DyadicHead database

Mike Brookes

22nd January 2020

This document discusses the dataset that is described in

[1] L. V. Hadley, W. O. Brimijoin, and W. M. Whitmer. Speech, movement, and gaze behaviours during dyadic conversation in noise. Scientific Reports, 9 (1), jul 2019. doi: 10.1038/s41598-019-46416-0

which tracks head and eye movements during dyadic conversations.

1 Format of database

Each of 15 talker pairs (dyads) recorded 3 conversations and data from each conversation is saved in a file whose name has the form "dyad01conv1.dyh". There are a total of only 44 files because conversation 3 from dyad 2 did not record properly and is omitted. The .dyh file consists of 21N+13 little-endian 16-bit integers where N is the number of frames. The 13-value header consists of the characters 'DyadicHead' followed by the File-version, Dyad and Conversation numbers. Of the 38 fields in the output of readdyh.m (see below), only the first 21 are saved explicitly in the database, the remaining 17 fields are calculated when the file is read.

2 MATLAB routines

The following MATLAB routines are available

readdyh This function reads the files from the database and calculates additional useful quantities

printdyad This function allows the user to select a conversation and frame and prints out the values of all the variables

plotdyad This function allows the user to select a conversation and plot one or more variables versus time over a chosen range of samples

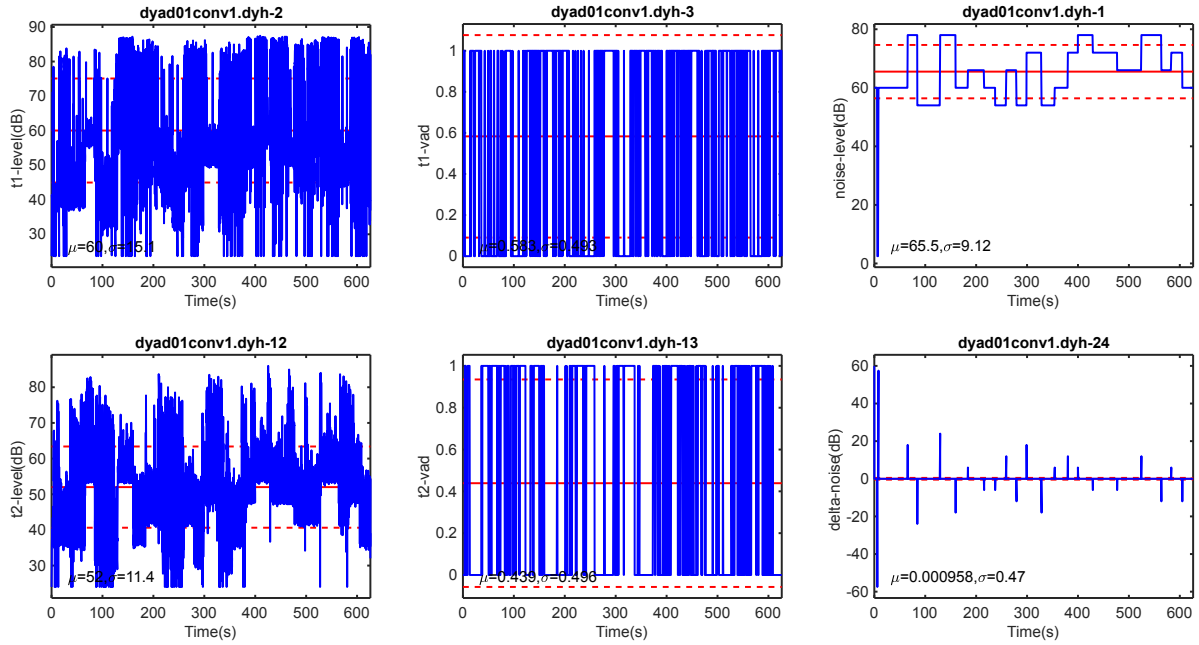
writedyh This function was used to read the raw text files available in [1] as supplementary informationand writes out binary files in a format that is more convenient to access

3 Experimental Setup and Coordinate System

The setup is shown in Fig. 1 of [1] and consists of two seated talkers facing each other at a distance of 1.5 m; a ring of eight loudspeakers provided speech-shaped background noise. From the viewpoint of talker-1, the positive x axis is to the right, y ahead and z upwards; the coordinate % origin is on the floor midway between the talkers. Head and gaze rotations are performed in the order yaw, pitch, roll using intrinsic coordinates that rotate with the head. A positive yaw turns to the right, a positive pitch tilts downwards and a positive roll tilts to the left.

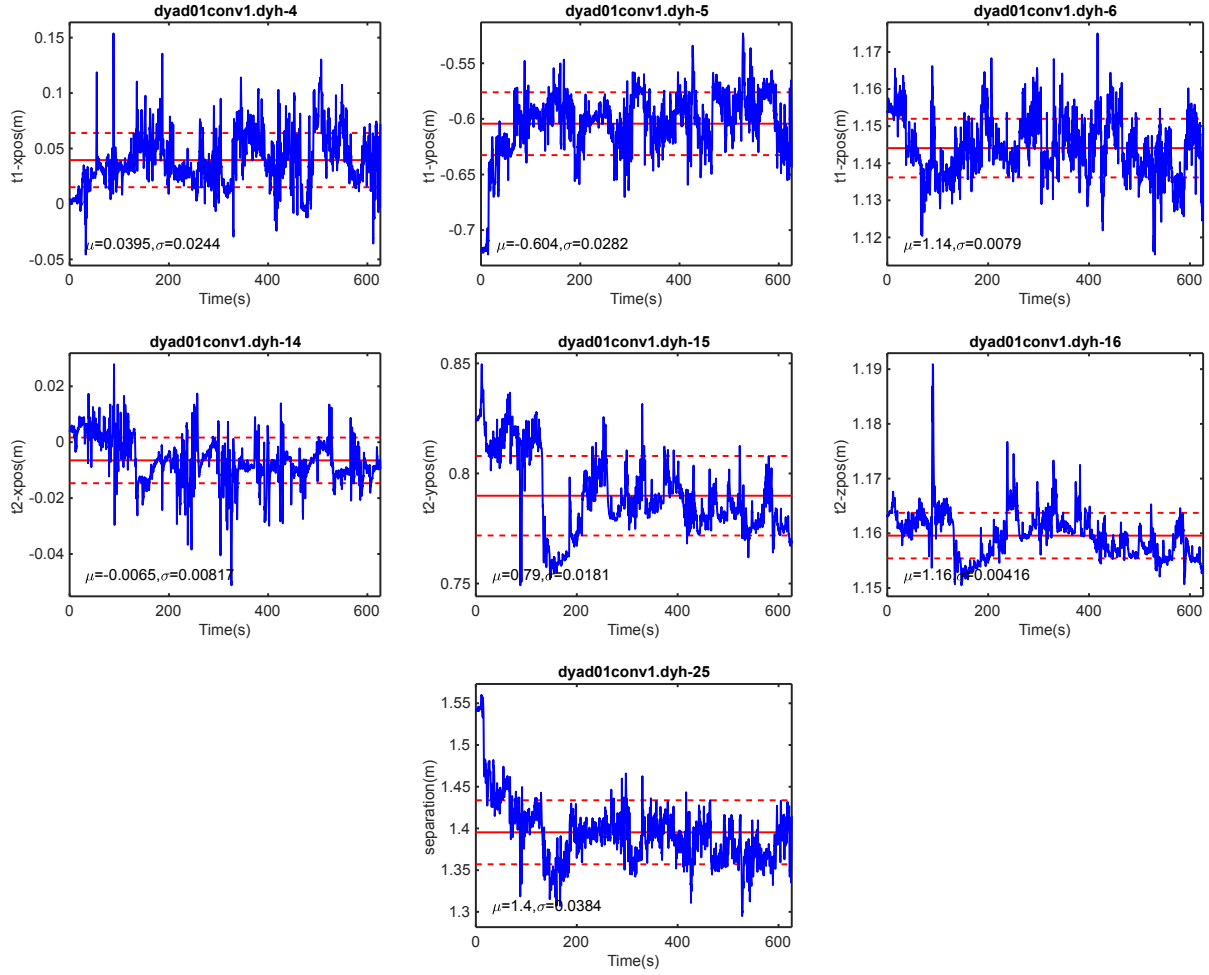
4 Speech and Noise

The plots below show data from the file dyad1conv1.dyh; this conversation lasts 10.44 minutes and contains 62625 samples @ 100 Hz sampling rate. The title of each plot gives the file name followed by the field number (in the range 1 to 38); fields 1 to 21 are extracted directly from the file while fields 22 to 38 are calculated from these values when the file is read. For some fields, some of the values are invalid (given as NaN in the files); the number of such values is indicated on the plot when it is non-zero. The mean ±standard deviation of the valid values are indicated by horizontal lines with numerical values given in the lower left corner. In most cases, the first row of each set of plots corresponds to talker-1 and the second row to talker-2.



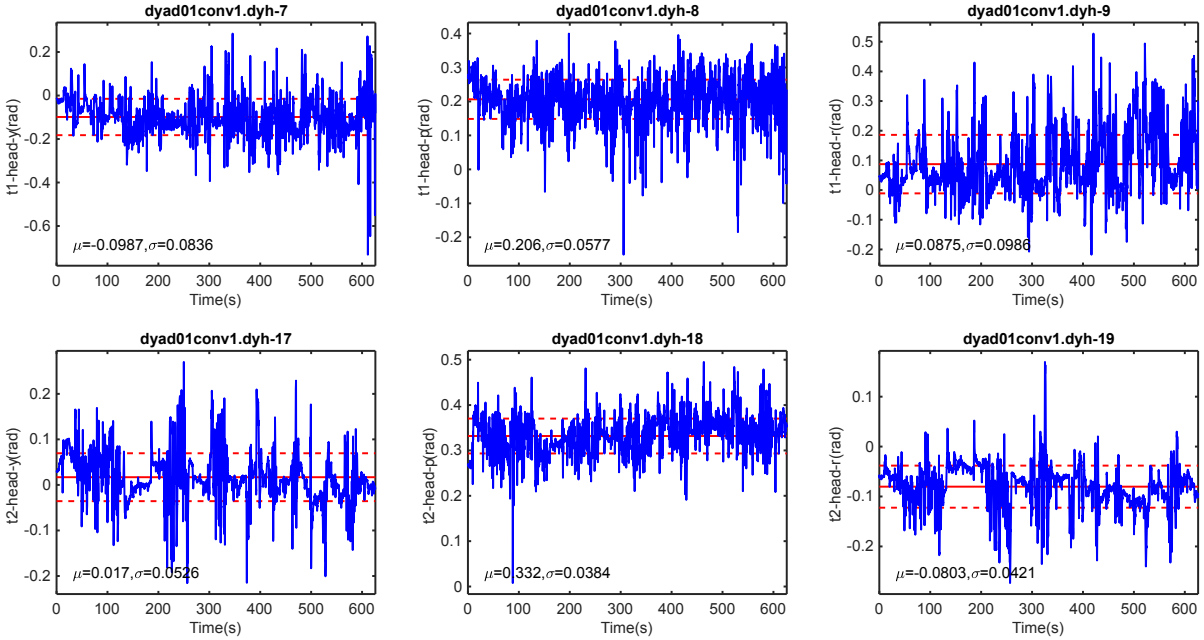
The first column shows the of each talker during speech; the upper and lower rows are talker-1 and talker-2 respectively. I assume that the units are dB SPL at the other speaker because this is the most relevant thing to measure. The second column and an indication of when they are talking (58% and 44% of the time respectively). The upper right plot shows the noise level over time (again presumably in dB SPL). There is an anomalous interval near the beginning where the level drops to 2.5 dB. The lower right plot is the change in noise level; non-zero values indicate the start of a new noise segment.

5 Head displacement

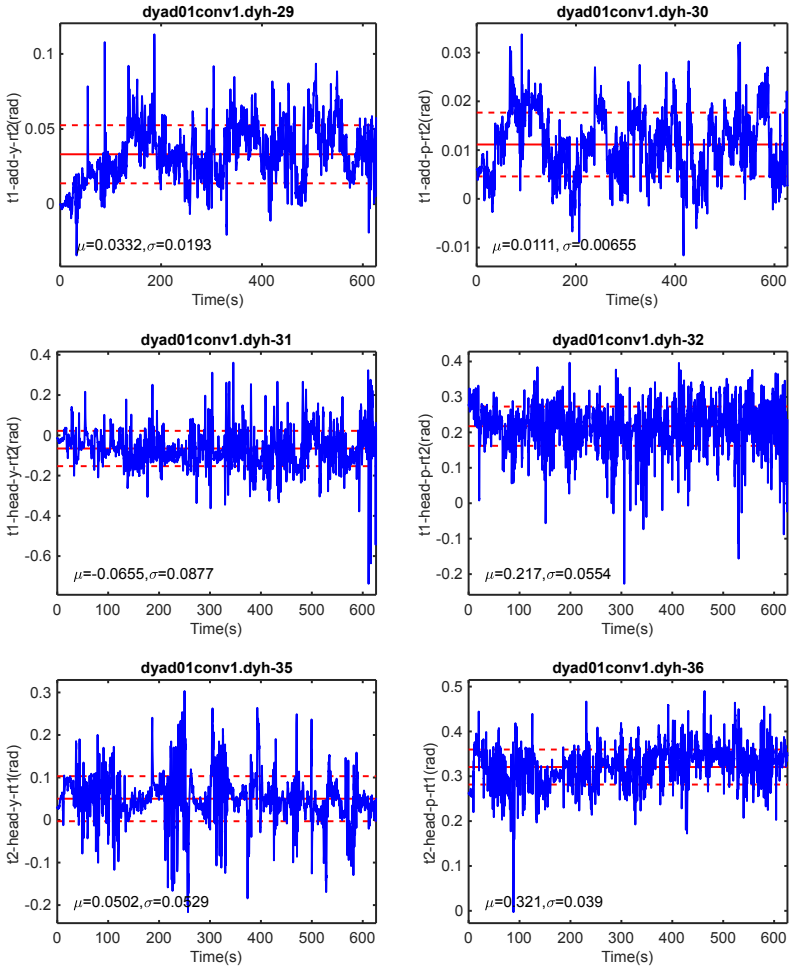


These plots show the (x,y,z) positions of talker-1 (top row) and talker-2 (middle row) as well as the distance between the two talkers. The centre of the head is defined as "between the ears in line with the bridge of the nose" [1, page 6, line 9].

6 Head Orientation

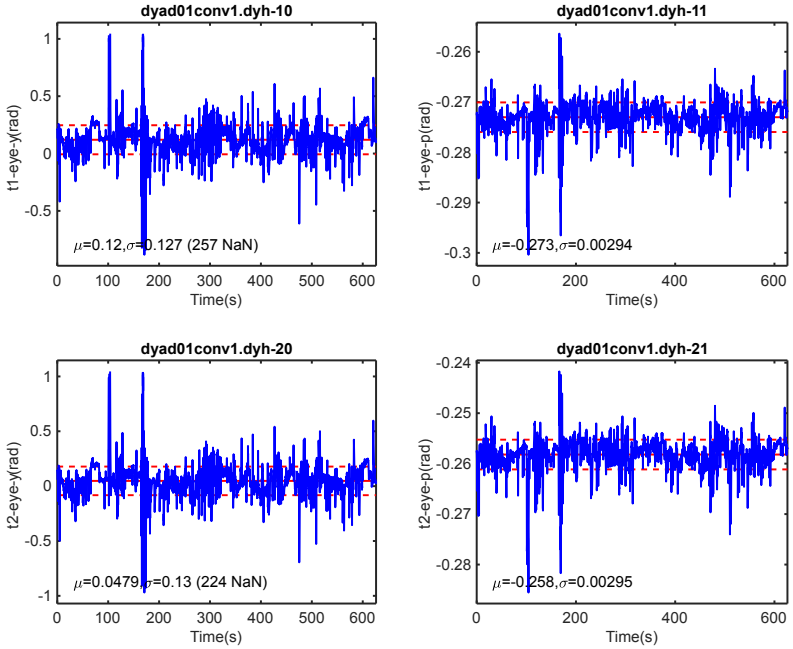


These plots show the head orientation using Tait-Bryan angles (yaw,pitch,roll) in degrees.. I assume that the rotation order is (yaw,pitch,roll) and that these are intrinsic rotations (i.e. they are relative to axes that move with the head). It appears that, looking along ae positive axis, a positive rotation is anti-clockwise.

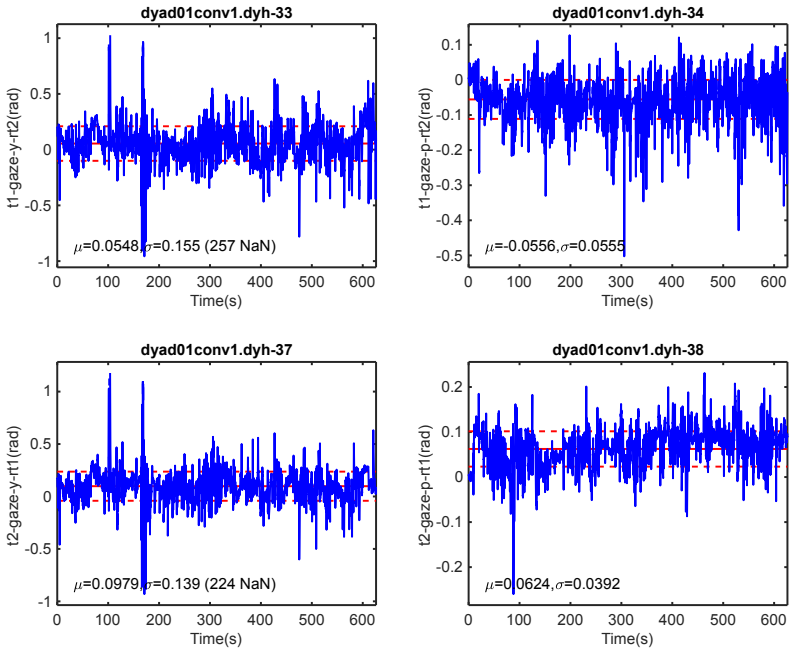


The top row shows how much must be added to the yaw and pitch to make the head rotation of talker-1 be relative to the centre of the talker-2's head rather than to the y-axis (the sign of the pitch offset is reversed for talker-2). The next two rows show the resultant relative head rotation for talker-1 and talker-2 respectively.

7 Eye Tracker and Gaze



These plots show the output of the eye tracker for talker-1 (upper plots) and talker-2 (lower plots). Apart from offsets of around 4°and 1° in ϕ and θ respectively, the plots for the two speakers look almost identical; this is unbelievable and so I presume there is an error in the data somehow. In addition, both files contain many invalid values in the regions near 100 s and 168 s where the large spikes occur in the graphs.



It is possible to add the eye-tracker information to the relative head orientation to find the gaze direction relative to the other talker. These plots suffer the same believability issues as the previous ones. The pitch gaze is dominated by head rotation since the eye movement is very small. The gaze values in [1, page 4, Figure 4b] use the ear tragi as the reference height according to [1, page 3, line 10] rather than the head centre defined earlier. It is unclear how the offset between the head centre and the tragi was calculated.

8 Issues

Here I summarize the issues and uncertainties in the data. Most of these have already been mentioned above.

1. The coordinate system, rotation polarities and rotation order are not precisely defined.
2. The point at which the speech and noise levels are measured is not clearly defined.
3. The eye tracker values are suspiciously identical for talker 1 and talker 2 (data indices 10/11 and 20/21) and contain some invalid values. I think this must be an error.
4. The head displacements define the centre of the head as "between the ears in line with the bridge of the nose" [1, page 6, line 9] whereas the relative gaze is calculated with respect to the tragi. I don't know how they mapped between these two and whether they measured a different vertical offset for each test subject.
5. The noise dips to 2.5 dB between 5.89 and 8.14 seconds. This contradicts the statement in [1, page 5, penultimate line] that the noise levels were taken from $\{54, 60, 66, 72, z, 78\}$ dB. I presume this is an error.