NN and the LDA performed similarly
  - Longer epoch lengths have higher classification accuracies than shorter epoch lengths
  - A higher amount of epochs used for training improves the results for all epoch lengths
- Person-independent:
  - Our chosen ET features do not generalize over participants
  - The sFBCSP-NN is more promising than an LDA with Power Spectral Density-related features
  - 4-second epochs have the best trade-off between classification latency and accuracy
  - A classifier training on a subset of selected participants with good classification results does not significantly improve person-independent classification compared to a classifier trained in a leave-1-out fashion

In Figure 12, the mean accuracy of the 4-second epochs classified by the NN for person-dependent, person-independent, selected training set, and pooled dataset are summarized.



Participant	4	5	9	13	14	11	12	7	2	3	6	10	1	8
PD	0.48	0.68	0.69	0.70	0.71	0.77	0.78	0.82	0.84	0.87	0.87	0.90	0.93	0.97
PI	0.52	0.50	0.60	0.51	0.51	0.43	0.54	0.52	0.76	0.66	0.75	0.75	0.66	0.62
selected set	0.50	0.55	0.51	0.52	0.57	0.38	0.52	0.54	0.80	0.65	0.66	0.69	0.70	0.88
pooled data	0.67	0.79	0.79	0.76	0.90	0.72	0.77	0.88	0.91	0.88	0.89	0.95	0.95	0.96

Fig. 12. Summarized results for the 4-second epochs and the sFBCSP-NN approach for person-dependent (PD), person-independent (PI), selected training set , and pooled dataset.

6 DISCUSSION

In this work, we explored different settings and processing options for internal and external attention detection in augmented reality. The goal is to work towards a real-time, person-independent brain-computer interface in combination with AR. We compared EEG and eye tracking as input modalities, four classification algorithms, and different time window lengths for the training and testing data. The individual differences between participants make general claims and average computations difficult. The person-independent classification worked well and much better than the chance level for half of the participants, where some scores reached as high as 88% classification accuracy. On the other hand, half of the participants' datasets could not be classified reliably by a person-independent classifier with the settings that we chose.

The possible explanations for the results are manifold. First of all, one always needs to consider the possibility that the recorded data was noisy for some participants, and artifacts disturb a good classification process. All of the participants that had lower person-independent scores also had comparatively low person-dependent scores. Possibly, their data quality or setup accuracy is responsible for the decreased classification accuracies. Additionally, the concept of BCI-illiteracy has evolved in this field of research (see Dickhaus et al. [13]). It describes the often observed phenomenon that some users are incompatible with BCIs, even if they are trained person-dependently. This concept also finds its critics (see Thompson [44]) but could be an explanation of why the setting works well for some participants while it fails for others.

Another source of problems might be label noise that results from the nature of attention. In the experiment, the participant was expected to direct it's attention internally or externally for 15 to 20 consecutive seconds. These trials were cut into smaller segments during our analysis. Short fluctuations of the attentional focus would induce a label noise on many segments. These are hard to control for and would require a different experimental setup or labeling mechanism. The different participant results might describe that some people were better at keeping their attentional focus than others, thus, their labeling was more accurate. Also, all claims that we made about the epoch length are dependent on the sampling frequency which was 500 Hz in this study. For other sampling frequencies, the results will differ.

Taking the data classification to the next level by making it person-independent is extremely important for the usability of such a BCI but might face the problem that there are interpersonal differences in EEG power spectra that are genetically dependent (see Smit et al. [43]). Generalizing over features that are based on power spectral densities or common spatial patterns in the brain could, therefore, be an approach with many obstacles.

We compared only two classifiers, a linear approach and a neural network that both work with power spectral features. Appriou et al. [2] explicitly suggest that the best results for mental workload were achieved with the shallow convolutional neural net. On the other hand, Schirrmeister et al. [39] suggested a deep convolutional neural network where the features are not fixed by the architecture. The results of such a network would be harder to understand but this black box might improve classification accuracies. In further studies, other classifiers could be implemented and tested to find further advantages in deep learning approaches.

Interestingly, the EEG data-based classification was not more accurate than the eye tracking-based classification in all participants. Especially for participants with a low EEG based accuracy, the eye tracking prediction worked well in the person-dependent fashion. Eye gaze behavior is highly influenced by the visual task in this paradigm. However, for the internal condition, we did not advise the participant to follow any specific strategy. Unfortunately, no questionnaire data is available on the task-solving approaches of the participants. We assume that during the external condition, the eyes were mainly fixated on the visible ball. During the internal conditions, participants could use different strategies for imagining the movement of the ball, for example, either fixation on the visible tube while performing the movement, or fixating where they imagine the ball to move and following it with their movement. If the gaze patterns during both conditions are very similar concerning fixations, saccades, and blinks, a classification is only possible with low accuracy.
The results of the pooled dataset for the eye tracking data also suggest that the gaze behavior and thus the strategies for solving the internal condition might have varied highly across participants. The classification accuracy for the pooled dataset was barely above chance level.
The conclusion that eye tracking data or gaze behavior is not usable for a person-independent classification of internal and external attention might arise from our study. However, our limited feature set might have been the reason for our results. The methods used to compute the features or the choice of features could have been sub-optimal and led to a classification barely above chance. This is one of the very interesting follow-up questions that arose from this study. Diving deeper into the topic of person-independent classification of eye tracking data and finding characteristic features for internal and external attention would have been out of the scope of this study but will be at the center of future research. We do not want to claim that eye tracking is unsuitable for person-independent attention classification.

Overall, the individual differences between participants make general claims difficult. Due to the fact that person-independent classification is only possible if we have interpersonal similarities, questions about the validity and possibility of the proposed approach arise. Also, considering that 50% of participants could not be reliably classified in the person-independent approach, we have to consider collecting data from more participants. A large and more varied data set might be better for dealing with outliers and improve prediction accuracy considerably.

One question that was not addressed in this study is the necessity and placement of EEG electrodes. As an example, Liu et al. [24] implemented a real-time approach to recognize 8 different emotions with only four electrodes. Fewer electrodes would be favorable for the usability of such a BCI but more electrodes might improve the accuracy. This trade-off is similar to the trade-off between 'real-time' windowing and accuracy. The evaluation of this balance is dependent on the usage context. For research and detailed lab studies, a more complicated setup with slower reaction times of the correct prediction could be endurable, whereas user applications highly depend on comfort and immediateness. It could also be considered whether other brain imaging techniques are suitable for attention detection (i.e. functional near-infrared spectroscopy).

The good results on the pooled dataset suggest that transfer learning is another promising approach that could improve the classification accuracy compared to pure person-independent model training. The results show that personalization of the model of some type might be necessary for reliable results. For example, Saeed et al. [36] propose a self-supervised learning approach that works with 5 labeled instances per class. Another idea would be that a pre-trained person-independent classifier could be retrained on only a few labeled trails of the user. However, a disadvantage would be that training data still need to be explicitly recorded. Whether or not this improvement is worth the effort and inconvenience of training data recordings is in the eye of the beholder.
An additional idea, instead of collecting explicitly labeled training data in a separate training session, would be a pre-trained self-correcting classifier. This option might be available for certain applications. The classifier could be person-independent in the beginning, and during the usage of the application, information about classification mistakes (i.e. error potentials) could be used to improve the classifier by giving feedback about the wrong prediction. While the wrong classification in the beginning would decrease the usability, the collection of the training data would be less obtrusive because it could happen during the normal usage of the application.

Our hope is that the results of this study can be transferred to other tasks. Since we assume that the only difference in the EEG data for both classes should be the attentional direction, we assume that the results hold for other internal/external attention tasks as well. Nevertheless, the task-dependency of the results is yet to be assessed.
In the preliminary proof-of-concept study, we tested whether an internal/external classifier would improve the perception of usability at all. The classifier required training and had low accuracy. Despite this sub-optimal setting, the attention-awareness was already rated as significantly better than an attention-unaware setting. This shows the importance of the performed study.

6.1 Contribution to the Field

The research on the suggested topic contributes to the field of Brain-Computer Interfaces, attention research, and specifically the improvement of Augmented Reality systems.
We are not aware of other studies that explore internal and external attention detection in a person-independent fashion. By suggesting these variations we point out an interesting direction of research that will improve the usability of certain Augmented Reality applications. With the wide scope and many different comparisons in this work, we test the effects of very basic settings. This can be an inspiration for a more detailed and task-independent analysis of some of the main research questions. Additionally, many follow-up questions arose from the results.

6.2 Future Work

As mentioned in the discussion, more data should be collected for more reliable results. Our next step will also include the collection of new data to test the task independence of the results. We will design other AR tasks with internally and externally directed attention and analyze open questions from this study. In the future, we will focus on the exploration of other classification approaches for eye tracking data that are more robust over participants. This will also have to be assessed for different paradigms to test how reliable the results are for different visual tasks. Afterward, the combination of multiple modalities in early, as well as late fusion approaches could be improved. The EEG results suggest that true person-independent classification will continue to be of high difficulty. Instead, we will consider classifier training approaches that use a minimum necessary amount of explicit training phases, such as supervised transfer learning or unsupervised self-correcting classifiers. Once reliable results were found, the re-implementation of a real-time classification paradigm will be the overall goal.

REFERENCES

[1] Sonja Annerer-Walcher, Simon Ceh, Felix Putze, Marvin Kampen, Christof Körner, and Mathias Benedek. 2020. How reliably do eye parameters indicate internal vs. external attentional focus? *PsyArXiv* (2020). https://psyarxiv.com/gxebf/

[2] Aurélien Appriou, Andrzej Cichocki, and Fabien Lotte. 2018. Towards Robust Neuroadaptive HCI: Exploring Modern Machine Learning Methods to Estimate Mental Workload From EEG Signals *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper LBW615, 6 pages. https://doi.org/10.1145/3170427.3188617

[3] Mathias Benedek, Rainer J. Schickel, Emanuel Jauk, Andreas Fink, and Aljoscha C. Neubauer. 2014. Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia* 56, 1 (2014), 393–400. https://doi.org/10.1016/j.neuropsychologia.2014.02.010

[4] Mathias Benedek, Robert Stoiser, Sonja Walcher, and Christof Körner. 2017. Eye behavior associated with internally versus externally directed cognition. *Frontiers in Psychology* (2017). https://doi.org/10.3389/fpsyg.2017.01092

[5] Robert Bixler and Sidney D'Mello. 2016. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction* 26, 1 (2016), 33–68.

[6] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. Attention-based Design of Augmented Reality Interfaces *(CHI EA '05)*. ACM, New York, NY, USA, 1228–1231. https://doi.org/10.1145/1056808.1056883

[7] Anne-Marie Brouwer, Maarten A. Hogervorst, Bob Oudejans, Anthony J. Ries, and Jonathan Touryan. 2017. EEG and Eye Tracking Signatures of Target Encoding during Structured Visual Search. *Frontiers in Human Neuroscience* (2017). https://doi.org/10.3389/fnhum.2017.00264

[8] Anne-Marie Brouwer, Boris Reuderink, Joris Vincent, Marcel A. J. van Gerven, and Jan B. F. van Erp. 2013. Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of Vision* 13, 3 (07 2013), 17–17. https://doi.org/10.1167/13.3.17 arXiv:https://jov.arvojournals.org/arvo/content_public/journal/jov/933551/i1534-7362-13-3-17.pdf

[9] Tad T Brunyé, Trafton Drew, Donald L Weaver, and Joann G Elmore. 2019. A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive research: principles and implications* 4, 1 (2019), 1–16.

[10] Simon Majed Ceh, Sonja Annerer-Walcher, Christof Körner, Christian Rominger, Silvia Erika Kober, Andreas Fink, and Mathias Benedek. 2020. Neurophysiological indicators of internal attention: An electroencephalography–eye-tracking coregistration study. *Brain and behavior* 10, 10 (2020), e01790.

[11] Marvin M. Chun, Julie D. Golomb, and Nicholas B. Turk-Browne. 2011. A Taxonomy of External and Internal Attention. *Annual Review of Psychology* (2011). https://doi.org/10.1146/annurev.psych.093008.100427

[12] Nicholas R. Cooper, Rodney J. Croft, Samuel J.J. Dominey, Adrian P. Burgess, and John H. Gruzelier. 2003. Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International Journal of Psychophysiology* 47, 1 (2003), 65–74. https://doi.org/10.1016/S0167-8760(02)00107-1

[13] Thorsten Dickhaus, Claudia Sannelli, Klaus-Robert Müller, Gabriel Curio, and Benjamin Blankertz. 2009. Predicting BCI performance to study BCI illiteracy. *BMC Neuroscience* 10, 1 (2009), P84.

[14] Ralf Engbert and Reinhold Kliegl. 2003. Microsaccades uncover the orientation of covert attention. *Vision research* 43, 9 (2003), 1035–1045.

[15] Myrthe Faber, Robert Bixler, and Sidney K D'Mello. 2018. An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods* 50, 1 (2018), 134–150.

[16] Siamac Fazli, Cristian Grozea, Marton Danoczy, Benjamin Blankertz, Florin Popescu, and Klaus-Robert Müller. 2009. Subject independent EEG-based BCI decoding. In *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Eds.). Curran Associates, Inc., 513–521. http://papers.nips.cc/paper/3671-subject-independent-eeg-based-bci-decoding.pdf

[17] Siamac Fazli, Florin Popescu, Márton Danóczy, Benjamin Blankertz, Klaus-Robert Müller, and Cristian Grozea. 2009. Subject-independent mental state classification in single trials. *Neural Networks* 22, 9 (2009), 1305 – 1312. https://doi.org/10.1016/j.neunet.2009.06.003 Brain-Machine Interface.

[18] Andrea Finke, Kai Essig, Giuseppe Marchioro, and Helge Ritter. 2016. Toward FRP-based brain-machine interfaces-single-trial classification of fixation-related potentials. *PLoS ONE* (2016). https://doi.org/10.1371/journal.pone.0146848

[19] Dongyu Gong and Jan Theeuwes. 2020. A saliency-specific and dimension-independent mechanism of distractor suppression. *Attention, Perception, & Psychophysics* (2020), 1–16.

[20] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* 7 (2013), 267. https://doi.org/10.3389/fnins.2013.00267

[21] Thalía Harmony, Thalía Fernández, Juan Silva, Jorge Bernal, Lourdes Díaz-Comas, Alfonso Reyes, Erzsébet Marosi, Mario Rodríguez, and Miguel Rodríguez. 1996. EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *International Journal of Psychophysiology* 24, 1 (1996), 161–171. https://doi.org/10.1016/S0167-8760(96)00053-0 New Advances in EEG and cognition.

[22] Stephen Hutt, Jessica Hardey, Robert Bixler, Angela Stewart, Evan Risko, and Sidney K D'Mello. 2017. Gaze-Based Detection of Mind Wandering during Lecture Viewing. *International Educational Data Mining Society* (2017).

[23] Ioanna Katidioti, Jelmer P. Borst, Douwe J. Bierens de Haan, Tamara Pepping, Marieke K. van Vugt, and Niels A. Taatgen. 2016. Interrupted by Your Pupil: An Interruption Management System Based on Pupil Dilation. *International Journal of Human-Computer Interaction* (2016). https://doi.org/10.1080/10447318.2016.1198525

[24] Yisi Liu and Olga Sourina. 2014. *Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm*. Springer Berlin Heidelberg, Berlin, Heidelberg, 199–223. https://doi.org/10.1007/978-3-662-43790-2_11

[25] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. 2018. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. *Journal of Neural Engineering* 15 (02 2018). https://doi.org/10.1088/1741-2552/aab2f2

[26] Weiquan Lu, Been Lirn Henry Duh, and Steven Feiner. 2012. Subtle cueing for visual search in augmented reality. In *ISMAR 2012 - 11th IEEE International Symposium on Mixed and Augmented Reality 2012, Science and Technology Papers*. https://doi.org/10.1109/ISMAR.2012.6402553

[27] Michael D Mrazek, Dawa T Phillips, Michael S Franklin, James M Broadway, and Jonathan W Schooler. 2013. Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in psychology* 4 (2013), 560.

[28] Pallavi Pandey and K. R. Seeja. 2019. Emotional State Recognition with EEG Signals Using Subject Independent Approach. In *Data Science and Big Data Analytics*, Durgesh Kumar Mishra, Xin-She Yang, and Aynur Unal (Eds.). Springer Singapore, Singapore, 117–124.

[29] Pallavi Pandey and K. R. Seeja. 2019. Subject-Independent Emotion Detection from EEG Signals Using Deep Neural Network. In *International Conference on Innovative Computing and Communications*, Siddhartha Bhattacharyya, Aboul Ella Hassanien, Deepak Gupta, Ashish Khanna, and Indrajit Pan (Eds.). Springer Singapore, Singapore, 41–46.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[31] Felix Putze, Jutta Hild, Rainer Kärgel, Christian Herff, Alexander Redmann, Jürgen Beyerer, and Tanja Schultz. 2013. Locating user attention using eye tracking and EEG for spatio-temporal event selection. *Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13* (2013), 129. https://doi.org/10.1145/2449396.2449415

[32] Felix Putze, Maximilian Scherer, and Tanja Schultz. 2016. Starring into the void? *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16* (2016), 1–4. https://doi.org/10.1145/2971485.2971555

[33] Felix Putze, Dennis Weiß, Lisa-Marie Vortmann, and Tanja Schultz. 2019. Augmented Reality Interface for Smart Home Control using SSVEP-BCI and Eye Gaze. 2812–2817. https://doi.org/10.1109/SMC.2019.8914390

[34] William J Ray and Harry W Cole. 1985. EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science* 228, 4700 (1985), 750–752.

[35] Mathieu Rodrigue, Jungah Son, Barry Giesbrecht, Matthew Turk, and Tobias Höllerer. 2015. Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking *(IUI '15)*. ACM, New York, NY, USA, 121–125. https://doi.org/10.1145/2678025.2701382

[36] Aaqib Saeed, Victor Ungureanu, and Beat Gfeller. 2020. Sense and Learn: Self-Supervision for Omnipresent Sensors. *arXiv preprint arXiv:2009.13233* (2020).

[37] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 71–78.

[38] SCCN. 2019. Lab Streaming Layer. https://github.com/sccn/labstreaminglayer. Accessed: 2020-09-15.

[39] Robin Schirrmeister, Jost Springenberg, Lukas Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional Neural Networks in EEG Analysis. *Human Brain Mapping* 38 (08 2017). https://doi.org/10.1002/hbm.23730

[40] Hakim Si-Mohammed, Ferran Argelaguet Sanz, Géry Casiez, Nicolas Roussel, and Anatole Lécuyer. 2017. Brain-Computer Interfaces and Augmented Reality: A State of the Art. In *Graz Brain-Computer Interface Conference*. Graz, Austria. https://doi.org/10.3217/978-3-85125-533-1-82

[41] Hakim Si-Mohammed, Jimmy Petit, Camille Jeunet, Ferran Argelaguet, Fabien Spindler, Andéol Evain, Nicolas Roussel, Géry Casiez, and Anatole Lécuyer. 2018. Towards BCI-based Interfaces for Augmented Reality: Feasibility, Design and Evaluation. *IEEE Transactions on Visualization and Computer Graphics* (2018), 1–1. https://doi.org/10.1109/TVCG.2018.2873737

[42] Jeroen BJ Smeets and Ignace TC Hooge. 2003. Nature of variability in saccades. *Journal of neurophysiology* 90, 1 (2003), 12–20.

[43] Dirk J. A. Smit, Dorret I. Boomsma, Hugo G. Schnack, Hilleke E. Hulshoff Pol, and Eco J. C. de Geus. 2012. Individual Differences in EEG Spectral Power Reflect Genetic Variance in Gray and White Matter Volumes. *Twin Research and Human Genetics* 15, 3 (2012), 384–392. https://doi.org/10.1017/thg.2012.6

[44] Margaret Thompson. 2018. Critiquing the Concept of BCI Illiteracy. *Science and Engineering Ethics* 25 (08 2018). https://doi.org/10.1007/s11948-018-0061-1

[45] Rob HJ Van der Lubbe, Carsten Bundt, and Elger L Abrahamse. 2014. Internal and external spatial attention examined with lateralized EEG power spectra. *Brain research* 1583 (2014), 179–192.

[46] Lisa-Marie Vortmann, Felix Kroll, and Felix Putze. 2019. EEG-Based Classification of Internally- and Externally-Directed Attention in an Augmented Reality Paradigm. *Frontiers in Human Neuroscience* 13 (2019), 348. https://doi.org/10.3389/fnhum.2019.00348

[47] Lisa-Marie Vortmann and Felix Putze. 2020. Attention-Aware Brain Computer Interface to Avoid Distractions in Augmented Reality *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 8. https://doi.org/10.1145/3334480.3382889

[48] Lisa-Marie Vortmann, Moritz Schult, Mathias Benedek, Sonja Annerer-Walcher, and Felix Putze. 2019. Real-Time Multimodal Classification of Internal and External Attention. 1–7. https://doi.org/10.1145/3351529.3360658

[49] Lisa-Marie Vortmann, Moritz Schult, Mathias Benedek, Sonja Walcher, and Felix Putze. 2019. Real-Time Multimodal Classification of Internal and External Attention *(ICMI '19)*. ACM, New York, NY, USA, Article 14, 7 pages. https://doi.org/10.1145/3351529.3360658

[50] Jue Wang, Nan Yan, Hailong Liu, Mingyu Liu, and Changfeng Tai. 2007. Brain-Computer Interfaces Based on Attention and Complex Mental Tasks. *Digital Human Modeling* (2007). https://doi.org/10.1007/978-3-540-73321-8

[51] D. Zhang, L. Yao, K. Chen, and J. Monaghan. 2019. A Convolutional Recurrent Attention Model for Subject-Independent EEG Signal Analysis. *IEEE Signal Processing Letters* 26, 5 (2019), 715–719.