Auxiliary Data and Tools

Synthetic Data Set

In this competition, we present 6 datasets to test your algorithm. The datasets have different diffculties in term of number of local EEG signals and artifact sources as well as white noise with variance 0.1 as shown in Table 1.

	# of Sources	# of Artifact	Noise Free	With Noise
A	4	2	ANF	AWN
B	12	6	BNF	BWN
C	19	6	CNF	CWN

Table 1: Six datasets with different difficulties

Data Set A

Data Set A assumes 6 sources of signals. Four sources (Singers in the Singers Party problem) are EEG data. Two of the sources are assumed to be EMG (instruments in the Singers Part problem) operating at high amplitude at high frequencies overlapping with High Beta and Gamma bands. Each source operates with a mixture of two frequencies representative of classical EEG bands as shown in Table 2.

The signals are sampled at 256Hz. Besides, the fifth source was activated in the last 250ms of every second, and the sixth source was activated in the last 500ms of every second. All other sources were activated from time 0. The reason for delayed activation of the EMG signals is that in real-world dataset, there is no guarantee that the clock for collecting data is synchronized with the clock of artifacts. For example, the drum may start in the middle of the window, while the saxophone may start at the end of that second. The signals are shown in

Source ID	Band	Amplitude	Frequency	Band	Amplitude	Frequency
s1	Delta	14	4	Beta	52	22
s2	Theta	23	7	Beta	70	19
s3	Delta	16	5	Alpha	43	11
s4	Alpha	44	9	Gamma	56	47
s5	EMG	144	31	EMG	337	51
s6	EMG	282	28	EMG	246	49

Table 2: The synthesis of the six signal sources.

Data Set B & C

Data Set B(C) assumes 18(25) sources of signals. 12(19) sources (Singers in the Singers Party problem) are EEG data. 6(6) of the sources are assumed to be artifacts as shown in Table 3.

Similarly, the signals are sampled at 256Hz. Besides, the source 20, 22, 24 were activated in the last 250ms of every second, and the source 21, 23, 25 was activated in the last 500ms of every second. All other sources were activated from time 0. The reason for delayed activation of the EMG signals is that in real-world dataset, there is no guarantee that the clock for collecting data is synchronized with the clock of artifacts. For example, the drum may start in the middle of the window, while the saxophone may start at the end of that second.

In these data set, a EEG model based on 10-20 system as shown in Figure 1, 2 will be used to generate the signals for this competition. The topographic map in Figurewas drawn using toolbox from EEGLAB, the radius of the head model 0.5dm. The exact location for each eletrode and artifact will be provided.

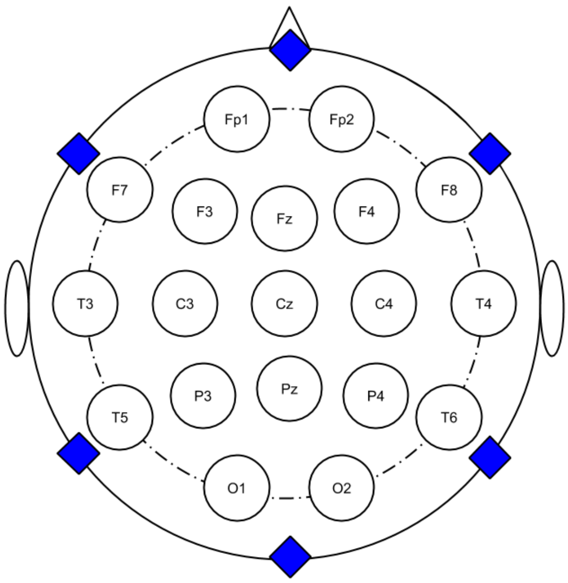

Figure 1: The 10/20 system with artifact location labelled.

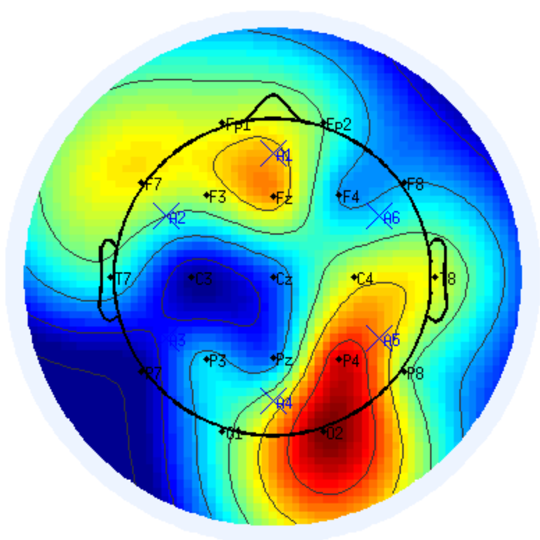

Figure 2: A topographic map of a scalp data field in a 2-D circular view.

The synthetic signals generated from the electrode locations are assumed to be local while 6 artifacts are assumed to be generated at labeled position. For Data set B, 12(*) out of 19 eletrodes will be used while all 19 electrodes will be used for Data set C. Similar to Data set A, the signals generated are assumed to have 2 frequencies ranging from 1 to 50 HZ. Figure 4 also shows the topographic map of a scalp data field in a 2-D circular view to show the contribution of each artifact to all electrodes.

Type	ID	Amplitude	Frequency	Amplitude	Frequency
EEG	1*	38	3	29	13
	2	43	5	28	11
	3	34	15	39	17
	4*	27	21	37	23
	5	26	24	23	39
	6*	25	2	54	33
	7*	55	4	41	20
	8*	20	22	45	36
	9*	24	16	52	32
	10*	53	18	21	41
	11*	33	25	35	45
	12	57	12	51	19
	13	49	6	47	38
	14*	22	42	32	46
	15*	48	26	36	34
	16	42	14	50	44
	17*	56	27	31	40
	18*	40	35	44	37
	19	30	43	46	47
Artifact	20*	177	8	268	29
	21*	151	10	229	30
	22*	281	7	164	49
	23*	216	9	255	52
	24*	242	28	190	51
	25*	294	31	203	50

Table 3: The synthesis of the 25 signal sources. (*) ID will be used in DataSet B.

The W_{ij} for each electrode from each artifact is proportional to the exponential of euclidean distance between them. The relationship is shown in Figure 3. The values are also shown in Table 4.

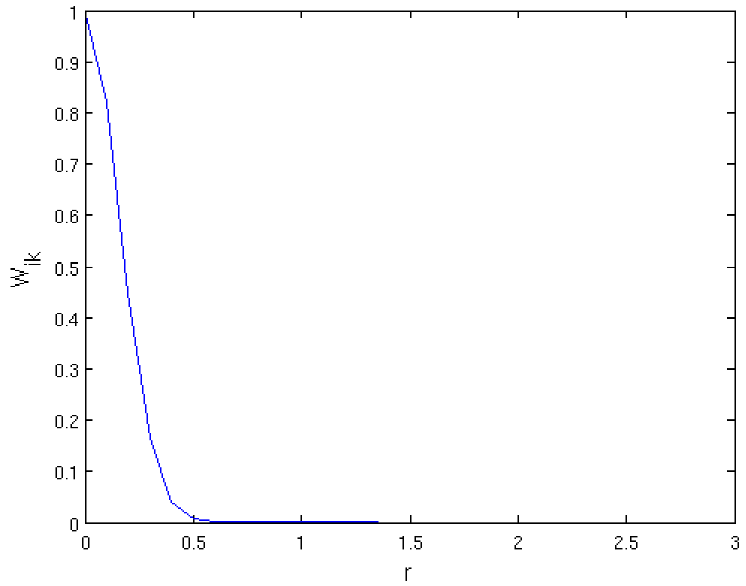

Figure 3 The relationship between W_{ij} *and r*

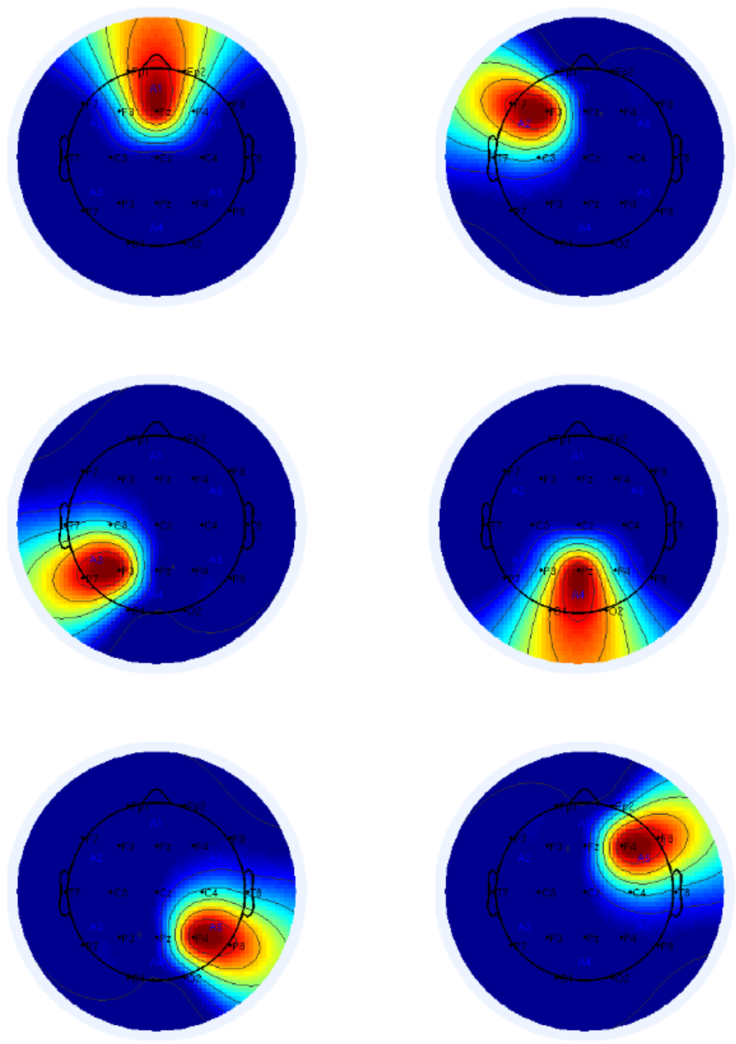

Figure 4: Topographic maps of scalp datas field in a 2-D circular view for each artifact.

Electrode / Artifact	1	2	3	4	5	6
1	0.3049	0.0433	0.0000	0.0000	0.0000	0.0003
2	0.3049	0.0003	0.0000	0.0000	0.0000	0.0433
3	0.0101	0.4586	0.0018	0.0000	0.0000	0.0000
4	0.1401	0.5004	0.0009	0.0000	0.0000	0.0001
5	0.4212	0.0143	0.0000	0.0000	0.0000	0.0143
6	0.1401	0.0001	0.0000	0.0000	0.0009	0.5004
7	0.0101	0.0000	0.0000	0.0000	0.0018	0.4586
8	0.0000	0.1372	0.1372	0.0000	0.0000	0.0000
9	0.0005	0.1703	0.1703	0.0005	0.0000	0.0000
10	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014
11	0.0005	0.0000	0.0000	0.0005	0.1703	0.1703
12	0.0000	0.0000	0.0000	0.0000	0.1372	0.1372
13	0.0000	0.0018	0.4586	0.0101	0.0000	0.0000
14	0.0000	0.0009	0.5004	0.1401	0.0001	0.0000
15	0.0000	0.0000	0.0143	0.4212	0.0143	0.0000
16	0.0000	0.0000	0.0001	0.1401	0.5004	0.0009
17	0.0000	0.0000	0.0000	0.0101	0.4586	0.0018
18	0.0000	0.0000	0.0433	0.3049	0.0003	0.0000
19	0.0000	0.0000	0.0003	0.3049	0.0433	0.0000

Table 4: The W_{ij} for i^{th} electrode and k^{th} artifact

Technical Details

This section provides information and code to generate the Data Sets.

Six data sets are provided to evaluate the algorithm, the matlab file for data genaration is provided as follow:

	# of Sources	# of Artifact	Noise Free	With Noise
A	4	2	ANF.m	AWN.m
B	12	6	BNF.m	BWN.m
C	19	6	CNF.m	CWN.m

Each Data Sets should contain the following information:

Data	File name
Original source signals	EEG.txt
Mixed Signals X	MixedSignals.txt
Components from FastICA S	FastICAcomponents.txt
Mixing Matrix from FastICA A	Amatrix.txt

Unmixing Matrix from FastICA W	Wmatrix.txt
Reconstructed signals from []	reX.txt

The exact location for electrodes and artifacts can be found in

Location file	Myloc.ced

Visualisation Tools (EEGLAB TOPOPLOT)

EEGLAB provides a toolbox to plot a topographic map of a scalp data field in a 2-D circular view.
http://sccn.ucsd.edu/eeglab/allfunctions/topoplot.html

Simple Usage:
topoplot(datavector, EEG.chanlocs); % plot a map using an EEG chanlocs structure
topoplot(datavector, 'my_chan.locs'); % read a channel locations file and plot a map

A example on how to visualise the contribution of each electrode can be found at

Visualise the contribution of each electrode	headmap.m

FastICA

The fastICA is a matlab program that implements fast fixes point algorithm for ICA. The matlab package and details can be found at:

http://research.ics.aalto.fi/ica/fastica/

Simple Usage:

[C, A, W] = fastica(R,'maxNumIterations',10000);

where

Variables	Description
C	Components
A	Mixing Matrix
W	Unmixing Matrix
R	Recorded Signals

FFT and IFFT
Matlab provides a function for Fast Fourier transform and its inverse

Simple Usage:

Y = fft(x)
y = ifft(X)

The details can be found at:
FFT : **http://www.mathworks.com/help/matlab/ref/fft.html**
IFFT : **http://www.mathworks.com/help/matlab/ref/ifft.html**

Evalution Tools

We provide a function to evaluate the goodness of the solution in term of information loss and artifact residue.

Usage:

[infoloss,residue] = infolossResidue(yourRecontructedSignals,dataid)

Example:

[infoloss,residue] = infolossResidue("myXforBWN","BWN")

filename	InfolossResidue.m