Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Project No. 611391

DREAM
Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Grant Agreement Type: Collaborative Project
Grant Agreement Number: 611391

D3.4.5 System Integration Progress Report

Due date: **1/4/2019**
Submission Date: **12/4/2019**

Start date of project: **01/04/2014** Duration: **54 months**

Organisation name of lead contractor for this deliverable: **University of Skövde**

Responsible Person: **E. Billing** Revision: **1**

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Contents

Executive Summary

Deliverable D3.4 is an annual progress report on the integration of the software developed in work packages WP4, WP5, and WP6. This is the Month 60 and final progress report. This report focus on two contributions during the past year. Firstly, we present an extension of the open source release with an open dataset comprising anonymous behavioral data processed from recorded interventions with autistic children. The dataset contains upper body skeleton data, head orientation, and eye-gaze, all specified in a common 3D frame of reference. Secondly, we present a method for traceable assessments of user's state, allowing the overall statistics constituting clinical outcomes from the 2017-2018 evaluation to be traced back to individual events in the recorded interventions. In this way we achieve both transparency and verifiableness in automatic assessments produced by the DREAM system.

Principal Contributors

The main authors of this deliverable are as follows (in alphabetical order).

Erik Billing, University of Skövde
Haibin Cai, University of Portsmouth
Robert Homewood, University of Skövde
Vipul Nair, University of Skövde
Emmanuel Senft, Plymouth University

1 Introduction

This is the Month 60 and final progress report for WP3, covering the period from April 2018 to the end of the project, March 2019. We came into this final year with a large clinical evaluation that was, at the time, almost finished. With that, the work with system integration was to a large degree over and WP3 has instead, during period five, focused on open source and support for WP2 and WP5 in the analysis of data. We have put a lot of energy into the open source release that was first presented at the P3 review (see D3.4.4). During this period, we have primarily engaged into an open release of the DREAM dataset, presented in Section 2. Parts of the dataset is now available at `github.com/dream2020/data` and the full dataset will be released together with a journal publication, which is currently in preparation. The data is licensed under a Creative Commons (CC) licence, see Section 2.2 for details.

WP3 has also supported the evaluations performed within WP2 and WP5 with data analysis, allowing the system's automatic assessment of the user's state to be traced back to the specific events at its precise time of the therapy. This work is presented in detail in Section 3.

There has been two integration meetings during the past year. The first meeting took place in Skövde, September 20-21, 2018 and the second meeting was hosted by SBR in Paris. See Section 4 for details. As previous years, there has also been other meetings. For example, large parts of the DREAM team was engaged in the Workshop on *Social Robots in Therapy and Care* (Hernández García et al., 2019), co-arranged with the 14th ACM/IEEE International Conference on Human Robot Interaction (HRI 2019).

Finally, in Section 5, two new utility applications, *ELAN export* and *JSON export*, are presented. Both utilities constitute important pieces in the production of the open dataset (Section 2) and the traceable assessments (Section 3). Also, many of the other utilities presented in earlier revisions of this deliverable have been used in the work with the open dataset.

2 Open dataset

During the 2017-2018 evaluation, we have recorded a total of 364 hours of therapy with 76 children. Therapy was conducted at several different places in Romania, but were all recorded using a the same sensorized therapy room design, described in D4.1. The setup is visible in Figure 1.

37 children interacted with a robot, referred to as the RET group. The other 39 children was given traditional ABA therapy, referred to as the SHT group. On average, each child took part in 50 sessions of therapy. The length of each session varied from a few minutes up to 40 minutes. To our knowledge, this is the largest data set of autism therapy involving robots and probably the largest recorded data set of children interacting with robots in general. With the exception of the PInSoRo dataset (Lemaignan et al., 2018), we have not found any other large open sets of data covering children's interactions with robots.

While we see a potentially great value in a public release of the DREAM dataset,

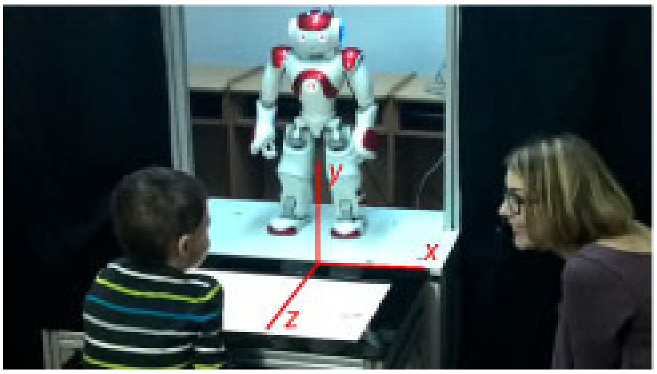


Figure 1: Example of the therapy environment. Red axes describe the orientation for the joint coordinate system for all data in the DREAM dataset.

we can not, according to the ethical approval and signed agreements with caregivers, distribute any primary data form the study. Primary data refers to direct measurements, e.g., video and audio recordings, including children in therapy. However, secondary data may be distributed freely. Secondary data refers processed measurements from the primary data.

Given that a large part of DREAM involves processing of sensory data into higher level perceptions, we have access to comprehensive secondary data from the evaluation. Any variables related to the clinical evaluation can however not yet be released, concerning for example the child's performances during therapy. This type of information may be included in a later release of the dataset. Other variables, e.g., face expressions, have relatively low reliability and were excluded from dataset for this reason. Finally, any variables that may reveal the child's identity has been excluded from the dataset.

After analysis and discussions with therapists, the following data has been selected for inclusion in the public release of the DREAM dataset:

1. Child id (numerical index),
2. Child's gender,
3. Child's age,
4. 3D skeleton comprising joint positions for upper body,
5. 3D head position and orientation,
6. 3D eye gaze vectors,
7. Therapy condition (RET or SHT),
8. Therapy task (Joint attention, Imitation, or Turn-taking),
9. Date of recording,
10. Initial ADOS score.

Table 1: Example structure of the open DREAM dataset, in JSON format. *"[...]"* corresponds to numeric arrays that are too long to include here.

```
{
  "participant": {"id": 1, "gender": "male", "age": 5},
  "date": "2017-04-23T18:25:43.511Z",
  "ados_initial": 13,
  "condition": "RET",
  "task": "turn-taking",
  "eye_gaze": {"rx": [...],"ry": [...],"rz": [...]},
  "head_gaze": {"rx": [...],"ry": [...],"rz": [...]},
  "skeleton": {
    "elbow_left": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "elbow_right": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "hand_left": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "hand_right": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "head": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "sholder_center": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "sholder_left": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "sholder_right": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "wrist_left": {"x": [...],"y": [...],"z": [...],"confidence": [...]},
    "wrist_right": {"x": [...],"y": [...],"z": [...],"confidence": [...]}
  }
}
```

This data comes primarily from offline analysis of recorded interventions, computed with the utility component *sensoryAnalysisOffline*. This utility application is an offline-version of the integrated camera system constituting a critical part of the live system. See (Cai et al., 2018) for details.

The output of the offline analysis is logged to event-based log files using the *sensoryInterpretationLogger*, specified in D3.4.2. The raw log files were however deemed unsuitable for a public release. Although following a basic comma separated values (csv) pattern, this format requires extensive processing in order to extract relevant variables. These log-files also lack relevant information coming from the user model.

With the ambition of releasing the dataset in an easily accessible, well-specified, and commonly used file format, JavaScript Object Notation (JSON)[1] was selected. JSON has many of the attractive attributes found in XML, including standard libraries for most programming languages, validation patterns, and human readability. However, JSON is less verbose than XML and includes standard notation for arrays, making it much more suitable for storing numeric data.

An example structure of the DREAM dataset JSON format is included in Table 1. The complete format is specified by the JSON Schema[2] found in Appendix A. An important

[1] JavaScript Object Notation (JSON) is a stripped form of the JavaScript programming language, intended for data representation. `https://www.json.org`

[2] JSON Schema is a vocabulary that allows annotation and validation of JSON documents. Please refer to `https://json-schema.org` for more information. A reference implementation and validation engine is

attribute of this dataset is that all attributes are defined in a common frame of reference using a Cartesian coordinate system. The orientation of the Cartesian space in relation to the therapy environment is visualized in Figure 1.

2.1 Public release

At the time of writing, about 1200 sessions has been processed and stored in the DREAM dataset format specified in the previous section. A small sample of the dataset has also been released publicly at `https://github.com/dream2020/data/`. The full dataset will be released once the dataset is published as a journal publication, which is currently in preparation.

2.2 Licence

The Open DREAM dataset is licensed under

> **Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License**.

This licence permits *copying and redistribution of the material in any medium or format* and states that *the licensor cannot revoke these freedoms as long as you follow the license terms*. This freedom is given under the following terms:

1. **Attribution** – You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
2. **NonCommercial** – You may not use the material for commercial purposes.
3. **NoDerivatives** – If you remix, transform, or build upon the material, you may not distribute the modified material.
4. **No additional restrictions** – You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices

- You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.
- No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

The complete licence is available at `https://creativecommons.org/licenses/by-nc-nd/4.0/`.

found at `https://www.jsonschemavalidator.net`.

3 Traceable assessment of user state

Information about users are commonly stored in a *User Model*[3], i.e., some kind of representation of the user's current state. In DREAM, the a user model was implemented to store prior diagnosis information about each child, as well as their ongoing intervention session performance with both the robot and the therapist. The information was intended for use primarily in WP2, constituting parts of the clinical results as well as input to the system performance analysis. Parts of the information in the user model is also displayed in the DREAM system GUI as it is useful information for the therapists to view when interventions are being conducted. Each user model is implemented as an XML structure comprising user id, name, gender, and a list of therapy sessions. For each session, the user model specifies a session id, session name, therapy condition (RET or SHT), script id, as well as the child's performance during the session as assessed by the combined SPARC robot-therapist team. For more information, please refer to D6.3.4.

To the most part, the user model implemented for the DREAM system fills its purpose. However, during the evaluation conducted during 2017-2018, a number of new use cases came up that were not foreseen at design time. Most importantly, the user model stores the *relevant* performances as assessed at runtime. The exact time of assessment and the reasons why the system made a certain assessment is not included in the user model.

During analysis, we realized that the lack of traceability in the user model was a problem. While the user model offered a convenient representation allowing the result of the automatic assessment to be compared to manual annotations of recorded interventions, it lacked information on exact causes. We argue that this issue is to a large degree an inherent problem in atomized tools for user assessment. There is always a negotiation part of any annotation or assessment process performed manually. When this process is atomized, this negotiation component is easily neglected, and so it was in DREAM. An atomized assessment tool that only provides access to the resulting assessment is unlikely to work in practise, since unexpected situations or events will require adjustments to the original assessment protocol.

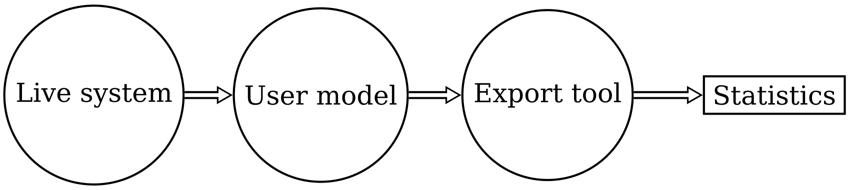
The approach we took to deal with the above described issues was to make the automatic assessment traceable through visualization. We realized that an important part of automatic assessment is to provide a platform for communication and negotiation between system developers and clinical staff. Visualization is critical when explaining why the system makes a certain interpretation and makes it easier for therapists to provide valuable feedback when the assessment needs to be adjusted. Visualization also work as an important tool for verification of the processed outcomes, e.g., in relation to the problems discussed in D3.4.4.

Through the whole project we have used the video annotation tool ELAN[4] for manual annotations of therapy sessions. ELAN is one of the most established tools for video annotations and is commonly used in clinical studies. From a technical point of view, ELAN stores data in an open and well-specified XML-based format referred to as the

[3]For background on user modelling, see the Springer LNAI series on User Modeling, Adaptation and Personalization, `http://www.um.org/umap2018/`

[4]ELAN Annotation Tool by the Max Planck Institute: `https://tla.mpi.nl/tools/tla-tools/elan/`

Figure 2: Visualization of the original, non-traceable, assessment process. Here, the *Live system* refers to the complete DREAM RET system, the *User model* is implemented as the userModel system component (D6.3.4), the *Export tool* is the UserDataExport presented in D3.4.3. *Statistics* refers to the complete intervention statistics used by clinicians.



ELAN Annotation Format (EAF).[5] We realized that ELAN would be a useful platform for integrating and visualizing automatic assessments so that each individual score can be traced back to its precise event in the recorded therapy session.

While ELAN is designed for manual annotation and provides an excellent interface to each individual recording, it does not offer an overview of large sets of recordings. In DREAM, the 2017-2018 evaluation alone comprise more than 4000 recorded therapy sessions, with a total of almost 30000 individual performances. While we need the traceability and detailed visualization offered by ELAN, manual handling of each video is not feasible. This challenge is further increased by that the total storage required for video recordings approaches 100 TB.

The original process for storing and extracting user data from the DREAM system is visualized in Figure 2. Here, the *relevant* performances as assessed at runtime are stored in a *User model* and can later be extracted from using a utility application *UserDataExport*, first presented in D3.4.3. The resulting statistics are represented as a spreadsheet with user performances for each session.

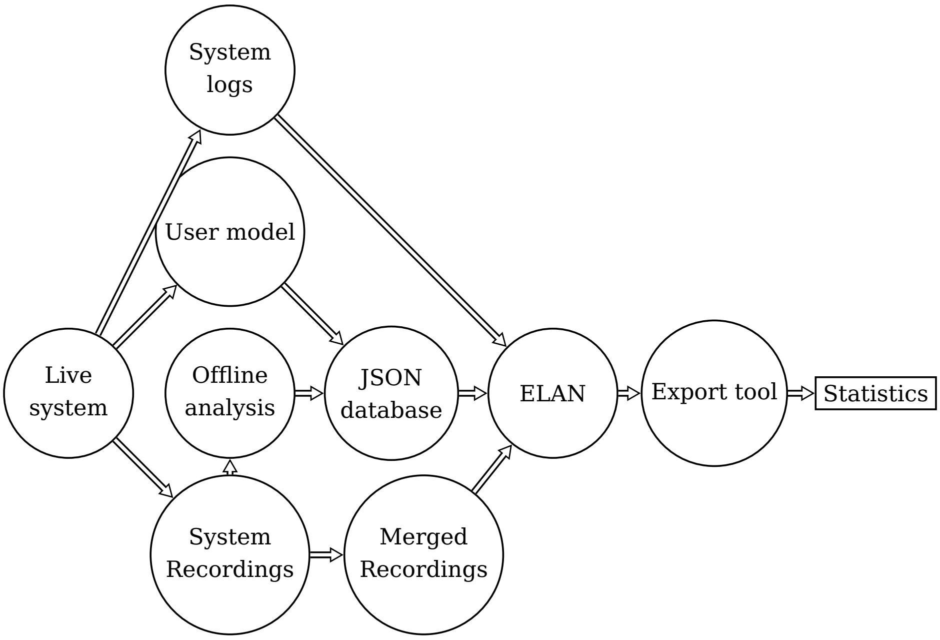
While simple, the process outlines in Figure 2 does not allow developers to go back and adjust assessments retrospectively. The exact time and cause for each individual assessment is not accessible. This developed into a serious problem when the assessment guidelines had to be adjusted for unexpected events during therapy.

The new, traceable, process for extracting user data is visualized in Figure 3. This process is agreeably much more complex, which is in several ways negative. However, this process does not prescribe the *interpretation* of the user's actions. Here, the system logs and video recordings provides access to most of the raw data allowing many aspects of the user's action to be traced back to the original events. Specifically, the system logs refer to text-based log files of port communication appearing between system components at run-time. Each row in the log file specifies the receiving port, a time-stamp with millisecond precision, and a text-representation of the message content. All log files were generated using the *sensoryInterpretationLogger* utility component, specified in D3.4.2.

The system recordings refer to the full recordings made with the *sensoryAnalysis* sys-

[5] Specification of the ELAN Annotation Format (EAF) is found at `https://tla.mpi.nl/tla-news/documentation-of-eaf-elan-annotation-format/`

Figure 3: Visualization of the traceable assessment process. This assessment process integrates information from the *User model* with data from system *logs* and *recordings*. ELAN video annotation tool works as an integration platform for all three data sources, combined with visualization. An updated version of the UserDataExport tool allows extraction of data from ELAN to overall statistics.



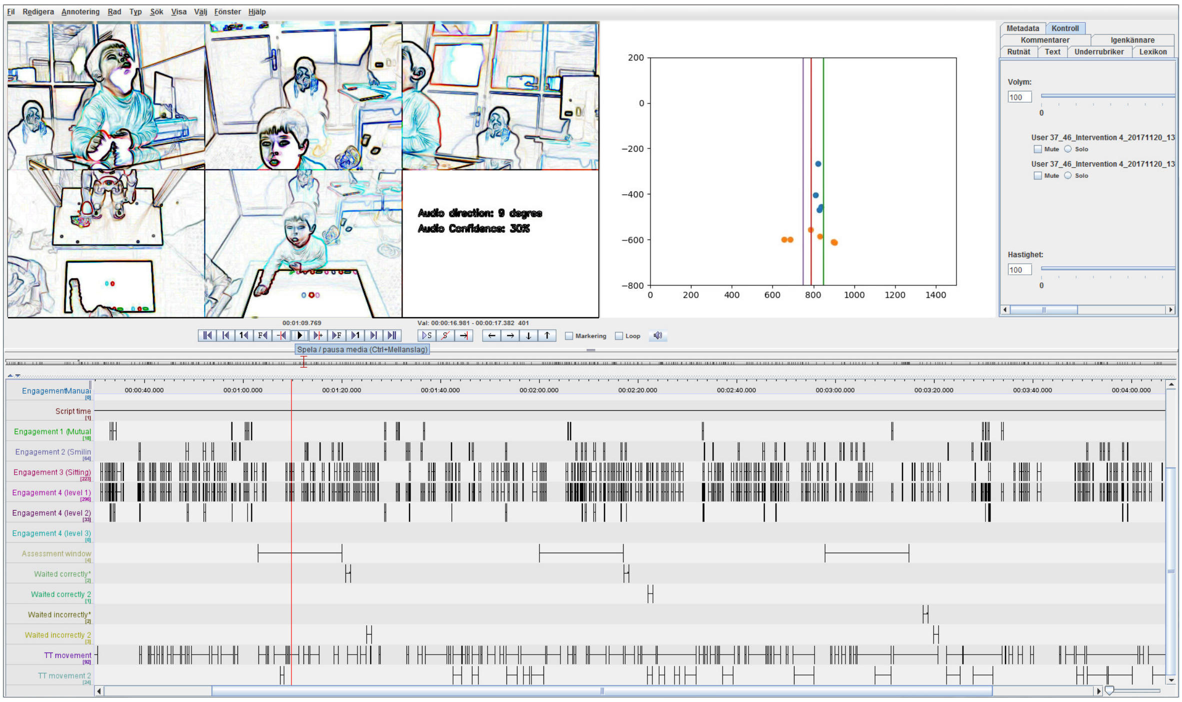


Figure 4: Example screen shoot from the traceable analysis using ELAN annotation tool. All tiers (i.e., rows) visible under the video display contain automatically generated annotations representing various aspects of the assessment process. The standard video is combined with a side-view of the Kinect skeleton data, visualized as a scatter plot with blue and orange markers for the joints. *Note that the video has been filtered in this figure in order to preserve the integrity of the child.*

tem component, specified in D4.1 and evaluated in D4.2. In addition to the live analysis of sensory data, which could be seen as the component's main task, it also stores raw data to disk. This comprise five streams of RGB video, two streams of depth data, Kinect skeleton information and one audio stream, all with a temporal resolution of 25 Hz. All data is synchronized in a frame basis allowing the combined recordings to be re-played and reanalyzed using the sibling component *sensoryAnalysisOffline* presented in D3.4.3.

The result of the offline analysis constitute the open DREAM dataset described in Section 2. This dataset is visible in Figure 3 as the *JSON database*.

As mentioned earlier, the full system recordings require a lot of storage. A single session takes between 10 and 100 GB, making it difficult to store and share large sets of recordings. For this reason, each recording was merged into a single, lower resolution, video file, containing sub-plots for each of the five RGB streams. The reduced image quality made these videos less useful for automatic sensory analysis, but they serve well for manual inspection. An example of one merged video inside the ELAN user interface is visible in Figure 4.

In sum, the use of ELAN and automatically generated annotations allow detailed analysis of the system's behavior at each point in time, in every recording. At the same time, the updated export tool, illustrated as the link from ELAN to Statistics in Figure 3, provides an overview of all 4000 recordings made within the scope of the 2017-2018 clinical evaluation.

4 Integration meetings

During period five, two physical integration meetings took place. September 20-21, 2018, partners from PLYM and HIS met in Skövde. The meeting focused on integration between WP3 and WP5, concerning the performance evaluation of the system. This meeting contributed both to the traceable assessment process described in Section 3 and to some of the results presented in D5.3.

A larger integration meeting was held at at the SoftBank Robotic's office in Paris, December 3-4, 2018. The meeting engaged HIS, PLYM, PORT, UBB, VUB, and SBR. Unfortunately, DMU could not attend this meeting. The meeting was planned directly after the extension of the project was confirmed, with the main target of coordinating the work during reminder of the the project.

5 Utilities

During the course of the whole DREAM project, eleven tools and utility applications have been developed for various purposes. The complete set of utilities in DREAM is listed below:

- componentChecker (D3.4.2)
- DREAM Boxology (D3.4.3)
- ELAN Export (new)

- JSON Export (new)
- Script Generator (D3.4.3)
- sensoryAnalysisOffline (D3.4.3/D4.3)
- sensoryInterpretationLogger (D3.4.2)
- User Data Export (D3.4.3, updated in D3.4.4)
- User Model Creator (D3.4.3)
- Video Merge (D3.4.5)
- YARP Component Generator (D3.4.2)

Two new utilities, *JSON Export* and *ELAN Export*, were introduced during this period. JSON export is effectively implementing the open DREAM dataset format, presented in Section 2 and Appendix A. ELAN Export was first developed as part of *User Data Export* (D3.4.4) but during this period separated into its own application. It takes information from system logs, recordings, and the new JSON database, and represents relevant information as annotated video (eaf) files. Statistics about childrens' performances is thereafter extracted from eaf files using User Data Export. The complete process is described in detail in Section 3 and visualized in Figure 3.

Finally, the utility application *Video Merge* was first developed during 2015 but never included in any deliverable. The purpose of the utility was to transform the large system recordings into a smaller format more easily managed during manual annotation of video. It has been updated on several occasions since then and can today automate the process of merging large sets of system recordings, including correct naming conventions based on the metadata. The tool has been a key component of the traceable analysis presented in Section 3 and is also visualized as the transformation from *System recordings* to *Merged recordings* in Figure 3.

Using *Video Merge*, the size of the complete set of system recordings is reduced from 100 TB to 0.5 TB. The reduced dataset includes merged versions of all five RGB video streams, but excludes RGBD data.

Appendix A Dataset specification

This is a JSON schema for the open DREAM dataset presented in Section 2. The schema specifies the format for a single therapy session with one child under therapy, including definitions of all mandatory attributes of the data. A JSON database file may however comprise additional attributes not defined here. The complete dataset comprise a large set of these sessions.

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id": "http://dream2020.eu/data/specification/dream.1.0.json",
  "title": "DREAM dataset format specification",
  "description": "This is a JSON schema for the DREAM dataset, an outcome of a European research project DREAM: Development of Robot-Enhanced therapy for children with AutisM spectrum disorders. More info at http://dream2020.eu.",
  "type": "object",
  "properties": {
    "date": {"type":"string","format":"date-time"},
    "ados": {
      "type": "object",
      "properties": {"initial": {"type": "number"}}
    },
    "condition": {"type": "string","pattern": "^(RET|SHT)$"},
    "task": {"type": "string"},
    "participant": {
      "type": "object",
      "properties": {
        "id": {"type": "integer"},
        "gender": {"type": "string","pattern": "^(male|female)$"},
        "age": {"type": "number","minimum": 0}
      },
      "required": ["id"]
    },
    "frame_rate": {"type": "number"},
    "eye_gaze": {"$ref": "#/definitions/rot"},
    "head_gaze": {"$ref": "#/definitions/rot"},
    "skeleton": {
      "type": "object",
      "properties": {
        "elbow_left": {"$ref": "#/definitions/pos"},
        "elbow_right": {"$ref": "#/definitions/pos"},
        "hand_left": {"$ref": "#/definitions/pos"},
        "hand_right": {"$ref": "#/definitions/pos"},
        "head": {"$ref": "#/definitions/pos"},
        "sholder_center": {"$ref": "#/definitions/pos"},
        "sholder_left": {"$ref": "#/definitions/pos"},
        "sholder_right": {"$ref": "#/definitions/pos"},
        "wrist_left": {"$ref": "#/definitions/pos"},
        "wrist_right": {"$ref": "#/definitions/pos"}
```

```
        },
        "required": [
          "elbow_left","elbow_right","hand_left","hand_right","head","sholder_center","sholder_left","sholder_right","wrist_left","wrist_right"
        ]
      }
    },
    "required": ["eye_gaze","head_gaze","skeleton","participant","ados","condition","task","frame_rate"],
    "definitions": {
      "pos": {
        "$id": "#pos",
        "description": "Absolute position in Cartesian space.",
        "type": "object",
        "properties": {
          "x": {"type": "array","items": {"type": ["number","null"]}},
          "y": {"type": "array","items": {"type": ["number","null"]}},
          "z": {"type": "array","items": {"type": ["number","null"]}},
          "confidence": {"type": "array","items": {"type": ["number","null"]}}
        },
        "required": ["x","y","z","confidence"]
      },
      "rot": {
        "$id": "#rot",
        "description": "Cartesian heading vector.",
        "type": "object",
        "properties": {
          "rx": {"type": "array","items": {"type": ["number","null"],"minimum": -1,"maximum": 1}},
          "ry": {"type": "array","items": {"type": ["number","null"],"minimum": -1,"maximum": 1}},
          "rz": {"type": "array","items": {"type": ["number","null"],"minimum": -1,"maximum": 1}}
        },
        "required": ["rx","ry","rz"]
      }
    }
  }
```

References

Cai, H., Fang, Y., Ju, Z., Costescu, C., David, D., Billing, E. A., Ziemke, T., Thill, S., Belpaeme, T., Vanderborght, B., Vernon, D., Richardson, K., and Liu, H. (2018). Sensing-enhanced Therapy System for Assessing Children with Autism Spectrum Disorders: A Feasibility Study. *IEEE Sensors Journal*.

Hernández García, D., Esteban, P. G., Lee, H. R., Romeo, M., Senft, E., and Billing, E. A. (2019). Social Robots in Therapy and Care. In *Workshop at the ACM/IEEE International Conference on Human Robot Interaction (HRI)*, Daegu, South Korea.

Lemaignan, S., Edmunds, C. E. R., Senft, E., and Belpaeme, T. (2018). The PInSoRo dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PLOS ONE*, 13(10):e0205999.