

# EasyDIAG Manual

## Calculating rater agreement

### Version 1.0.0

©Robert Rein, Henning Holle

[gerrobrein@yahoo.com.au](mailto:gerrobrein@yahoo.com.au)

[h.holle@hull.ac.uk](mailto:h.holle@hull.ac.uk)

April Cologne 2014

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.



## Content

<b>Introduction .....</b>	<b>3</b>
<b>Installing EasyDIAG using the source files.....</b>	<b>4</b>
<b>Installing EasyDIAG using the provided binary files.....</b>	<b>4</b>
<b>Exporting from ELAN.....</b>	<b>5</b>
<b>Using EasyDIAG.....</b>	<b>7</b>
<b>Analyzing rater agreement.....</b>	<b>7</b>
Individual scores.....	8
Investigating global agreement .....	10
Setting overlap.....	12
<b>Analyzing problems.....</b>	<b>13</b>
Agreement plot.....	13
Rater disparity .....	14
Annotation lengths.....	15
<b>Exporting the results.....</b>	<b>15</b>
<b>References.....</b>	<b>16</b>

## Introduction

The EasyDIAG Toolbox provides routines to calculate the rater agreement for annotations of video data made by two raters. The routine takes into account the overlap and type of units at the same time and calculates kappa values for global agreement as well as kappa values for individual types.

The EasyDIAG toolbox has been developed using annotation data from the ELAN annotation tool (<http://www.lat-mpi.eu/tools/elan>). If you use other software to produce annotation data, as the needed file is a text file it should be possible to produce the desired result using a text editor or Microsoft Excel®-likes.

When you use this software please cite the following article:

Holle and Rein, in preparation

The EasyDIAG toolbox is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

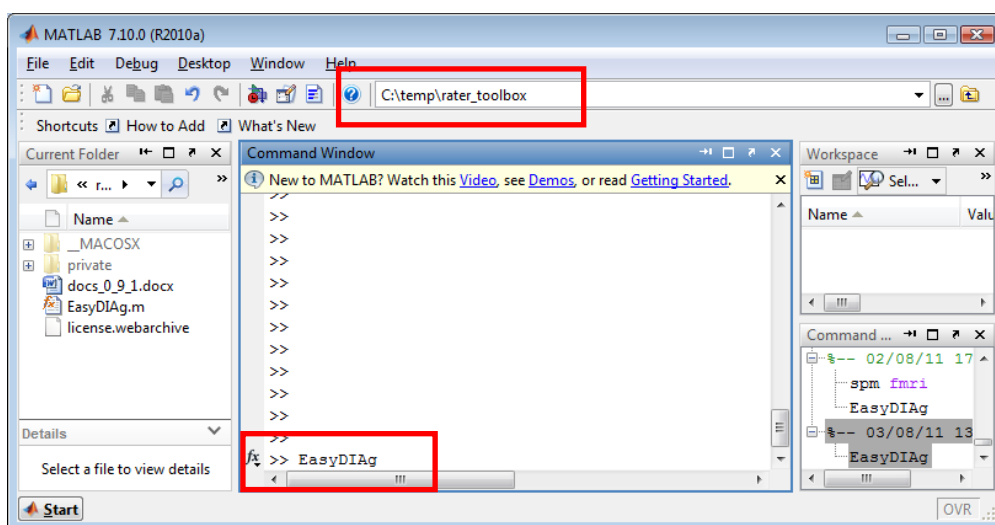
You should have received a copy of the GNU General Public License along with this program. If not, see <<http://www.gnu.org/licenses/>>.

If you find a bug or you think that the toolbox produces erroneous results please either send a report to [gerrobrein@yahoo.com.au](mailto:gerrobrein@yahoo.com.au) or file a ticket at Sourceforge, describing your MATLAB version, the error which occurred (screenshot or copy-and-paste the error into the mail) and including maybe the first few lines of your data.

### Installing EasyDIAG using the source files

If you have a working installation of MATLAB™ on your computer, you can directly download the source code and run it within MATLAB™. The toolbox has been developed on MATLAB™ R2009a and has been successfully tested on any version since. The latest version of the toolbox is always available on Sourceforge (<http://sourceforge.net/projects/easydiag/>). Just browse to the Code section and click on "Download Snapshot".

Create a new folder and extract the provided zip file into the folder. The folder contains all necessary files for EasyDIAG.m to work. You do not need to add the folder to your working directory. Change the working directory in MATLAB to the newly created folder. Type EasyDIAG into the command window of MATLAB and execute the command by pressing ENTER.



**Figure 1: Changing the working directory in Matlab™. Make sure, that the working directory is the one containing the EasyDIAG.m file (see top red box). Next, type EasyDIAG into the command window (see bottom red box) and execute the command by pressing 'ENTER'.**

### Installing EasyDIAG using the provided binary files

If you don't have MATLAB™ license, you can run the toolbox by downloading a binary file. To download the binary, go to (<http://sourceforge.net/projects/easydiag/>), click on the 'Files' section and download the appropriate version for your computer. You will have to download the MATLAB Compiler Runtime (MCR) from the Mathworks site. Please see the readme.txt accompanying the binary.

## Using EasyDIAG

When using the EasyDIAG binary all you have to do after installing the MCR is start the binary.

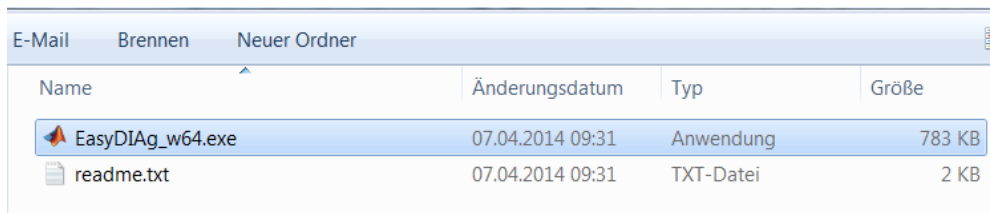


Figure 2: EasyDIAG binary in Windows explorer window

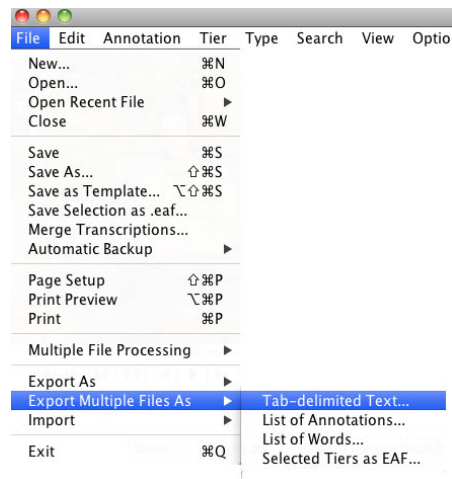
## Exporting from ELAN

In order for EasyDIAG to work, you need to create a text file containing your annotations. Here, we show how such a text file can be created automatically using the export function of ELAN (Version used here: 4.1.1).



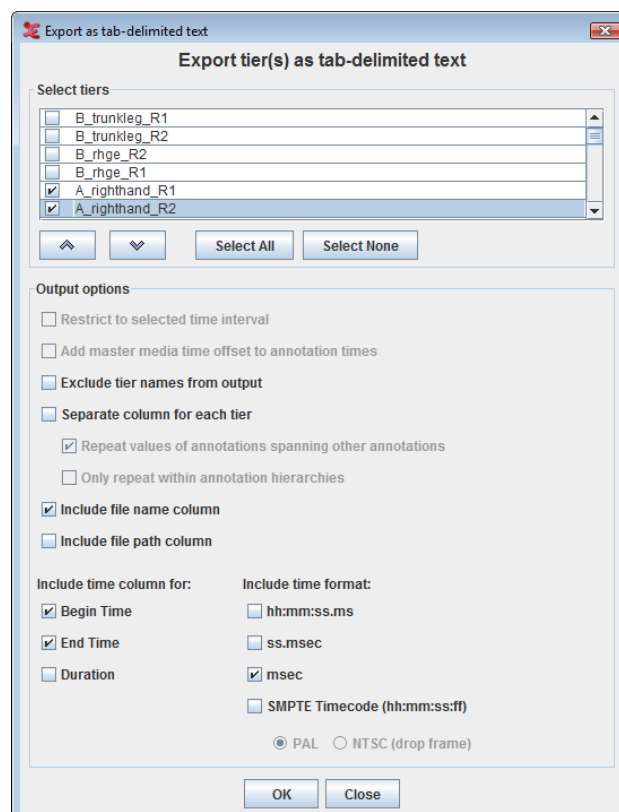
Figure 3: ELAN version

Within ELAN, select the Export Multiple Files As → Tab-delimited Text... menu entry (even if you work with a single file only). Please consult the ELAN-documentation for further details about exporting tab-delimited files.



**Figure 4: Exporting as tab-delimited file**

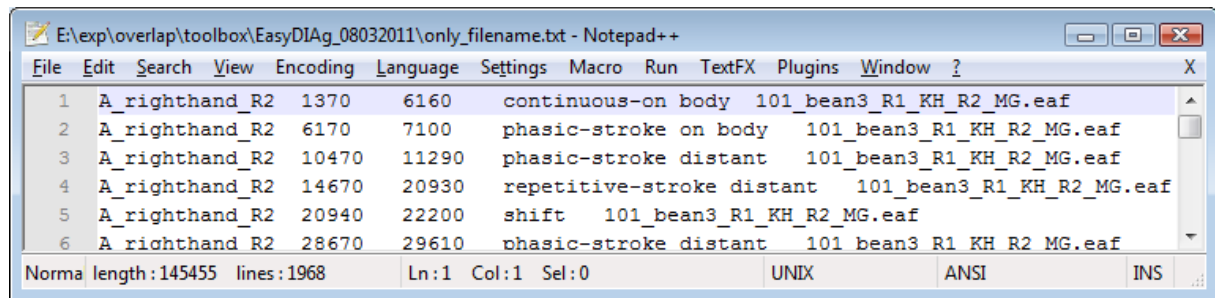
Mark the tiers you want to export from ELAN. Have a look at the output options below. Make sure the output options are exactly as shown in Figure 3, otherwise EasyDIAG will not be able to properly load the file. Only **'include file name column'**, **'Begin time'**, **'End time'** and **'msec'** should be **ticked**, everything else should be **unticked** (see also Figure 3).



**Figure 5: Specifying tiers and output options (in ELAN 4.1.1). Make sure you select the Output options exactly as shown here.**

## Using EasyDIAG

The EasyDIAG toolbox expects a tab-delimited text file where the **first column specifies the tier**, the **second and third columns the beginning and the end of the annotation (in msec)**, the **fourth column the annotation category**, and the last **fifth column the file name** (see Figure 4). Annotations from two different raters for one media file should also be contained in the same file as the filename is used to pair up ratings.



1	A_righthand_R2	1370	6160	continuous-on body	101_bean3_R1_KH_R2_MG.eaf
2	A_righthand_R2	6170	7100	phasic-stroke on body	101_bean3_R1_KH_R2_MG.eaf
3	A_righthand_R2	10470	11290	phasic-stroke distant	101_bean3_R1_KH_R2_MG.eaf
4	A_righthand_R2	14670	20930	repetitive-stroke distant	101_bean3_R1_KH_R2_MG.eaf
5	A_righthand_R2	20940	22200	shift	101_bean3_R1_KH_R2_MG.eaf
6	A_righthand_R2	28670	29610	phasic-stroke distant	101_bean3_R1_KH_R2_MG.eaf

Figure 6: Example of an exported tab-delimited file

**Tier names (in column 1)** must contain the indicators **R1** for the first rater and **R2** for the second rater! If these indicators are not contained in the tier names the toolbox is not able to determine which annotations belong to which annotator. The order is not important but should be consistent throughout the files which ELAN takes care of.

If you use another naming convention you can easily produce an appropriate file by doing a 'search & replace' in the exported text file, using your favorite text editor (e.g. Notepad++, <http://notepad-plus-plus.org/>) or Microsoft Excel®.

**Attention!!!** The current version of ELAN (v4.6.2) has a bug where after the tiername two tabs (instead of one) are included. This triggered an error in earlier version of EasyDIAG but should be fixed in the current. If you encounter any problems, please use a text editor or Excel to remove the superfluous tab. The bug has been filed with the ELAN maintainers and hopefully will be fixed in the next version.

## Using EasyDIAG

### Analyzing rater agreement

Once EasyDIAG is started, an empty EasyDIAG window will pop up.

## Using EasyDIAG

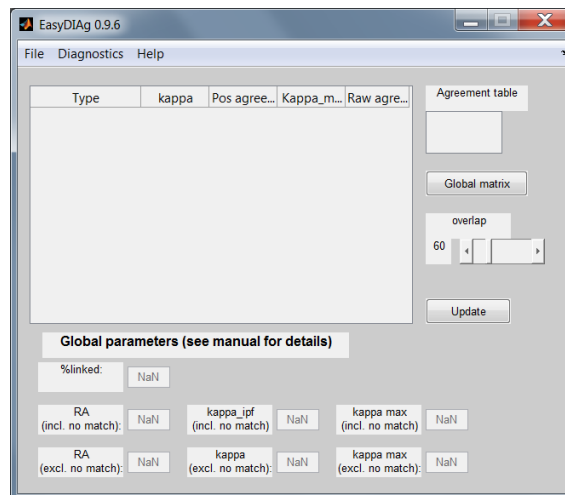


Figure 7: Empty EasyDIAG window

To import a file select '**File => Import ELAN file**' from the menu bar. A standard file dialog will appear allowing you to specify the file you want to load into EasyDIAG. Select the text file that you exported from ELAN.



Figure 8: Opening a file

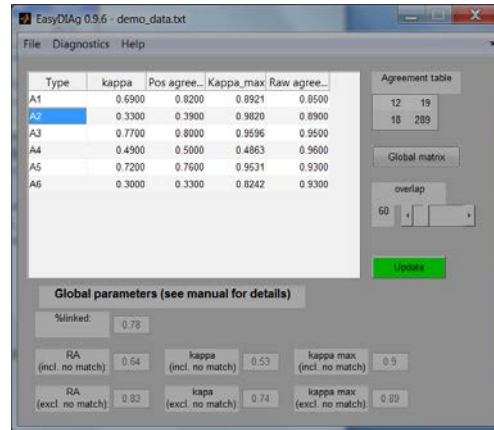
EasyDIAG first checks the ELAN file and tests whether the file adheres to the expected format. In case an error message is displayed, please read the message and check whether the exported ELAN file follows the format describe above. Also check whether the file contains any additional errors.

EasyDIAG also prints some quick facts about the file on the command line. If the export file is correct, the interface automatically calculates several rater agreement scores and shows the results in the graphical user interface.

### Individual scores

The table shows several agreement scores calculated for each individual annotation type (compare Figure 9).





**Figure 9: Agreement scores for individual units**

The first column specifies the annotation type name, the second column indicates the calculated Cohen's kappa measure (Cohen, 1960). Columns 3 to 5 indicate the positive agreement according to Cicchetti and Feinstein (1990), maximum kappa value (Cohen, 1960), and the raw agreement accordingly.

Cohen's kappa is calculated according to:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where  $p_0$  = the proportion of agreed upon units,  $p_e$  = proportion agreement expected by chance.

Positive agreement is calculated according to:

$$p_{pos} = \frac{2a}{f_1 + g_1}$$

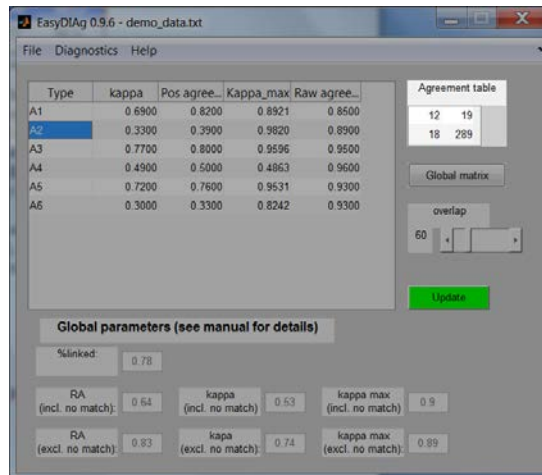
where  $a$  = top-left entry in agreement table,  $f_1$  = total of first column of agreement table,  $g_1$  = total of first row of agreement table.

$\kappa_{max}$  is calculated according to:

$$\kappa_{max} = \frac{p_{0M} - p_e}{1 - p_e}$$

$$p_{oM} = \sum_{i=1}^K \min(p_{i\bullet}, p_{\bullet i})$$

EasyDIAG allows to select a row in the table and the interface displays the calculated agreement table (compare Figure 10).

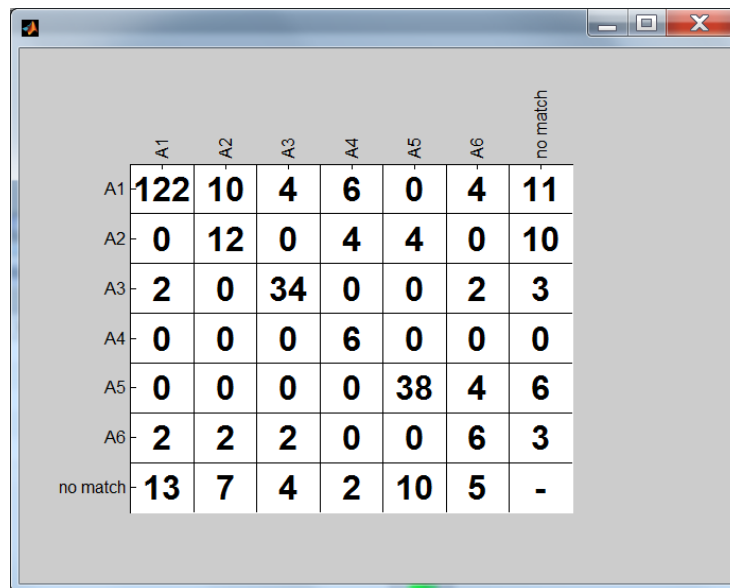


**Figure 10: Agreement table**

The agreement table contains information about agreement and disagreement between the raters. The top left number represents the number of agreements between raters (top left). The top right and bottom left numbers indicated those units where only one of the raters identified the unit. Finally, the number in the bottom corner indicates the number of occurrences when both raters identified a unit different from the specified unit. The agreement table represents the basis for the calculation of the Cohen's kappa value for individual annotation units.

### Investigating global agreement

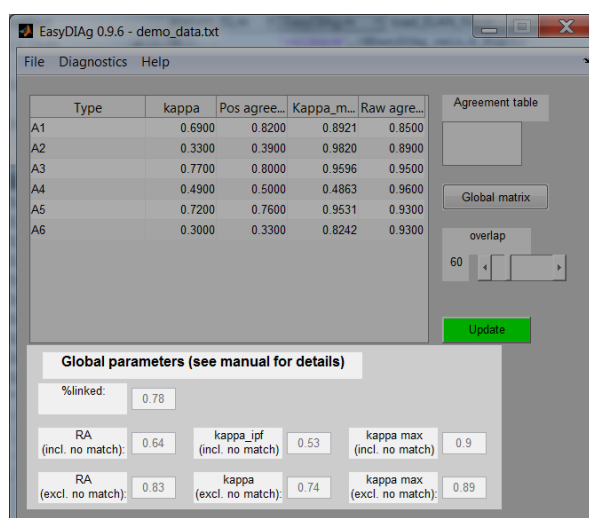
In addition to agreement scores at the annotation unit level, EasyDIaG allows to investigate the global agreement across all annotation types. To investigate the global agreement table press the "Global matrix" button. Another window will appear depicting the agreement between the different annotation types.



	A1	A2	A3	A4	A5	A6	no match
A1	122	10	4	6	0	4	11
A2	0	12	0	4	4	0	10
A3	2	0	34	0	0	2	3
A4	0	0	0	6	0	0	0
A5	0	0	0	0	38	4	6
A6	2	2	2	0	0	6	3
no match	13	7	4	2	10	5	-

Figure 11: Global agreement matrix

The global agreement matrix allows you to inspect in detail which annotation types have led to disagreements between the raters. Top units belong to Rater 1, left units belong to Rater 2. The annotation units are augmented by a "no match" type to indicate annotation units which have not been linked. In the bottom-right corner a hyphen is displayed as there cannot be a value in this combination of no match versus no match. To account for this structural zeros, kappa values including the no match category are fitted by an iterative proportional fitting algorithm to calculate the expected frequencies needed for the calculation of  $\kappa_{ipf}$  (Bakeman & Robinson, 1994, p.106). Global agreement scores are displayed at the bottom of the EasyDIAG GUI.



Type	kappa	Pos agree...	Kappa_m...	Raw agre...
A1	0.6900	0.8200	0.8921	0.8500
A2	0.3300	0.3900	0.9820	0.8900
A3	0.7700	0.8000	0.9596	0.9500
A4	0.4900	0.5000	0.4863	0.9600
A5	0.7200	0.7600	0.9531	0.9300
A6	0.3000	0.3300	0.8242	0.9300

Global parameters (see manual for details)

%linked: 0.78

RA (incl. no match): 0.64    kappa\_ipf (incl. no match): 0.53    kappa max (incl. no match): 0.9

RA (excl. no match): 0.83    kappa (excl. no match): 0.74    kappa max (excl. no match): 0.89

Figure 12: Global agreement parameters

In total seven different global parameters are displayed.

## Using EasyDIAG

"%linked" describes the proportion of linked annotation units to the total number of annotation units. Thus, this parameter estimates how often units were not matched.

The following six parameter are divided into agreement scores including no match and agreement scores excluding no match. By comparing these two sets of scores one can investigate the influence of categorization error (excluding no match) versus segmentation and categorization error (including no match). The agreement scores, RA (raw agreement),  $\kappa$ , and  $\kappa_{\max}$  are calculated using the same formulas as for individual annotation unit scores. The  $\kappa_{\max}$  value gives you an upper bound for the maximum  $\kappa$  obtainable given the marginal distributions of the agreement table. The most important value for reporting will be  $\kappa_{\text{ipf}}$ .

### Setting overlap

Per default, EasyDIAG uses an overlap of 60 percent to match annotation to each other. If a more strict measure is desired, the value can be freely adjusted from 51% up to 90% using the slider on the right of the interface. After adjustment of the overlap value the update button must be pressed to initiate a recalculation of the agreement scores. EasyDIAG will indicate that the values must be recalculated by coloring the update button red (compare Figure 13).

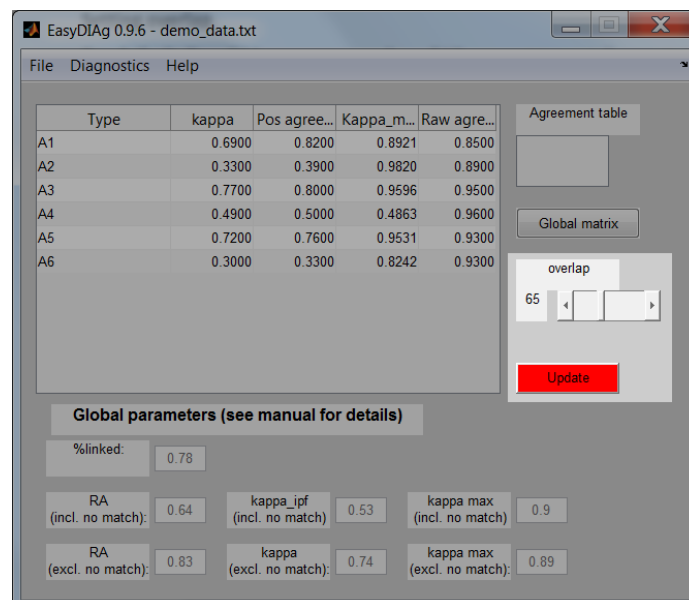


Figure 13: Adjusting overlap value and recalculating agreement scores

## Analyzing problems

EasyDIAG provides graphical user interfaces to better identify problematic annotation units. Currently three different diagnostic diagrams are available through the diagnostics menu.

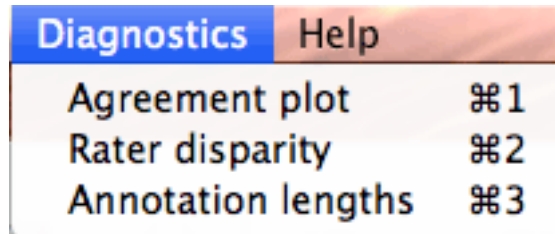
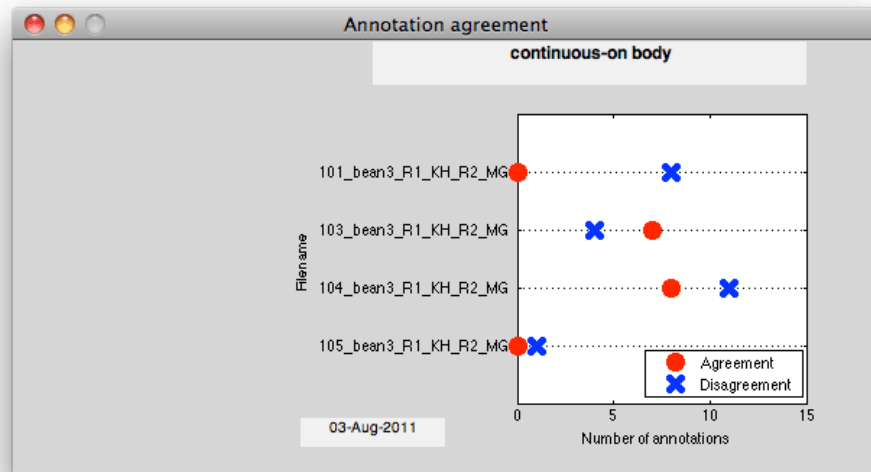


Figure 14: Diagnostics menu

Each of these diagrams shows diagnostic measures for each file individually. This makes it easy to identify problematic files, which may be revisited for reevaluation. For each diagnostic diagram a new window will appear displaying the diagram. Each diagram can be saved to a file for later analysis or storage.

### Agreement plot

The agreement plot shows the difference between the agreement and disagreement for each unit or collapsed over all units for each file. The raters agree when both found the same unit and the two units have the chosen overlap and they disagree if when rater identified the specific unit but the other rater did not. Thus, the values are the upper left value of the agreement table for agreement and the sum of the upper right and lower left value of the agreement for each file. The current value of the overlap is used. Thus, if the overlap value is changed, a new diagram should be produced, as there is no update mechanism.



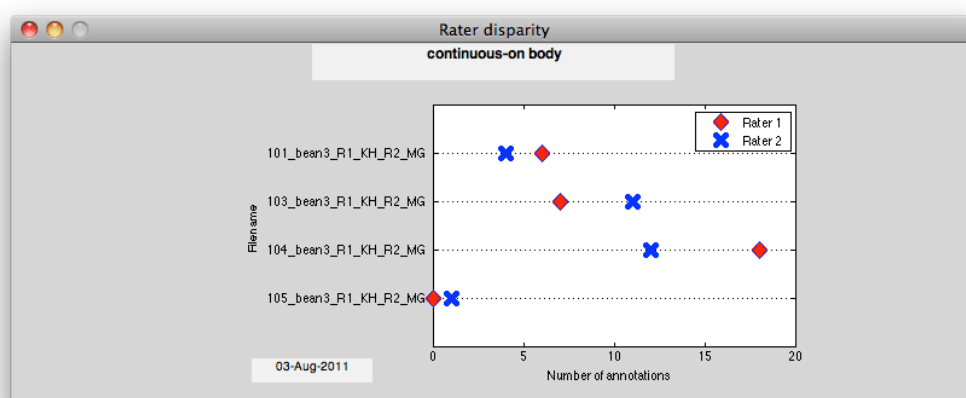
**Figure 15: Agreement plot comparison. Each file uses one row.**

The diagram can be used to browse through the different unit categories and identify problems for specific files, which potentially should be revisited.

The date in the lower left corner of the diagram shows the current date and will be also saved into the file to keep track of when the analysis was done.

### Rater disparity

The rater disparity plot shows the number of annotations made by each rater for each file for each unit category or collapsed over all unit categories.



**Figure 16: Rater disparity**

The rater disparity plot provides a quick way to scan for large differences in the number of annotations between the two raters. No information regarding the timing of the annotations is used. If there are large disparities between the two raters they can be

## Using EasyDIAG

quickly identified using this diagram. Large differences between the raters could indicate that one rater may tend to break up gestures into shorter units whereas the other tends to identify large chunks of gestures as belonging together.

### Annotation lengths

The annotation length plot displays the total duration for each unit type for each rater for each file.

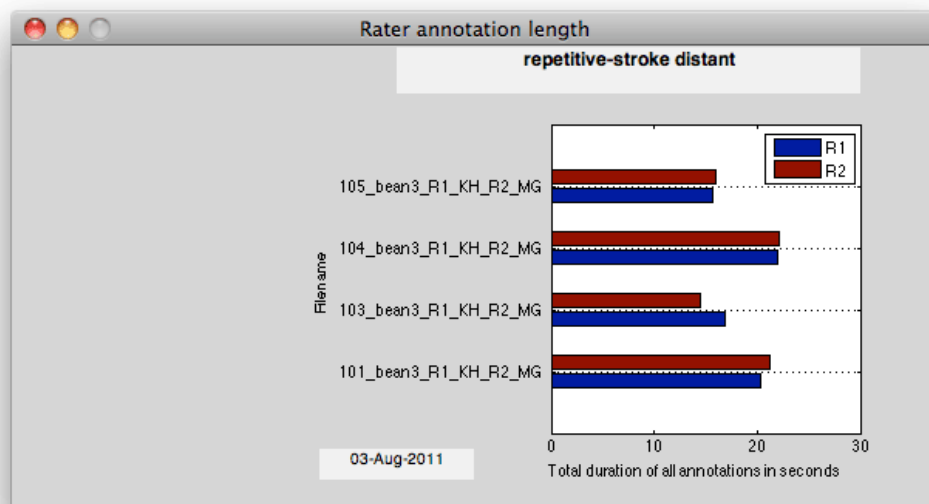


Figure 17: Annotation length for each rater

Differences between the raters could indicate that one rater identifies either multiple units whereas the other rater sees only few annotations or that there are differences for the offsets or onsets of the units.

### Exporting the results

The agreement results produced by EasyDIAG, including global and category-specific agreement indices, can be exported into a text file using the File->Save menu of EasyDIAG.

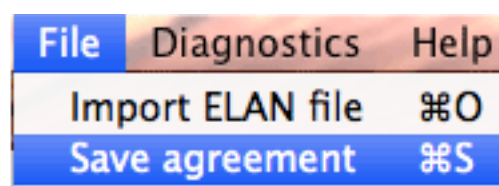


Figure 18: Saving the agreement results

EasyDIAG will produce a text file containing all results and some additional information.

## Using EasyDIAG

```
1 EasyDIAG version: 0.9.6
2 File created: 09-Apr-2014
3
4 ELAN data export filename:
5 D:\eigene_dateien\Dropbox\analysis\interrater\upload\demo_data.txt
6 Contained 338 annotations.
7 Overlap value used = 60 %
8
9 1) Percentage of linked units:
10 linked = 0.78 %
11
12 2) Overall agreement indicies (including no match):
13 Raw agreement = 0.64
14 kappa = 0.53
15 kappa_max = 0.9
16
17 3) Overall agreement indicies (excluding no match):
18 Raw agreement = 0.83
19 kappa = 0.74
20 kappa_max = 0.89
21
22 4) Global agreement matrix:
23
24      A1      122      10      4      6      0      4      11
25      A2      0      12      0      4      4      0      10
26      A3      2      0      34      0      0      2      3
27      A4      0      0      0      6      0      0      0
28      A5      0      0      0      0      38      4      6
29      A6      2      2      2      0      0      6      3
30      no match 13      7      4      2      10      5      0
31
32 5) Individual agreement scores:
33
34      Type      kappa      Pos agreement      Kappa_max      Raw agreement
35      A1      0.69      0.82      0.89      0.85
36      A2      0.33      0.39      0.98      0.89
37      A3      0.77      0.80      0.96      0.95
38      A4      0.49      0.50      0.49      0.96
39      A5      0.72      0.76      0.95      0.93
40      A6      0.30      0.33      0.82      0.93
```

Figure 19: Exported result text file

We hope that you find EasyDIAG useful for your research however we do not guarantee valid results.

## References

- Bakeman, R., & Robinson, Byron F. (1994). *Understanding log-linear analysis with ILOG: an interactive approach*. Hove, UK: Lawrence Erlbaum Associates.
- Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43(6), 551-558. doi: 0895-4356(90)90159-M [pii]
- Cohen, Jacob. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. doi: 10.1177/001316446002000104