# MindBigData the MNIST of Brain Digits v1.01

**Data** · September 2015

DOI: 10.13140/RG.2.1.3312.9441



**1 author:**

David Vivancos
Artificiology.com
**5** PUBLICATIONS **9** CITATIONS

SEE PROFILE

# MindBigData

## The "MNIST" of Brain Digits

The version 1.01 of the open database contains **1.160.193 brain signals** of **2 seconds each**, captured with the stimulus of **seeing a digit (from 0 to 9)** and thinking about it, over the course of almost 2 years between 2014 & 2015, from a single Test Subject David Vivancos.

All the signals have been captured using commercial EEGs (not medical grade), NeuroSky MindWave, Emotiv EPOC, Interaxon Muse & Emotiv Insight, covering a total of **19 Brain (10/20) locations.**

Four files are available for download:

| DataBase | File | Zip size | Uncompressed File size |
|---|---|---|---|
| **MindWave** | MindBigData-MW-v1.0.zip | 62,6 MB (65,663,303 bytes) | 297 MB (311,994,495 bytes) |
| **EPOC** | MindBigData-EP-v1.0.zip | 409 MB (429,732,466 bytes) | 2,66 GB (2,859,712,035 bytes) |
| **Muse** | MindBigData-MU-v1.0.zip | 62,6 MB (65,663,303 bytes) | 297 MB (311,994,495 bytes) |
| **Insight *** | MindBigData-IN-v1.01.zip | 8,49 MB (8.903.725 bytes) | 53,2 MB (55.817.340 bytes) |

We built our own tools to capture them, but there is no post-processing on our side, so they come raw as they are read from each EEG device, in total **383,751,556 Data Points.**

Feel free to test any machine learning, deep learning or whatever algorithm you think it could fit, we only ask for acknowledging the source and please let us know of your performance!

We choose not to differentiate the signals into training/test/validation sets at this point so pick the distribution you prefer.

A small portion of the signals were captured without the stimulus of seeing the digits for contrast, all are random actions not related to thinking or seeing digits, you can decide to use them or not in your tests, they use the code -1.

**SIGNAL DISTRIBUTION:**

This is the distribution of the signals per device and digit:

| Device / Digit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | -1 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **MindWave (MW)** | 5,531 | 5,498 | 5,517 | 5,416 | 5,381 | 5,568 | 5,476 | 5,552 | 5,545 | 5,450 | 12,701 | 67,635 |
| **EPOC (EP)** | 91,224 | 88,914 | 90,930 | 92,652 | 88,886 | 91,994 | 91,322 | 88,718 | 91,728 | 91,882 | 2,226 | 910,476 |
| **Muse (MU)** | 11,904 | 11,632 | 11,920 | 11,832 | 11,536 | 12,052 | 12,368 | 12,080 | 12,208 | 11,988 | 44,412 | 163,932 |
| **Insight (IN)*** | 1.820 | 1.860 | 1.805 | 1.825 | 1.765 | 1.905 | 1.845 | 1.805 | 1.690 | 1.830 | 0 | 18.150 |
| **Total** | 110.479 | 107.904 | 110.172 | 111.725 | 107.568 | 111.519 | 111.011 | 108.155 | 111.171 | 111.150 | 59,339 | **1.160.193** |

* Insight captures started in September 2015, so soon will be updated with more brain signals, **last update 09/16/2015 v1.01**

**FILE FORMAT**:

The data is stored in a very simple text format including:

**[id]:** a numeric, only for reference purposes.

**[event]** id, a integer, used to distinguish the same event captured at different brain locations, used only by multichannel devices (all except MW).

**[device]:** a 2 character string, to identify the device used to capture the signals, "MW" for MindWave, "EP" for Emotive Epoc, "MU" for Interaxon Muse & "IN" for Emotiv Insight.

**[channel]:** a string, to indentify the 10/20 brain location of the signal, with possible values:

| | |
|---|---|
| **MindWave** | "FP1" |
| **EPOC** | "AF3, "F7", "F3", "FC5", "T7", "P7", "O1", "O2", "P8", "T8", "FC6", "F4", "F8", "AF4" |
| **Muse** | "TP9","FP1","FP2", "TP10" |
| **Insight** | "AF3,"AF4","T7","T8","PZ" |

**[code]:** a integer, to indentify the digit been thought/seen, with possible values 0,1,2,3,4,5,6,7,8,9 or -1 for random captured signals not related to any of the digits.

[**size]:** a integer, to identify the size in number of values captured in the 2 seconds of this signal, since the Hz of each

device varies, in "theory" the value is close to 512Hz for MW, 128Hz for EP, 220Hz for MU & 128Hz for IN, for each of the 2 seconds.

**[data]:** a coma separated set of numbers, with the time-series amplitude of the signal, each device uses a different precision to identify the electrical potential captured from the brain: integers in the case of MW & MU or real numbers in the case of EP & IN.

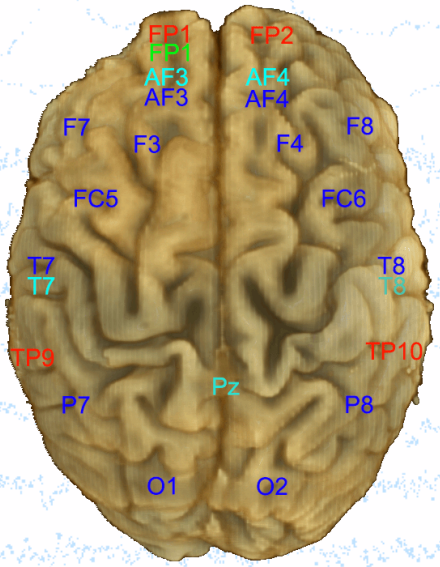
There is no headers in the files, every line is a signal, and the fields are separated by a tab

For example one line of each device could be (without the headers)

| [id] | [event] | [device] | [channel] | [code] | [size] | [data] |
|---|---|---|---|---|---|---|
| 27 | 27 | MW | FP1 | 5 | 952 | 18,12,13,12,5,3,11,23,37,36,26,24,35,42…… |
| 67650 | 67636 | EP | F7 | 7 | 260 | 4482.564102,4477.435897,4484.102564……. |
| 978210 | 132693 | MU | TP10 | 1 | 476 | 506,508,509,501,497,494,497,490,490,493…… |
| 1142043 | 173652 | IN | AF3 | 0 | 256 | 4259.487179,4237.948717,4247.179487,4242.051282…… |

**BRAIN LOCATIONS:**

Each EEG device capture the signals via different sensors, located in these areas of my brain, the color represents the device: **MindWave**, **EPOC**, **Muse**, **Insight**




Contact us if you need any more info.

Let's decode My Brain!
September 2015
David Vivancos
vivancos@vivancos.com