*Multimodal Technologies and Interaction*

MDPI

*Article*

# EEG Correlates of Distractions and Hesitations in Human–Robot Interaction: A LabLinking Pilot Study

**Birte Richter [1,*,†], Felix Putze [2,†], Gabriel Ivucic [2], Mara Brandt [1], Christian Schütze [1], Rafael Reisenhofer [2], Britta Wrede [3] and Tanja Schultz [2]**

[1] Medical Assistance Systems, Medical School OWL, Bielefeld University, 33615 Bielefeld, Germany
[2] Cognitive Systems Lab, University of Bremen, 28359 Bremen, Germany
[3] Software Engineering for Cognitive Robots and Systems, University of Bremen, 28359 Bremen, Germany
* Correspondence: birte.richter@uni-bielefeld.de; Tel.: +49-521-106-67883
† These authors contributed equally to this work.

**Abstract:** In this paper, we investigate the effect of distractions and hesitations as a scaffolding strategy. Recent research points to the potential beneficial effects of a speaker's hesitations on the listeners' comprehension of utterances, although results from studies on this issue indicate that humans do not make strategic use of them. The role of hesitations and their communicative function in human-human interaction is a much-discussed topic in current research. To better understand the underlying cognitive processes, we developed a human–robot interaction (HRI) setup that allows the measurement of the electroencephalogram (EEG) signals of a human participant while interacting with a robot. We thereby address the research question of whether we find effects on single-trial EEG based on the distraction and the corresponding robot's hesitation scaffolding strategy. To carry out the experiments, we leverage our LabLinking method, which enables interdisciplinary joint research between remote labs. This study could not have been conducted without LabLinking, as the two involved labs needed to combine their individual expertise and equipment to achieve the goal together. The results of our study indicate that the EEG correlates in the distracted condition are different from the baseline condition without distractions. Furthermore, we could differentiate the EEG correlates of distraction with and without a hesitation scaffolding strategy. This proof-of-concept study shows that LabLinking makes it possible to conduct collaborative HRI studies in remote laboratories and lays the first foundation for more in-depth research into robotic scaffolding strategies.

**Keywords:** human–robot interaction; LabLinking; electroencephalography; neural correlates of distraction; role of hesitations in spoken communication

check for updates

**Citation:** Richter, B.; Putze, F.; Ivucic, G.; Brandt, M.; Schütze, C.; Reisenhofer, R.; Wrede, B.; Schultz, T. EEG Correlates of Distractions and Hesitations in Human–Robot Interaction: A LabLinking Pilot Study. *Multimodal Technol. Interact.* **2023**, *7*, 37. https://doi.org/10.3390/mti7040037

Academic Editor: Myounghoon Jeon (Philart)

Received: 21 February 2023
Revised: 17 March 2023
Accepted: 23 March 2023
Published: 29 March 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

This research is rooted in the vision that robots will adequately support humans in their everyday tasks and that they will continuously learn and adapt to human needs over time. Engaging in joint action to achieve a task requires a common understanding of the ongoing interaction [1,2]. This can be established through grounding [3] or alignment processes [4], which are typically established by verbal and non-verbal communication. Allwood and colleagues [5] identified four basic requirements for human communication: (1) willingness to communicate, (2) willingness to perceive, (3) ability to understand, and (4) ability to react attitudinally or behaviorally. The willingness to communicate and to perceive has been investigated and modeled in the context of HRI in terms of engagement and joint attention [6–8], which is mostly measured based on gazing behavior. Problems of communication generally arise in one of these areas and need to be identified and remedied. In order to achieve a shared understanding between robots and humans, it is, therefore, necessary for the robot to monitor the state of the interaction partner regarding these four

levels and to provide appropriate scaffolding strategies in the event of problems such as inattentiveness or non-understanding.

Here, we focus on hesitations as a specific scaffolding strategy—a measure to support learners in acquiring new knowledge and skills. Our previous investigations have focused on the benefits of hesitations based on behavioral data such as correctly memorized information or correctly oriented gaze [9,10]. However, the previous approach focused on interaction results but did not shed light on the processes that actually lead to better memorization.

The role of hesitations in human-human interaction is a much-discussed topic in current research [11,12], and there is evidence for the hypothesis that they have a communicative function and influence the interaction partner. As hesitations occur more often prior to long utterances [13] and infrequent words [14], the frequency distribution of hesitations in spontaneous speech is assumed to be caused by the level of cognitive load of the speaker. The strategic use of hesitations as a scaffolding strategy for the listener has also been discussed by some authors [15]. Yet, results from studies on this issue indicate that humans do not make strategic use of hesitations [11], although several studies point in the direction of the beneficial effects on the listeners' comprehension of utterances [16,17].

To better understand how scaffolding strategies, such as hesitations, affect the cognitive processes of a human interaction partner and to work toward an adaptive robot behavior that takes the attentional state of its user into consideration, we developed an HRI setup that allows the measure of EEG signals of a human participant while interacting with a robot. The proper measurement of high-dimensional EEG biosignals from humans in interaction and the development of a robot system capable of interacting with people in a contingent way are both complex approaches that require dedicated hard- and software as well as specific expertise and knowledge. Both together are not commonly found in a single lab.

However, in today's digitally connected world, the close collaboration across disciplines, distances, languages, and cultures have become the rule rather than the exception. For example, in research and development, tightly interconnected interdisciplinary groups benefit from each other's diverse experiences and perspectives to jointly create innovations. With the technological invention of *LabLinking*, we take the concept of close collaboration to the next level. LabLinking is a technology-based interconnection of experimental laboratories with a defined level of connection tightness (LabLinking Level—LLL) [18]. We argue that linked labs provide a unique platform for a continuous exchange between scientists and experimenters, thereby enabling a time-synchronous execution of experiments performed with and by decentralized users and researchers, improving the outreach and ease of human participant recruitment, allowing the establishment of new experimental designs jointly and to incorporate a panoply of complementary biosensors, devices, hard- and software solutions to capture human behavior [18]. Furthermore, LabLinking supports the increasing demand for sustainability and hybrid events in the post-COVID-19 world.

The following study would not have been conducted without LabLinking, since it builds on the complementary expertise and equipment of two laboratories: the Medical Assistance Systems Group (MAS) at Bielefeld University with its rich expertise in social robotics based on robots such as Pepper, Nao, or Flobi [19–21], and the Cognitive Systems Lab (CSL) at University of Bremen with vast experience in biosignal-adaptive cognitive systems [22] based on multimodal biosignal acquisition [23] and processing using machine learning methods [24], including the recording and interpretation of spoken communication [25] and high-density EEG in the context of intelligent robots and systems [26].

## 2. State of the Art

The use of EEG as a method to measure the impact of distractions and hesitations on the user is motivated by related work, in which some researchers found an effect of speaker's hesitations on the listener's EEG during listening to continuous speech. Corley et al. [27] showed that event-related potentials (ERPs) associated with the meaningful

processing of language are affected by a preceding hesitation. In their experiment, the N400 effect (measuring difficulties in the integration of a word into its linguistic context) was found as unpredictable in contrast to predictable words. They compared fluent to disfluent utterances and found that the N400 effect was reduced by a hesitation before the unpredictable word, indicating that linguistic integration difficulties were reduced. In addition, a memory test indicated that words preceded by a disfluency were more likely to be remembered [27].

Collard [16] showed that ERPs associated with attention (mismatch negativity (MMN) and P300 effect) are affected by a preceding hesitation vowel in a similar experiment. Infrequently occurring, acoustically manipulated target words resulted in typical MMN and P300 components compared to a non-manipulated baseline. Furthermore, a prolonged pause between the hesitation vowel and continuing speech appeared to impair covert attention to the post-disfluent content and the subsequent memory performance for this content. This indicates an immediate effect of hesitations on the listener's overt attention to the upcoming speech [16].

For a robot to respond to lapses in attention, e.g., through a hesitation strategy, requires it to detect the lapse of attention. In many situations, this is caused by external, exogenous shifts of attention [28], which we refer to as distractions. Distractions can be detected from EEG signals based on a few seconds of data by applying machine learning techniques that are commonly used in Brain–Computer Interfaces (BCIs). While average-case analysis of EEG, as discussed above, can often identify subtle effects from ERP analysis, such single-trial classification in interactive scenarios more often relies on frequency-based features, which are not as susceptible to small latency shifts as ERPs. Vortmann and Putze [29] showed, for example, how such a detection (in their case of visual distractions) can be incorporated beneficially into an interactive system to adapt its behavior to the attentional state of the user. In the field of driver safety, multiple studies have investigated the detection of auditory distractions via EEG-based BCIs, for example, by Beltrán et al. [30] and Salous et al. [31]. Another field in which the detection of distractions has been applied successfully is the medical field, which requires long periods of sustained attention, for example, during rehabilitation exercises [32].

These results indicate that hesitations may serve as a good scaffolding strategy that, on the one hand, provides time for processing while, on the other hand, guides the attention to the upcoming difficult part when explaining new information to an interaction partner. They also show that the EEG signal likely contains relevant information that could help us to identify the need for such guidance and measure the impact of hesitations on neural processing.

However, it is still unclear if this positive effect of hesitations can also be used as a communicative tool in human-agent interaction. In our research, we could already show that hesitations can be used as a non-intrusive intervention strategy for dealing with inattentive interaction partners [33]. In our smart-home setting, a robotic virtual agent used hesitations in an explanation scenario whenever it lost the visual attention of the explainee. We could already show that in short interactions, without a change in the discourse, unfilled pauses based on missing mutual gaze have a positive effect on the gazing behavior, i.e., the visual focus of attention, of the interlocutor [34]. In further studies, we could show that such a hesitation intervention strategy could also lead to higher task performance of the human at the cost of less positive subjective ratings regarding the artificial agent [33,35].

Using (1) mutual gaze and task-related features to detect inattentiveness due to missing engagement or difficulties in understanding and (2) different strategies to deal with these improve the task performance without negative side effects on the interaction [33]. However, besides these positive effects of the explainer's hesitations on the task performance of the explainee, it is still an open question if they also affect the EEG responses of the listener and, therefore, provide further information about the understanding process.

This study investigates EEG responses to distractions and hesitations in human–robot interaction. It integrates previous findings from the MAS Lab with the work of CSL

and colleagues within the DFG CRC 1320 Everyday Activity Science and Engineering (EASE), where we provide unique and critical contextual background for robots based on the recording, processing, modeling, and interpretation of human activities, perceptions, and feedback [26]. For this purpose, biosignals resulting from the activity of the brain, the muscles, and the eyes, which are correlated to motion, communication, and other mental tasks, are recorded and interpreted to provide insight into diverse aspects of human behavior that enable them to masterfully perform everyday activities with little effort or attention [24]. Furthermore, these findings are brought into the context of AI explanations, as investigated in the DFG TRR318 Constructing Explainability.

The LabLinking concept, as outlined above, of course, draws inspiration from related work. For example, a development in recent years has been the distribution of standardized, well-documented experiments across multiple labs to ensure the reproducibility of the reported results. Lücking et al. [36] is an example from the robotics domain, while the EEGManyLabs initiative by Pavlov et al. [37] is a similar endeavor from the neuroscience perspective. This approach shares similarities with LabLinking in that it conducts experiments in multiple labs and requires a level of formalized documentation to enable the replication of experiments. A key difference is that for reproducibility efforts, every lab on its own needs to be able to perform the experiment independently and asynchronously. A similar role in the research landscape is played by multicentre studies (e.g., [38]), which also focus on the harmonization of experiment protocols, but with the goal of creating a uniform data set. An alternative approach, which considers real-time interaction between robots and EEG setups at different sites, is teleoperation. Several examples [39–41] show that EEG-based Brain–Computer Interfaces can be used to control robots across a distance, with real-time transmission of the detected control signals. In contrast to LabLinking, the focus in teleoperation is on establishing a tight control loop between the EEG user and robot, while LabLinking supports a wide range of other scenarios (e.g., verbal HRI) and basic research (analyzing the EEG signal instead of using it for control purposes).

## 3. Materials and Methods

### *3.1. Scenario*

We investigate the impact of a robot's hesitations to distractions on the EEG signals of a human listener in the context of an explaining scenario. This involves everyday actions, such as laying out dishes or building blocks on a table with unusual configurations. To investigate how cognitive processes are affected by scaffolding signals, we address situations where humans receive instructions and explanations from a robot. Our envisioned scenario corresponds to a robot scaffolding a human partner who is impaired, for example, by dementia and needs support to carry out sequential everyday tasks. To avoid the influence of prior knowledge on the processing of the instructions, we used a scenario consisting of fictitious new rules for setting a table with standard dishes and cutlery and a more abstract scenario with building blocks. Furthermore, we introduce a distraction signal to impair attention and understanding processes at predefined points in time during the interaction. This will allow us to measure brain signal responses to the scaffolding effects of hesitations.

### *3.2. Research Questions*

We address the following research questions in our HRI scenario:

- EEG responses to scaffolding in HRI
  - **SRQ1:** Can we find effects on single-trial EEG based on the distraction or non-understanding in HRI?
  - **SRQ2:** Can we find effects on single-trial EEG of scaffolding strategies in HRI?
- LabLinking Evaluation
  - **MRQ1:** Does LabLinking enable interdisciplinary joint research approaches between multiple labs?
  - **MRQ2:** What are the best practices to achieve benefits through LabLinking?

### *3.3. LabLinking Method*

Combining a social robot and a high-density EEG setup within the same experiment poses a significant challenge for the proposed research project. Both components are expensive and require extensive expertise to be properly operated. Performing such an experiment often requires two or more groups to combine their unique technical equipment and their respective expertise. However, collaborations like this are difficult to realize because it is often not feasible to move delicate and unique equipment between two spatially distributed labs or to keep important items from one lab at a different location for an extended period of time.

To overcome this issue, we employ the LabLinking paradigm [18] for our experiment, depicted in Figure 1. In this paradigm for experimental research, it is possible for the robot Pepper to remain at its usual location at Bielefeld University while interacting in real-time with a human participant at the BioSignals Lab at the University of Bremen. To realize this setup, we implemented a technical infrastructure with the following main capabilities: (1) streaming of audio, video, and other data for human–robot interaction, (2) synchronized recording of multimodal data streams, and (3) control of experiment flow in a multi-site experiment.

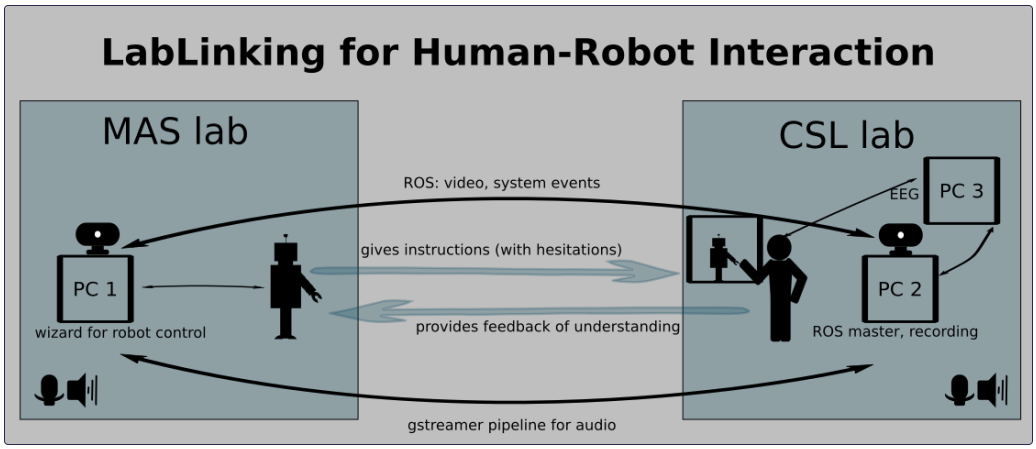


**Figure 1.** The LabLinking setup for conducting HRI studies between the laboratory of the Medical Assistance System Group (MAS) at Bielefeld University and the Cognitive Systems Lab (CSL) at the University of Bremen: The Pepper robot gives instructions to the human interaction partner, which provides feedback of understanding. The laboratories are connected via a GStreamer pipeline for audio and additional communication via ROS for video and further system events.

For the cross-site communication of events, we used the Robot Operating System (ROS) [42], which was also employed to control the robot Pepper at the Bielefeld lab. ROS provides a flexible messaging interface that allows us to establish a consistent and robust data flow between multiple machines across the two sites. A graphical user interface at each lab allowed the respective experimenters to communicate the state of the experiment (e.g., whether a trial was completed successfully) to the other side. Furthermore, all events were logged in ROS bags for later analysis of the temporal structure of the experiment (e.g., to identify trial beginnings). All involved machines at both sites were synchronized to the same NTP server to ensure a reliable alignment of timestamps. For reproducibility, the code for controlling the robot and parts of the lab linking can be installed using a distribution in the cognitive interaction toolkit (CITK) [43].

#### 3.3.1. Streaming Study-Relevant Data

For streaming video data, we used OpenCV-based plugins in ROS (video_stream_opencv, image_view) [44]. For streaming audio data, we used the GStreamer software [45], which supports highly configurable, low-latency streaming pipelines. A video stream capturing visual components from the robot in Bielefeld was streamed to Bremen, while video and

audio components of the participant and a video of the table were streamed back to Bielefeld for the operators of the robot. Furthermore, we implemented a custom GStreamer plugin to store accurate timestamps of the beginning and end of audio recordings. All video frames and other event data were assigned timestamps within the ROS framework. This allowed us to precisely align all collected data types and modalities during analysis. The EEG recorder used a different middleware (Lab Streaming Layer), for which we implemented a custom bridge component to convert the respective data packages into ROS messages.

Instead of streaming the generated voice output of Pepper directly, we used an array microphone to record the voice output of Pepper and streamed this recording from Bielefeld to Bremen. This produced an acoustic setting reminiscent of a real video meeting, as opposed to a cleanly generated, text-to-speech audio experience.

#### 3.3.2. Synchronized Recording

Pepper itself was connected to ROS via a customized version of the C++ naoqi ROS driver [46], as the original implementation does not support timestamps. We added timestamps at the beginning and end of speech output to be able to synchronize the different modalities used at the two sites of the LabLinking. This way, the integrated animated speech functionality of Pepper, which analyses the text and tries to produce contextually appropriate movements of Pepper, could also be added and used. A fixed offset yielding from a delay of audio recording in the two locations was calculated, along with a transmission delay of the spoken sentences from Bielefeld to Bremen.

#### 3.3.3. Experiment Control Flow

To maintain a robust experiment flow in a multi-site experiment, we formalized the steps of one experiment trial into a state machine that was implemented by two different applications, which communicated between the sites. Through a graphical user interface, the experimenters could trigger steps of the experiment or mark them as completed. Marking a step as completed would then notify the other site and unlock the following stages of the experiment. This procedure ensured that experimenters at both sites had a matching understanding of the state of the experiment, avoiding premature or redundant steps. Besides this formalized interaction, a video conference channel was kept open during the experiment to coordinate in the case of unforeseen events.

### *3.4. Experimental Setup*

To collect data on the neural responses to distractions in HRI as well as potential strategies to remedy such distractions, we conducted a LabLinking experiment.

#### 3.4.1. Participants

Participants were recruited via postings on online message boards and through paper flyers. On average, participants were 22.92 years old (SD = 3.82). Four participants identified as female, and eight participants identified as male. Fluent German language skills were an inclusion criterion for participant selection. As compensation, participants received 30 EUR after completing the experiment (participants coming through a dedicated course on empirical human-computer interaction methods also received partial class credit). Each participant was informed about the procedure of the data collection and signed a consent form.

#### 3.4.2. Setup

Participants were seated in front of an empty table at the CSL lab, with the EEG recording equipment on a smaller table on the left side and an assortment of items to place on the table on the right side. In front of the participant (behind the table) was the projection of the live video stream of the Pepper robot. The image was placed and scaled in a way that participants had the impression of Pepper standing behind the table. A microphone

hanging from the ceiling recorded an audio stream from the CSL lab, and two cameras captured a top-view of the table as well as a frontal recording of the participant.

#### 3.4.3. Instructions

At the beginning of the experiment (after all sensors were set up, see next section), participants received instruction about the interaction with the robot and the task procedure and then went through one trial block to clear up any misunderstandings, to familiarize the, with the voice of the robot and the style of instructions. In the main experiment, participants went through a number of blocks. In each block, they were asked to set up the table in a unique way, as instructed by the robot. Each block used items from one of two item sets: One set (KIT) included typical everyday kitchen objects, such as cups, plates, cutlery, or napkins. The other set (BLC) contained colored wooden blocks of different shapes. For each block, we created a custom sequence of instructions. Each instruction asked the participant to place one item in relation to one of the previously placed items or the table (e.g., "place the green cube left of the blue cylinder"). In the middle of the experiment, participants took a 15-min break. Instructions (especially in the case of the kitchen items) were purposefully designed to *not* resemble a realistic table layout to minimize the influence of assumptions on likely item locations. The duration of speech content in each instruction varied between 3 and 5 s. We piloted and adjusted all instructions to ensure that all words were intelligible in the synthesized voice.

#### 3.4.4. Distractions

Pepper's voice came from the front (where the robot was displayed), while distractions were played from the front-left, back-left, front-right and back-right locations. Distractor sounds were sampled from freely available radio and television documentaries with the keyword "robot" in their title. The sampled segments usually contained a mixture of narration, interviews, music, and sound effects. For each block, distractors were taken continuously from a single recording to give a sense of an ongoing conversation. For instructions without distraction, we played ambient sound of road noise from the same directions as the distractors. This was performed to avoid the main difference between distracted and not distracted conditions to be the presence or absence of source separation in the brain. The volume level of distractors and ambient sound was kept constant during all experiments.

#### 3.4.5. Hesitations

Pepper's hesitations were predefined and generated during the normal synthesis process with Pepper's *ALAnimatedSpeech* interface. As a hesitation strategy, we adapted the hesitation strategy for synthetic speech proposed by Betz et al. [47], which has already been tested in an HRI scenario [10]. To make sure the hesitations are recognized as hesitations and not as a normal break, we decided to use an additional silent pause of 1500 ms before the filler and 1000 ms after it. However, the actual total pause in Pepper's speech synthesis was about 3.5 s before the filler and 2.5 s after it (this included Pepper's normal speech pauses, synthesis processing delays, and delay in feedback on the current status of the end of synthesis). Pepper kept gesturing during the breaks. The German filler word "ähm" was played back 50% slower and with a pitch of 80% of the normal voice. In addition, the word before the first silent pause was reduced in speed rate (50%) as well to lengthen the word as an initiation of the hesitation.

#### 3.4.6. Behavioral Data

After each instruction, participants replied to the robot by indicating that they "understood" the instruction, that they "not understood" and needed the robot to repeat the instruction, or that they were "uncertain" about the instruction (but would still try to execute it). Following a potential repetition of the instruction (if requested), the participants executed it by picking up one item and placing it. After the execution of one instruction,

the resulting table was checked by the experimenter at the CSL for correctness by comparing the table layout (as seen through the overhead camera) to a reference picture for each step. In case of a deviation, this trial was marked, and the table setup was manually corrected by the experimenter to the expected position to avoid conflicts with subsequent instructions. Due to the deliberate ambiguity of some instructions, we did not correct every mistake, only those which were not compatible with the given instruction.

#### 3.4.7. Conditions

Within each block, we derived three conditions from the combination of two factors (cf. Table 1): (i) *distraction* (present/absent) and (ii) *hesitation* (employed/not employed). From the four different combinations, we excluded the combination of absent distraction and employed hesitation, as this does not reproduce the expected robot behavior and removing it still allowed us to study the most relevant comparisons while dedicating more trials to the remaining three combinations. We call these combinations NoDist (distraction absent), DistNoHes (distraction present, hesitation not employed), and DistHes (distraction present, hesitation employed).

**Table 1.** Three conditions with a number of utterances utilized in the experiment (DIS: distraction, HES: hesitation, KIT: kitchen item, BLC: wooden block item).

| | **NO DIS** | **DIS** | **Sum** |
|---|---|---|---|
| NO HES | 15 KIT<br>15 BLC<br>→ NoDist | 15 KIT<br>15 BLC<br>→ DistNoHes | 30 KIT without HES<br>30 BLC without HES |
| HES | | 15 KIT<br>15 BLC<br>→ DistHes | 15 KIT with HES<br>15 BLC with HES |

The participants performed two sets of interaction—one with the kitchen objects and one with the wooden blocks. Each set consisted of five interaction blocks, and each block consisted of nine instructions. In total, each participant has thus carried out 90 instructions overall (30 per condition, cf. Table 2). In order for Pepper to behave consistently over a certain period of time, it reacted either with or without hesitations per set (kitchen objects (KIT), wooden blocks (BLC)) and changed its behavior for the second set. This resulted in four different interaction sequences ((1). KIT with hesitations, BLC without hesitations; (2). KIT without hesitations, BLC with hesitations; (3). BLC with hesitations, KIT without hesitations; (4). BLC without hesitations, KIT with hesitations). The participants are randomly assigned to one of these four scripts to balance between the order of the appearance of hesitations, and the order of the presented set and reduce interaction effects between them.

Figures 2 and 3 show the experimental setup.

**Table 2.** Four different scripts to mitigate order effects. Each participant was assigned one script with 90 utterances (DIS: distraction, HES: hesitation, KIT: kitchen item, BLC: wooden block item).

| | | | |
|---|---|---|---|
| Script 1: | 30 KIT DistHes + 15 KIT NoDist | 30 BLC DistNoHes + 15 NoDist | 90 |
| Script 2: | 30 KIT DistNoHes + 15 NoDist | 30 BLC DistHes + 15 NoDist | 90 |
| Script 3: | 30 BLC DistHes + 15 NoDist | 30 KIT DistNoHes + 15 NoDist | 90 |
| Script 4: | 30 BLC DistNoHes + 15 NoDist | 30 KIT DistHes + 15 NoDist | 90 |

**Figure 2.** Setup from the MAS lab perspective with the Pepper robot.

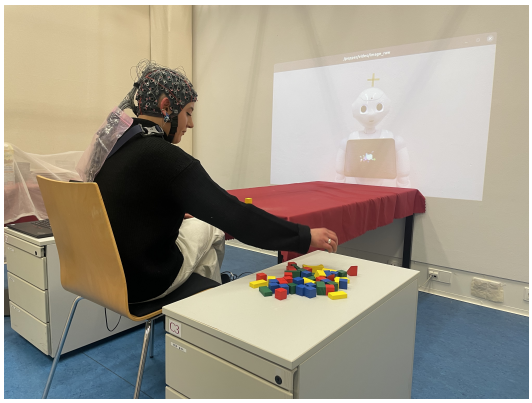

**Figure 3.** Setup from the CSL perspective.

#### 3.4.8. Questionnaire

After the interaction, the participants filled out a questionnaire to gain further insights. The questionnaire consisted of six parts, including questions regarding (i) general demographics, (ii) the self-reported distractibility of the participants, (iii) the perception of the distractions, (iv) the intelligibility of the instructions, (v) the perception of the hesitations, and (vi) the synthesis quality.

### *3.5. EEG Processing and Classification*

In traditional, strongly controlled experiment setups, we would analyze the EEG for event-related effects, such as Event-Related Potentials in the time domain or Event-Related Spectral Perturbations in the frequency domain. However, the uncontrolled nature of our approach (which we chose deliberately to study a realistic HRI scenario) makes it difficult to align events exactly, as onset, content, and acoustic properties of speech and distractors varied from trial to trial. A machine learning-based approach is more flexible in capturing these differences and also prepares us to eventually support the real-time adaptation of the robot.

#### 3.5.1. Preprocessing

Throughout the 10 blocks, continuous EEG recordings were taken with a sample rate of 512 Hz and 64-channel EEG using a g.HIAMP 256 Biosignal Amplifier(g.tec). Two participants were excluded from the analysis due to technical issues in EEG recording. To obtain the EEG data during the participants' listening to Pepper, the audio recorded in Bielefeld was utilized as a precise reference for identifying speech onsets and offsets, corresponding to each of the nine instructions in the block. The fixed recording delay between the audio file in Bremen and the start of the EEG recording was subtracted, as detailed in Section 3.3.2. Furthermore, to account for the transmission delay, a correction

was added to the onsets and offsets, calculated as the average time lag across each block based on the cross-correlation between the clear Pepper audio and the mixed Bremen audio. This yielded 30 trials of EEG data for each of the three conditions with a duration of 3–5 s for normal instructions and distracted instructions and 8–11 s for trials with hesitations. The EEG data were first rereferenced by the average of the two reference electrodes on the left and right earlobes, which were then excluded for further analysis, reducing the number of electrodes from 64 to 62. Subsequently, the data were bandpass filtered between 1 and 32 Hz using an FIR filter from the MNE python library [48], designed as a one-pass, zero-phase, non-causal bandpass filter, followed by a downsampling of the signal to 64 Hz. The filter was created using the firwin method and a Hamming window, with a passband ripple of 0.0194 and a stopband attenuation of 53 dB. The lower passband edge was set to 1.00 Hz with a transition bandwidth of 1.00 Hz and a lower $-6$ dB cutoff frequency of 0.50 Hz. The upper passband edge was set to 32.00 Hz with a transition bandwidth of 8.00 Hz and an upper $-6$ dB cutoff frequency of 36.00 Hz. The filter length was 1691 samples, equivalent to 3.303 s. To allow for consistent comparisons across the conditions and varying trial lengths, the EEG data of each trial were segmented into 1 s decision windows with a 0.5 s overlap. For the hesitation condition, only the EEG data following the hesitation phase was taken into consideration to allow for a direct comparison between the cognitive state of the participants while listening to Pepper with distracting background speech and for the same condition, after their attention was redirected to Pepper. The pre-hesitation audio and silent phases were excluded from further analysis. This procedure resulted in approximately 150 windows per condition. For each of the 1 s windows, spectral power features were computed in 2 Hz bins for each channel using the Welch method for spectral power density calculation, yielding a $62 \times N$ dimensional feature vector, where N is determined by the number of binned features (which was varied for different setups).

#### 3.5.2. Classification

In accordance with research questions SRQ1 and SRQ2, we evaluated the impact of auditory distractions on cognitive processing by comparing the control condition (NoDist) to condition DistNoHes and the distraction condition to condition DistHes. Our goal was to assess the quantifiable effects of these distractions on cognitive processing and to determine if the robot's hesitation strategy leads to any significant changes in the EEG signal, which could indicate a shift in the participant's perception of the robot. The discrimination between NoDist and DistHes was omitted for this analysis, as potential findings could not be attributed to either manipulation, namely distraction and hesitation as an intervention.

To answer our research question, we utilized a Random Forest model from the scikit-learn library [49] to discriminate the EEG windows for the two comparisons. Through a shallow grid search, we obtained optimized results for all participants with a parameter setting of 500 trees, a maximum tree depth of 10, a minimum sample split of 3, and a minimum of 4 samples per leaf. Further, a combination of all features from the delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), low beta (12–20 Hz), and high beta (20–30 Hz) bands binned in 2 Hz led to the optimal performance for 4–12 Hz in the first classification task (ambient vs. distraction) and for 4–20 Hz in the second classification (distraction vs. hesitation), and $62 \times 5 = 310$ and $62 \times 9 = 558$ dimensional feature vectors, respectively.

## 4. Results

### *4.1. EEG*

The classification between the undistracted baseline condition and the distraction condition is depicted in Figure 4. The Random Forest model was trained and tested in a person-dependent manner using a stratified ten-fold cross-validation while making sure that no overlapping window appeared in the training and test sets and was run 20 times and averaged for each participant to account for random factors.

Using a combination of the theta and alpha band (4–12 Hz) yielded the best results for the SRQ1 (NoDist vs. DistNoHes). The average accuracy across all participants was

approximately 60% (STD 5%). A two-tailed t-test comparing our results to the baseline accuracy of 50% revealed a statistically significant difference for all participants ($t(11) = 2.719$, $p < 0.05$). The effect size, as measured by Cohen's d, was 0.43, indicating a medium effect. Using a combination of the theta, alpha, and low-beta band (4–20 Hz) yielded the best results for the SRQ2 (DISTNOHES vs. DISTHES). The average accuracy across all participants approximately reaches 73% (STD, 10%), significantly outperforming the baseline for each participant ($t(11) = 7.757$, $p < 0.001$) with robust classification results for more than half of the participants. The effect size, as measured by Cohen's d, was 1.7, indicating a large effect.

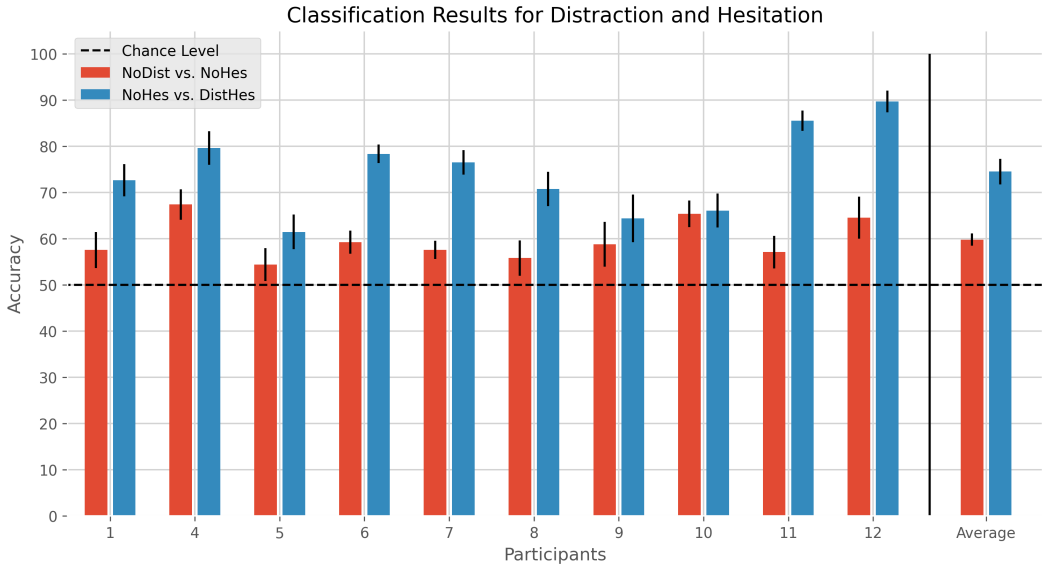


**Figure 4.** Classification results for 1-second EEG windows of the ambient and distraction conditions (red) and the distraction and hesitation conditions (blue) with standard error bars (black) calculated between the 10 splits of the cross-validation and the random runs.

The confusion matrix in Figure 5 shows consistent predictions for both classifications over all participants, with no clear preferences toward one class.

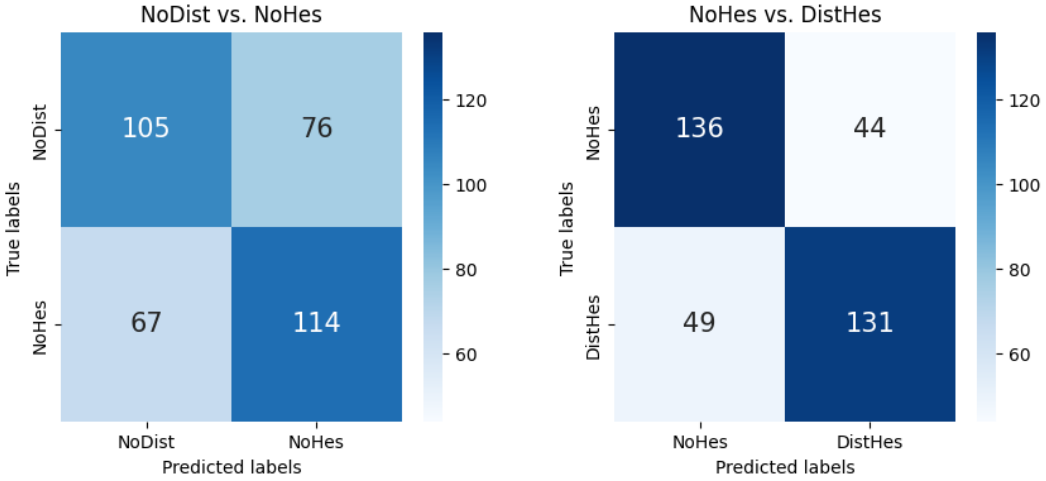


**Figure 5.** Confusion matrix for 1-second EEG windows of the ambient and distraction conditions and the distraction and hesitation conditions averaged over all folds of the 10-fold cross-validation and all participants. The color bar denotes the amount of 1-second windows.

### *4.2. Behavioral Data*

After each instruction, participants replied to the robot by indicating that they "understood" the instruction, that they "not understood" and needed the robot to repeat the

instruction, or that they were "uncertain" about the instruction (but would still try to execute it). Figure 6 visualizes the participant's self-assessment of the understanding of Pepper's instructions.

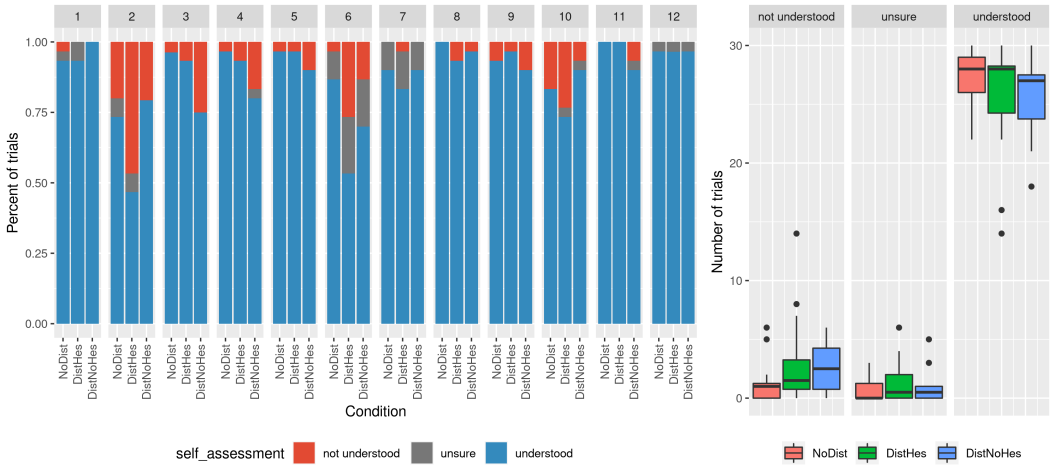


**Figure 6.** Self-assessment of understanding the instruction for each participant (**left**) and over all participants (**right**).

The participant's self-reported non-understanding of the instructions differed on average. In the NoDist condition, the average of not understanding instructions was $Mdn = 1$; in the DistHes condition, $Mdn = 1.5$; and in the DistNoHes condition, $Mdn = 2.5$. However, this finding did not reach statistical significance, $F(2, 22) = 2.0, p = 0.157$, $\eta^2 = 0.06$.

Figure 7 depicts the understanding (problems) for each participant on the left and over all participants on the right. The blue portion of the bar plots represents all trials in which the participant "understood" Pepper and correctly performed the corresponding action. The red area subsumes all cases in which the interaction was "unsuccessful", so an understanding problem occurred: the person was unsure whether they understood the actual instruction, asked for a repeat, or performed the action incorrectly. The number of "unsuccessful" interactions was statistically significantly different for the three conditions, $F(2, 22) = 4.41, p = 0.025, \eta^2 = 0.08$. In the NoDist condition, the participants had, on average, $Mdn = 2.5$ "unsuccessful" instructions, whereas in the DistHes and the DistNoHes conditions, the average "unsuccessful" instructions were higher ($Mdn = 4.5$, $Mdn_{DistNoHes} = 4.5$). The post-hoc test showed only a significant difference between the ambient and the conditions with distraction (whether with or without hesitation) and no difference in the distraction conditions with and without hesitation (see Table 3).

**Table 3.** Post-hoc pairwise comparison with Bonferroni correction of "unsuccessful" interaction between the conditions.

| Group 1 | Group 2 | df | $p_{adj}$ |
|---|---|---|---|
| NoDist | DistHes | 11 | 0.010 |
| NoDist | DistNoHes | 11 | 0.027 |
| DistHes | DistNoHes | 11 | 0.827 |

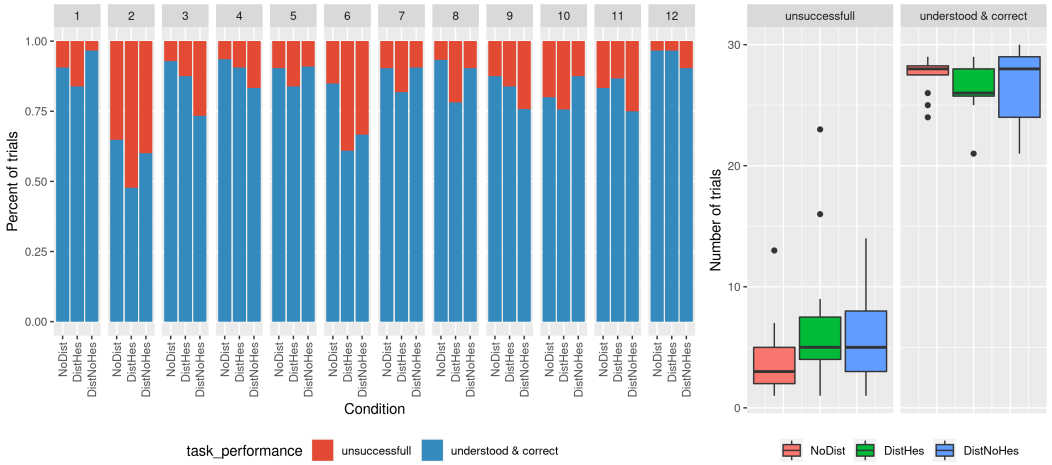


**Figure 7.** Understanding (problems) for each participant (**left**) and over all participants (**right**).

### *4.3. Questionnaires*

#### 4.3.1. Distractibility of the Participants

The general self-reported distractibility of the participants was measured with the Mind-Wandering-Questionnaire [50] on a 6-point likert scale with five items. The mean distractibility of all participants was 3.81 ($SD = 0.66$, Cronbach's $\alpha = 0.78$).

#### 4.3.2. Perception of the Distractions

The perception of the distraction was measured on a 6-point likert scale with four items (see Figure 8). The participants rated the provided distraction (background speech) as more disruptive than the general background noise ($M_{NoDist} = 2.5, M_{Dist} = 4; V = 0$, $p < 0.01, r = -0.5$). In addition, they stated that they could understand Pepper's voice less well during the provided distractions ($M_{NoDist} = 5, M_{Dist} = 4; V = 0, p < 0.05, r = 0.47$). As the participants stated that they found the background speech more disturbing than the general background noise, the manipulation test was successful.

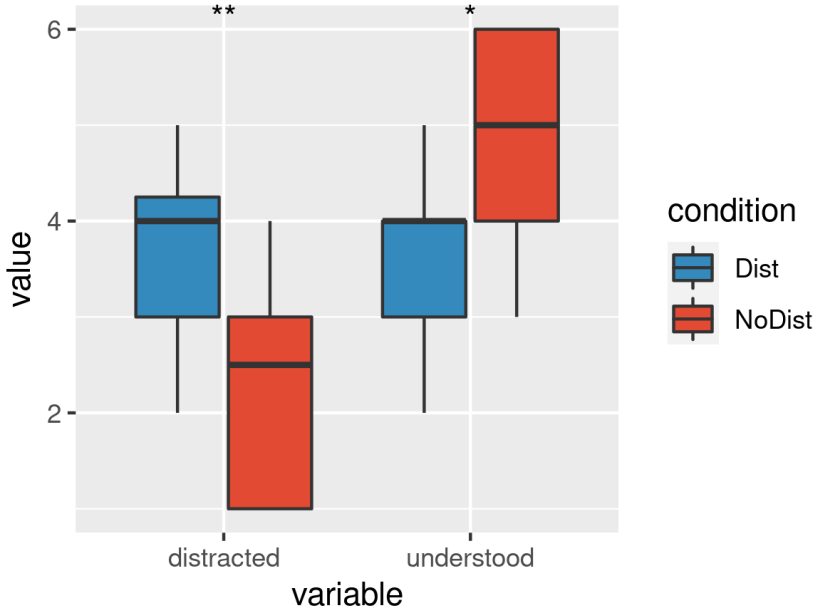


**Figure 8.** Perception of the distractions (significance level *: $p <= 0.05$, **: $p <= 0.01$ ).

#### 4.3.3. Intelligibility of Pepper's Instructions

The intelligibility of Pepper's instructions was rated on a 6-point likert scale with seven items (see Figure 9). Pepper's information was mostly perceived as appropriate

according to length, time, and comprehension. Only one-third of the participants perceived the instruction as rather untimely. They were felt to be a bit too slow and slightly delayed, which could be attributed to the hesitations.

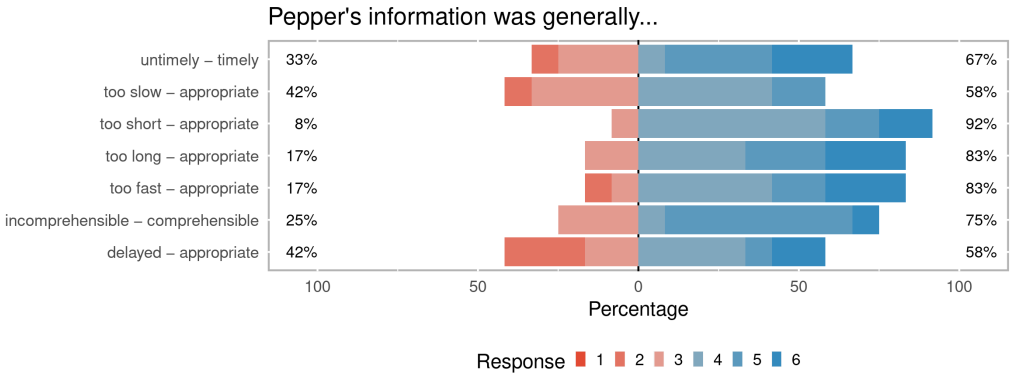


**Figure 9.** Intelligibility of Pepper's instructions.

#### 4.3.4. Perception of the Hesitations

At first, participants were asked, "Pepper didn't always react in the same way to the appearance of the background voices. What did you notice?" to find out whether they noticed the hesitations. Two participants stated that they had not noticed anything. Another two participants noticed the hesitations but did not recognize them as such but as errors in synthesis (words "swallowed"). Most participants recognized the hesitations ($n = 7$). The last participant only stated that the background voices were heavily demanding her concentration. Afterward, the hesitations were explained and rated on a 6-point likert scale with seven items (see Figure 10). The results indicate that most participants noticed the hesitations, but most of the participants rated them as rather unnatural and too long. We decided to use an additional silent pause of 1500 ms before the filler and 1000 ms after it so that the pauses are also recognized as hesitations and not as a normal break. Finding the right length of unfilled pauses is still an open research topic and should be addressed in further research.

Two-thirds of the participants stated that the hesitation did not cause them to stop listening to Pepper. Additionally, one-third of the participants said that the hesitation reattended them to Pepper's speech when they were distracted. Interestingly, none of the hesitation ratings significantly correlate with task performance.

#### 4.3.5. Synthesis Quality

Figure 11 shows the subjective ratings of Pepper's synthesis on the MOS-X2 questionnaire [51]. Pepper's synthetic voice received rather high values for intelligibility ($M = 8.0, SD = 1.76$), prosody ($M = 5.67, SD = 1.97$), and social impression ($M = 6.92$, $SD = 2.54$). However, the naturalness of Pepper's voice was perceived differently by the participants and received a rather low value on average ($M = 4.83, SD = 3.43$).

Splitting the results by the interaction script suggests that the participants who ended with the hesitation block (scripts 2 and 4) rated the naturalness rather low. (see Figure 12) It should be noted here that only three participants are available per script, and, therefore, a statistical evaluation is not possible. However, it could indicate that the hesitation leads to a less natural synthesis. This would be consistent with our previous research (e.g., [35]), as the natural synthesis of hesitations in human–robot interaction is still an important field of research and should be addressed in further research.

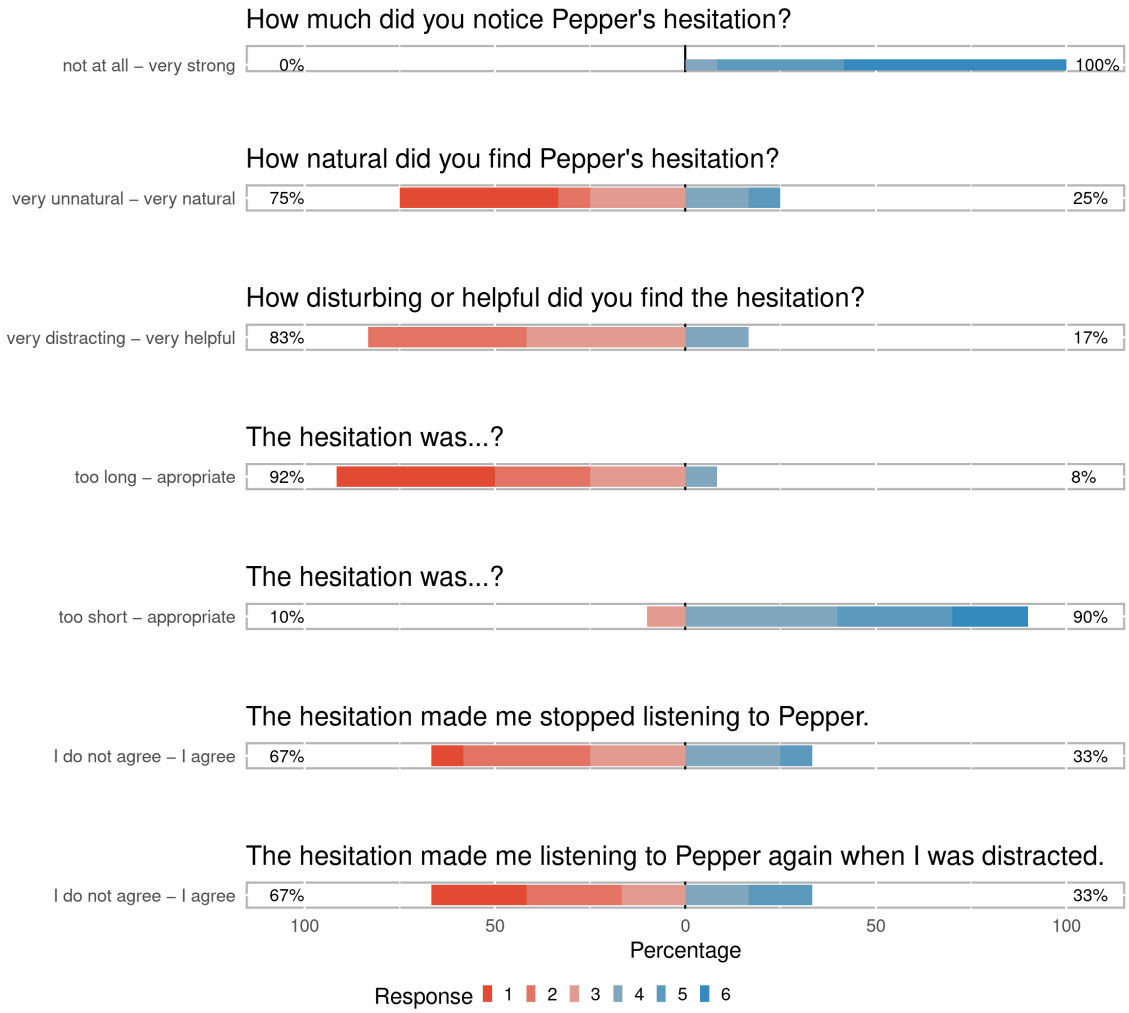


**Figure 10.** Participant's perception of the hesitations.

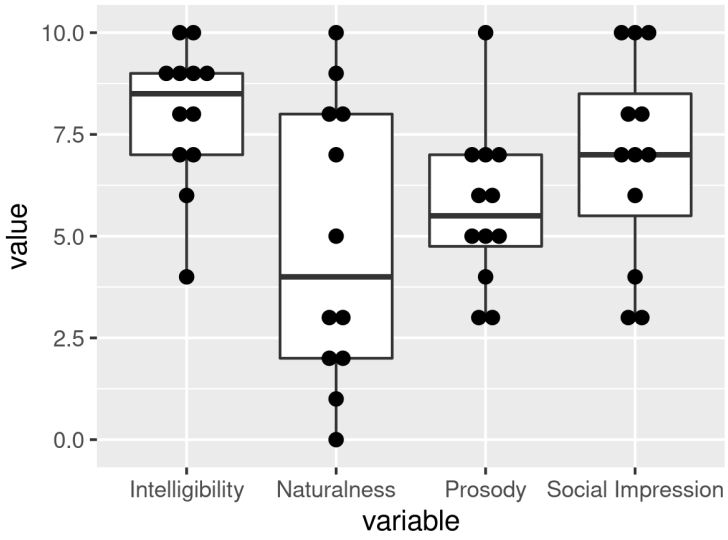


**Figure 11.** Subjective ratings of Pepper's synthetic voice.

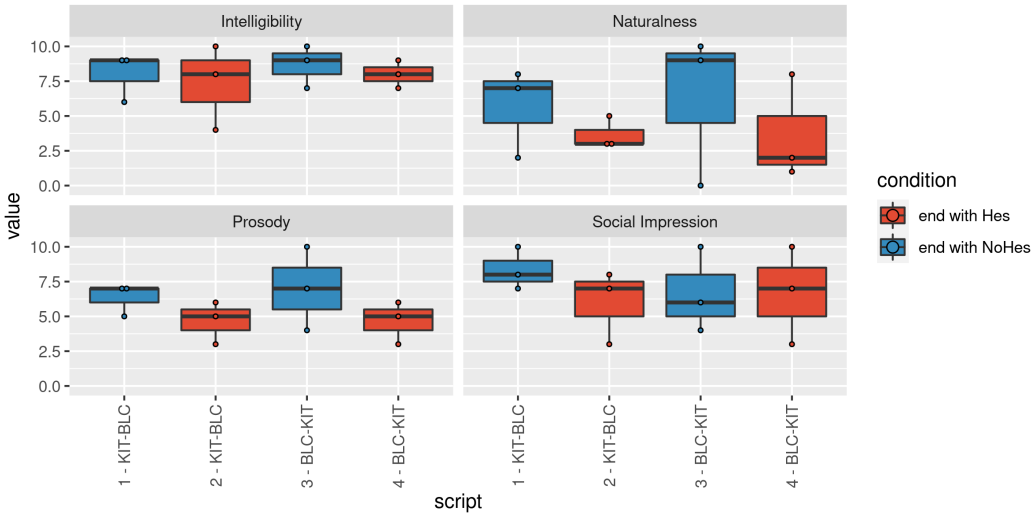


**Figure 12.** Synthesis quality separated by script.

## 5. Discussion

In our study, we conducted an experiment to investigate the distractibility in a human–robot-interaction setting by simulating daily activities instructed by a robot in a remote live setting. The results of our experiment provide some insights regarding EEG responses to scaffolding in HRI and the methodology of LabLinking.

### *5.1. SRQ1: Effects of Distraction*

Participants self-reported that they perceived the provided distraction (background voices) as more disruptive than the general ambient background noise. Thus, the manipulation check for our experiment was successful. In addition, the participants had significantly more understanding of problems in the disruptive condition. This was shown by participants' own insecurity about whether what was said was understood correctly, the request for repetition, or incorrect task execution.

Furthermore, we were able to classify between the undistracted baseline and the distracted condition. Hence, we could find effects on single-trial EEG based on the distraction or non-understanding condition in the human–robot interaction (**SRQ1**). The single-trial EEG analyses of neural responses during speech perception showed a statistically significant average discriminability of 60% between trials where background voices were used as a source of distraction, compared to the control trials with ambient noise. These findings were supported by the feedback of the participants, who reported greater difficulty in understanding and maintaining focus on the instructions in the distraction condition. Together, these results provide additional validation of the effect of distraction through speech as compared to a noisy environment on a neural level [52]. Our analysis demonstrated that the best classification results were achieved through a combination of the theta and alpha band power features. Notably, an increase in theta band power and a decrease in alpha band power are associated with reduced attention and increased distractibility, and high correspondence to the cortical tracking of speech [53–55]. Therefore, our results imply that the auditory distraction was successful and could potentially be detected using neural signals, particularly when using small window sizes. While the classification accuracies achieved in our study may still be considered quite low and not yet applicable for real-world applications, they do open up the possibility of building better systems and implementing adaptive approaches via real-time feedback to enhance human–robot interaction in dynamic environments. Critically, it cannot be excluded that the effects arise due to non-understanding of the participants as opposed to distraction. Further analyses are needed to identify the underlying cognitive processes as reflected by the EEG data. In this first approach, we compared the trials without disruption with trials with

disruption (but no hesitation). As a next step, we will analyze the successfully understood and correctly performed trials and compare them with the unsuccessful interaction trials.

### *5.2. SRQ2: Effects of Hesitation Scaffolding Strategy*

Our second research question (**SRQ2**) addresses whether it is possible to find EEG responses related to scaffolding strategies in HRI, in our example, the hesitation strategy. Again, we were able to classify the conditions of distraction vs. hesitation. However, in the hesitation condition, participants had no fewer understanding problems than in the distraction condition without hesitations. Thus, this experiment was unable to reproduce the positive effects of hesitations from previous studies (e.g., [33]). The results of the understanding for each participant in Figure 7 show that the success of the hesitation scaffolding strategy could be very person-dependent. The hesitations only seem to be beneficial for some individuals; other participants seem more likely to be further distracted by the hesitations. This is also reflected by the results of the subjective ratings of the hesitations. The possible benefit of hesitation as a scaffolding strategy will be addressed in further studies, where the subjective ratings should be assessed additionally between blocks with different conditions. In doing so, we could gain more insights into the subjective perception of Pepper's hesitations. Furthermore, the hesitation strategy itself needs to be investigated further. Finding the right length of unfilled pauses is still an open research topic and should be addressed in further research. To make sure the hesitations are recognized as hesitations and not as a normal break, we decided to use an additional silent pauses before and after the filler. The actual total pause in Pepper's speech synthesis was about 3.5 s before the filler and 2.5 s after it. Pepper kept gesturing during the breaks. However, most participants rated this as too long. In addition, some participants did not recognize the hesitations as such but as errors in synthesis. Further research is needed to synthesize hesitations in a robot's live system.

The EEG data analysis revealed significant differences in neural responses between the two conditions. Averaging over all participants, the single-trial classification accuracy was 73% for trials following the scaffolding strategy compared to those where no such strategy was employed. These findings are consistent with the perception of the participants in Figure 10, which reflects that the robots' hesitation was clearly noticeable for every participant. However, the majority of participants reported perceiving Peppers's hesitation as unnatural and even distracting. A limitation of the current study is the ability to identify the specific cognitive processes underlying the observed neural responses. It remains unclear whether the observed difference is a direct result of the scaffolding strategy, hence an improvement in attention, or simply a reflexive response to the preceding action of the robot. In particular, the participants provided rather negative feedback regarding the effectiveness of redirecting attention. To definitely answer this question, more in-depth analyses of the EEG signals are required to find evidence for the involvement of specific cognitive processes that may be affected by the robot's scaffolding strategies. Nevertheless, the robust detection of the robot's intervention in the neural responses supports the use of scaffolding strategies such as hesitation for enhancing human–robot interaction with neural assistance.

### *5.3. MRQ1: Benefits of LabLinking*

Beside the research regarding EEG responses to scaffolding in HRI, we investigated the LabLinking method (**MRQ1**). It enabled the interdisciplinary joint HRI research between our two laboratories: the Medical Assistance Systems Group (MAS) at Bielefeld University and the Cognitive Systems Lab (CSL) at the University of Bremen. The method allowed the presented HRI study to be carried out without having the expensive hardware in one place. This saves resources in many ways. First, (i) neither in Bremen nor in Bielefeld new hardware had to be bought. Additionally, the hardware did not have to be transported to the other laboratory, (ii) which always involves a risk of damage during transport. Furthermore, since the hardware did not have to be taken to the other laboratory, (iii) it

could be used for further on-site studies meanwhile. Moreover, (iv) several resources for traveling were saved since the labs are connected via LabLinking. It was possible to carry out the presented study successfully, although we only visited each other's laboratory once. We had approximately 25 meetings over one year (excluding the study itself) and 9 days on which the study was conducted. It was not necessary to travel to the other laboratory to test the joint setup, which saved travel costs and $CO_2$ emissions. In addition, we were able to conduct the research despite the (travel) restrictions imposed by the COVID-19 pandemic. The connection of the two laboratories allows a further (v) economical use of human resources regarding their time. Experts in the individual areas were able to share their knowledge more quickly and easily. This enabled an interdisciplinary research exchange in a large team.

### *5.4. MRQ2: Challenges of LabLinking*

In addition to these advantages, however, the LabLinking method also revealed some challenges (**MRQ2**), which required us to develop a number of best practices during iterations of the experiment development. First, it is important to maintain a joint experiment state and an unambiguous communication protocol between the experimenters on both sites. This can be supported through the combination of multiple channels: live monitors of the video and audio recordings on the different sites, explicit terminology, and an explicit flowchart, which determines which site is responsible for confirmations or taking the initiative to advance the experiment. We also formalized this flowchart in experiment control programs on both ends, which enforce this protocol programmatically. Second, it is important to also take care of temporal synchronization from a technical perspective. For this purpose, we use a unified messaging protocol (ROS in our case) through which all information is passed, including timestamps for each event. These also allow us to monitor data transmission latency. The latency for the transmission of the speech synthesis of the Bielefeld robot and arrival to the microphones in the Bremen laboratory was measured to be approximately $415 \pm 170$ ms. This process of synchronization creates a single repository of all data occurring within the experiment, all aligned according to a common clock.

### *5.5. Further Research*

Apart from the already mentioned future work, we want to address the following research directions.

**Early detection of non-understanding:** If it is possible to detect problems of understanding early in the interaction based on the EEG correlates, the Pepper robot could use these to identify the unsuccessful interactions during the interaction and correct them through linguistic strategies (e.g., hesitations).

**Adaption:** Hesitations only seem to be beneficial for some individuals, whereas other participants seem more likely to be further distracted by the hesitations. A more detailed analysis of the data could provide information here. If, for example, the EEG data show differences in the groups, an appropriate (hesitation) strategy for the particular person could be selected.

**Pipeline Automatization:** To facilitate the experiment process, we want to further automatize the pipeline for experiment execution and data analysis. In particular, we want to add automatic object detection to identify and locate the objects on the table. This will enable an automatic, fine-grained scoring of item placement and will, furthermore, provide additional context information to the robot.

## 6. Conclusions

In this paper, we presented the first results of a LabLinking pilot study investigating EEG correlates of distractions and hesitations in human–robot interaction. We were able to show that (i) the EEG correlates in the distracted condition are different from the baseline condition without distractions, and we can classify them. In addition, (ii) we could differentiate the EEG correlates of distraction with and without hesitations. Finally, (iii) we

presented the benefits and challenges of the LabLinking method for enabling interdisciplinary joint research HRI experiments between multiple labs. This proof-of-concept study showed that it is possible to conduct HRI studies via LabLinking and lays a first foundation for more in-depth research into robotic scaffolding strategies.

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

**Funding:** The research reported in this paper has been partially supported by the German Research Foundation DFG, as part of the Collaborative Research Center (Sonderforschungsbereich) 1320 Project-ID 329551904 "EASE—Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org/ (accessed on 21 February 2023)). The research was conducted in subprojects H03 and H04. It has also been partially supported by the TRR 318 "Constructing Explainability" TRR 318/1 2021 – 438445824. Furthermore, this project was partly funded by via the High-Profile Area Minds, Media, Machines, at the University of Bremen.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and is covered by the approval for the CRC EASE by the Ethics Committee of the University of Bremen (issued on 7 April 2017).

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

**Data Availability Statement:** All data and code for experiments and processing are available in the following repository hosted by the Open Science Framework: https://osf.io/jyhdq/ (accessed on 21 February 2023). The data are documented according to the experiment model outlined by Putze et al. [56].

**Acknowledgments:** The authors thank Ayla Luong and Mustafa Mosavy for their support in the data collection and the anonymous reviewers for their constructive feedback.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| BCI | brain-computer interfaces |
| BLC | wooden blocks |
| CSL | Cognitive Systems Lab |
| EASE | Everyday Activity Science and Engineering |
| EEG | electroencephalogram |
| ERPs | event-related potentials |
| HRI | human–robot interaction |
| KIT | kitchen objects |
| LLL | LabLinking Level |
| MAS | Medical Assistance Systems Group |
| MMN | mismatch negativity |
| ROS | Robot Operating System |

## References

1. Belhassein, K.; Fernández-Castro, V.; Mayima, A.; Clodic, A.; Pacherie, E.; Guidetti, M.; Alami, R.; Cochet, H. Addressing joint action challenges in HRI: Insights from psychology and philosophy. *Acta Psychol.* **2022**, *222*, 103476. [CrossRef] [PubMed]
2. Tomasello, M.; Carpenter, M. Shared intentionality. *Dev. Sci.* **2007**, *1*, 121–125. [CrossRef]
3. Clark, H.H. *Using Language*; Cambridge University Press: Cambridge, UK, 1996.
4. Garrod, S.; Pickering, M.J. Joint action, interactive alignment, and dialog. *Top. Cogn. Sci.* **2009**, *1*, 292–304. [CrossRef]
5. Allwood, J.; Nivre, J.; Ahlsén, E. On the semantics and pragmatics of linguistic feedback. *J. Semant.* **1992**, *9*, 1–26. [CrossRef]

6. Klotz, D.; Wienke, J.; Peltason, J.; Wrede, B.; Wrede, S.; Khalidov, V.; Odobez, J.M. Engagement-based Multi-party Dialog with a Humanoid Robot. In Proceedings of the SIGDIAL 2011 Conference, Portland, OR, USA, 17–18 June 2011; pp. 341–343.
7. Rogers, T.E.; Sekmen, A.S.; Peng, J. Attention Mechanisms for Social Engagements of Robots with Multiple People. In Proceedings of the ROMAN 2006—The 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, UK, 6–8 September 2006; pp. 605–610. [CrossRef]
8. Salam, H.; Chetouani, M. A multi-level context-based modeling of engagement in Human-Robot Interaction. In Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 3, pp. 1–6. [CrossRef]
9. Carlmeyer, B.; Schlangen, D.; Wrede, B. Exploring self-interruptions as a strategy for regaining the attention of distracted users. In Proceedings of the 1st Workshop on Embodied Interaction with Smart Environments, Tokyo, Japan, 16 November 2016; pp. 1–6.
10. Carlmeyer, B.; Betz, S.; Wagner, P.; Wrede, B.; Schlangen, D. The Hesitating Robot—Implementation and First Impressions. In Proceedings of the Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI '18, New York, NY, USA, 5–8 March 2018; pp. 77–78. [CrossRef]
11. Corley, M.; Stewart, O.W. Hesitation Disfluencies in Spontaneous Speech: The Meaning of um. *Lang. Linguist. Compass* **2008**, *2*, 589–602. [CrossRef]
12. Finlayson, I.R.; Corley, M. Disfluency in Dialogue: An Intentional Signal from the Speaker? *Psychon. Bull. Rev.* **2012**, *19*, 921–928. [CrossRef] [PubMed]
13. Shriberg, E. Disfluencies in Switchboard. In Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia, PA, USA, 3–6 October 1996.
14. Merlo, S.; Mansur, L. Descriptive discourse: Topic familiarity and disfluencies. *J. Commun. Disord.* **2004**, *37*, 489–503. [CrossRef]
15. Smith, V.; Clark, H. On the course of answering questions. *J. Mem. Lang.* **1993**, *32*, 25–38. [CrossRef]
16. Collard, P. Disfluency and Listeners' Attention: An Investigation of the Immediate and Lasting Effects of Hesitations in Speech. Ph.D. Thesis, University of Edinburgh, Edinburgh, UK, 2009.
17. Fraundorf, S.H.; Watson, D.G. The disfluent discourse: Effects of filled pauses on recall. *J. Mem. Lang.* **2011**, *65*, 161–175. [CrossRef]
18. Schultz, T.; Putze, F.; Fehr, T.; Meier, M.; Mason, C.; Ahrens, F.; Herrmann, M. Linking Labs: Interconnecting Experimental Environments. *bioRxiv* **2021**.
19. Groß, A.; Schütze, C.; Wrede, B.; Richter, B. An Architecture Supporting Configurable Autonomous Multimodal Joint-Attention-Therapy for Various Robotic Systems. In Proceedings of the Companion Publication of the 2022 International Conference on Multimodal Interaction, New York, NY, USA, 7–11 November 2022; pp. 154–159. [CrossRef]
20. Schütze, C.; Groß, A.; Wrede, B.; Richter, B. Enabling Non-Technical Domain Experts to Create Robot-Assisted Therapeutic Scenarios via Visual Programming. In Proceedings of the Companion Publication of the 2022 International Conference on Multimodal Interaction, ACM, New York, NY, USA, 7–11 November 2022; pp. 166–170. [CrossRef]
21. Hegel, F.; Muhl, C.; Wrede, B.; Hielscher-Fastabend, M.; Sagerer, G. Understanding Social Robots. In Proceedings of the Int. Conf. Advances in Computer-Human Interactions (ACHI), Cancun, Mexico, 1–7 February 2009; pp. 169–174. [CrossRef]
22. Schultz, T.; Maedche, A. Biosignals meet Adaptive Systems. *Springer Nat. Appl. Sci.* 2023, *in press*.
23. Meier, M.; Mason, C.; Putze, F.; Schultz, T. Comparative Analysis of Think-Aloud Methods for Everyday Activities in the Context of Cognitive Robotics. In Proceedings of the Interspeech, Graz, Austria, 15–19 September 2019; pp. 10–14.
24. Schultz, T. Biosignal Processing for Human-Machine Interaction. In Proceedings of the Interspeech, Graz, Austria, 15–19 September 2019. Available online: https://www.youtube.com/watch?v=F0-r6V6wNRA (accessed on 21 February 2023).
25. Schultz, T.; Kirchhoff, K. (Eds.) *Multilingual Speech Processing*; Academic Press: Burlington, MA, USA, 2006.
26. Mason, C.; Gadzicki, K.; Meier, M.; Ahrens, F.; Kluss, T.; Maldonado, J.; Putze, F.; Fehr, T.; Zetzsche, C.; Herrmann, M.; et al. From Human to Robot Everyday Activity. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 8997–9004. [CrossRef]
27. Corley, M.; MacGregor, L.J.; Donaldson, D.I. It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition* **2007**, *105*, 658–668. [CrossRef] [PubMed]
28. Chun, M.M.; Golomb, J.D.; Turk-Browne, N.B. A taxonomy of external and internal attention. *Annu. Rev. Psychol.* **2011**, *62*, 73–101. [CrossRef] [PubMed]
29. Vortmann, L.M.; Putze, F. Attention-aware brain computer interface to avoid distractions in augmented reality. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–8.
30. Beltrán, E.T.M.; Pérez, M.Q.; Bernal, S.L.; Pérez, G.M.; Celdrán, A.H. SAFECAR: A Brain–Computer Interface and intelligent framework to detect drivers' distractions. *Expert Syst. Appl.* **2022**, *203*, 117402. [CrossRef]
31. Salous, M.; Küster, D.; Scheck, K.; Dikfidan, A.; Neumann, T.; Putze, F.; Schultz, T. SmartHelm: User Studies from Lab to Field for Attention Modeling. In Proceedings of the 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Prague, Czech Republic, 9–12 October 2022; pp. 1012–1019.
32. Apicella, A.; Arpaia, P.; Frosolone, M.; Moccaldi, N. High-wearable EEG-based distraction detection in motor rehabilitation. *Sci. Rep.* **2021**, *11*, 5297. [CrossRef]

33. Richter, B. The Attention-Hesitation Model. A Non-Intrusive Intervention Strategy for Incremental Smart Home Dialogue Management. Ph.D. Thesis, Bielefeld University, Bielefeld, Germany, 2021.
34. Carlmeyer, B.; Schlangen, D.; Wrede, B. "Look at Me!": Self-Interruptions as Attention Booster? In Proceedings of the Fourth International Conference on Human Agent Interaction, HAI '16, New York, NY, USA, 4–7 October 2016; pp. 221–224. [CrossRef]
35. Betz, S.; Carlmeyer, B.; Wagner, P.; Wrede, B. Interactive Hesitation Synthesis: Modelling and Evaluation. *Multimodal Technol. Interact.* **2018**, *2*, 9. [CrossRef]
36. Lücking, P.; Lier, F.; Bernotat, J.; Wachsmuth, S.; Ŝabanović, S.; Eyssel, F. Geographically distributed deployment of reproducible HRI experiments in an interdisciplinary research context. In Proceedings of the Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 181–182.
37. Pavlov, Y.G.; Adamian, N.; Appelhoff, S.; Arvaneh, M.; Benwell, C.S.; Beste, C.; Bland, A.R.; Bradford, D.E.; Bublatzky, F.; Busch, N.A.; et al. # EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex* **2021**, *144*, 213–229.
38. Prado, P.; Birba, A.; Cruzat, J.; Santamaría-García, H.; Parra, M.; Moguilner, S.; Tagliazucchi, E.; Ibáñez, A. Dementia ConnEEGtome: Towards multicentric harmonization of EEG connectivity in neurodegeneration. *Int. J. Psychophysiol.* **2022**, *172*, 24–38. [CrossRef]
39. Li, J.; Li, Z.; Feng, Y.; Liu, Y.; Shi, G. Development of a human–robot hybrid intelligent system based on brain teleoperation and deep learning SLAM. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 1664–1674. [CrossRef]
40. Liu, Y.; Habibnezhad, M.; Jebelli, H. Brain-computer interface for hands-free teleoperation of construction robots. *Autom. Constr.* **2021**, *123*, 103523. [CrossRef]
41. Beraldo, G.; Tonin, L.; Millán, J.d.R.; Menegatti, E. Shared intelligence for robot teleoperation via bmi. *IEEE Trans. Hum.-Mach. Syst.* **2022**, *52*, 400–409. [CrossRef]
42. Quigley, M.; Gerkey, B.; Conley, K.; Faust, J.; Foote, T.; Leibs, J.; Berger, E.; Wheeler, R.; Ng, A. ROS: An open-source Robot Operating System. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) Workshop on Open Source Robotics, Kobe, Japan, 31 May 2009.
43. Lier, F.; Wienke, J.; Nordmann, A.; Wachsmuth, S.; Wrede, S. The cognitive interaction toolkit–improving reproducibility of robotic systems experiments. In Proceedings of the Simulation, Modeling, and Programming for Autonomous Robots: 4th International Conference, SIMPAR 2014, Bergamo, Italy, 20–23 October 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 400–411.
44. Ros-Drivers. Ros-drivers/video_stream_opencv: A package to open video streams and publish them in Ros using the OPENCV videocapture mechanism. Available online: https://github.com/ros-drivers/video_stream_opencv Branch: master, commit: 65949bd (accessed on 21 February 2023).
45. GStreamer. GStreamer/gstreamer: Gstreamer open-source multimedia framework. Available online: https://github.com/GStreamer/gstreamer (accessed on 21 February 2023).
46. Ros-Naoqi. Ros-naoqi/naoqi_driver: C++ Bridge based on libqi. Available online: https://github.com/ros-naoqi/naoqi_driver Branch: master, commit: a2dd658 (accessed on 21 February 2023).
47. Betz, S.; Wagner, P.; Voße, J. Deriving a strategy for synthesizing lengthening disfluencies based on spontaneous conversational speech data. In Proceedings of the Tagungsband Der 12, Tagung Phonetik und Phonologie im Deutschsprachigen Raum, München, Germany, 12–14 October 2016.
48. Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D.A.; Strohmeier, D.; Brodbeck, C.; Parkkonen, L.; Hämäläinen, M.S. MNE software for processing MEG and EEG data. *Neuroimage* **2014**, *86*, 446–460. [CrossRef] [PubMed]
49. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
50. Mrazek, M.D.; Phillips, D.T.; Franklin, M.S.; Broadway, J.M.; Schooler, J.W. Young and restless: Validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Front. Psychol.* **2013**, *4*, 560. [CrossRef] [PubMed]
51. Lewis, J. Investigating MOS-X Ratings of Synthetic and Human Voices. *Assoc. Voice Interact. Des.* **2018**, *2*, 1–22.
52. Olguin, A.; Bekinschtein, T.A.; Bozic, M. Neural encoding of attended continuous speech under different types of interference. *J. Cogn. Neurosci.* **2018**, *30*, 1606–1619. [CrossRef]
53. Kerlin, J.R.; Shahin, A.J.; Miller, L.M. Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J. Neurosci.* **2010**, *30*, 620–628. [CrossRef] [PubMed]
54. Hambrook, D.A.; Tata, M.S. Theta-band phase tracking in the two-talker problem. *Brain Lang.* **2014**, *135*, 52–56. [CrossRef]
55. Ding, N.; Simon, J.Z. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* **2012**, *107*, 78–89. [CrossRef]
56. Putze, F.; Putze, S.; Sagehorn, M.; Micek, C.; Solovey, E.T. Understanding hci practices and challenges of experiment reporting with brain signals: Towards reproducibility and reuse. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2022**, *29*, 1–43. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.