Electroencephalography and clinical Neurophysiology, 82 (1992) 391–393

EEG 91604

Short communication

A simple format for exchange of digitized polygraphic recordings *

Bob Kemp [a], Alpo Varri [b], Agostinho C. Rosa [c], Kim D Nielsen [d] and John Gade [d]

[a] *Dept of Clinical Neurophysiology, University Hospital Leiden, Leiden (The Netherlands),* [b] *Dept Electrical Engineering, Tampere University of Technology, Tampere (Finland),* [c] *CAPS, IST, Technical University of Lisbon, Lisbon (Portugal) and* [d] *Dept of Medical Informatics and Image Analysis (MIBA), Aalborg University, Aalborg (Denmark)*

(Accepted for publication 10 January 1992)

Summary A simple digital format supporting the technical aspects of exchange and storage of polygraphic signals has been specified Implementation of the format is simple and independent of hard- or software environments It allows for any local montages, transducers, prefiltering, sampling frequencies, etc

At present, 7 laboratories in various countries have used the format for exchanging sleep-wake recordings These exchanges have made it possible to create a common database of sleep records, to compare the analysis algorithms local to the various laboratories to each other by applying these algorithms to identical signals, and to set up a computer-aided interlaboratory evaluation of manual and automatic analysis methods

Key words Standard format, Digital signal

A large variety of commercial and home-made computer-based systems are being applied for automatic and computer-assisted manual analyses and archival of polygraphic recordings These systems are based on an equally large variety of formats for digital storage of the polygraphic signals In addition, automatic as well as computer-assisted manual analysis methods differ greatly between laboratories These methodological differences make pooling or comparison of results from different laboratories difficult

We have designed a simple standard format for digital storage and exchange of polygraphic recordings Such exchanges make possible the creation of a common archive (i e , database) of recordings, the comparison of algorithms (e g , K-complex detectors, slow-wave quantifiers etc) from different laboratories when applied to identical signals, and also the evaluation of inter-laboratory differences in manual analysis procedures such as sleep classification and K-complex detection

This article describes the format in full detail, making implementation possible in other laboratories Preliminary results of practical exchanges are also reported

The format

One data file contains one uninterrupted digitized polygraphic recording A data file consists of a header record followed by data records The variable-length header record identifies the patient and specifies the technical characteristics of the recorded signals The data records contain consecutive fixed-duration epochs of the polygraphic recording

The first 256 bytes of the header record specify the version number of this format, local patient and recording identification, time information about the recording, the number of data records and finally the number of signals (ns) in each data record Then for each signal another 256 bytes follow in the header record, each specifying the type of signal (e g , EEG, body temperature, etc), amplitude calibration and the number of samples in each data record (from which the sampling frequency can be derived since the duration of a data record is also known) In this way the format allows for different gains and sampling frequencies for each signal The header record contains $256+(ns \times 256)$ bytes Fig 1 shows its detailed format and Fig 2 shows an example

The information in the ASCII strings must be left justified and filled out with spaces Midnight time is 00 00 00 The duration of each data record is recommended to be a whole number of seconds and its size (number of bytes) is recommended not to exceed 61,440 Only if a 1 sec data record exceeds this size limit, the duration is recommended to be smaller than 1 sec (e g , 0 01)

The digital minimum and maximum of each signal should specify the extreme values that can occur in the data records These often are the extreme output values of the A/D converter The physical (usually also physiological) minimum and maximum of this signal should correspond to these digital extremes and be expressed in the also specified physical dimension of the signal These 4 extreme values specify offset and amplification of the signal

Following the header record, each of the subsequent data records contains 'duration' seconds of 'ns' signals, with each signal being represented by the specified (in the header) number of samples In order to reduce data size and adapt to commonly used software for acquisition, processing and graphical display of polygraphic signals, each sample value is represented as a 2-byte integer in 2s complement format Fig 1 shows the detailed format of each data record See also the example in Fig 2

Correspondence to Dr Bob Kemp, University Hospital, Building 20, Dept of Clinical Neurophysiology, Rijnsburgerweg 10, NL-2333 AA Leiden (The Netherlands)
Tel +31 71 262188, Fax +31 71 213152

* This work was financially supported by the European Communities (COMAC-BME) and the Leiden Foundation for Sleep and Biorhythms

HEADER RECORD

8 ascii	version of this data format (0)
80 ascii	local patient identification
80 ascii	local recording identification
8 ascii	startdate of recording (dd mm yy)
8 ascii	starttime of recording (hh mm ss)
8 ascii	number of bytes in header record
44 ascii	reserved
8 ascii	number of data records (-1 if unknown)
8 ascii	duration of a data record, in seconds
4 ascii	number of signals (ns) in data record
ns * 16 ascii	ns * label (e g EEG FpzCz or Body temp)
ns * 80 ascii	ns * transducer type (e g AgAgCl electrode)
ns * 8 ascii	ns * physical dimension (e g µV or degreeC)
ns * 8 ascii	ns * physical minimum (e g -500 or 34)
ns * 8 ascii	ns * physical maximum (e g 500 or 40)
ns * 8 ascii	ns * digital minimum (e g -2048)
ns * 8 ascii	ns * digital maximum (e g 2047)
ns * 80 ascii	ns * prefiltering (e g HP 0 1Hz LP 75Hz)
ns * 8 ascii	ns * nr of samples in each data record
ns * 32 ascii	ns * reserved

DATA RECORD

nr of samples[1]	* integer	first signal in the data record
nr of samples[2]	* integer	second signal
nr of samples[ns]	* integer	last signal

Fig 1 Detailed digital format of the header record (upper block, ascii's only) and of each subsequent data record (lower block, integers only) Note that each one of the ns signals is characterized separately in the header

```
0         Free local patient identification: take care of priv
acy regulations!              Free local recording identificat
ion                                                16 09 87 20
35 00 768      Reserved                                        2
880      30       2    EEG FpzCz       Body temperature Ag-Ag
Cl cup electrodes
    Rectal thermistor
                          µV      Degree C -440       34 4
 510      40 2      -2048    -2048    2047     2047      Time
constant 1s, First order lowpass at 75Hz
     DC to 0 1Hz (first-order)
                           15000    3       Reserved
               Reserved
```

Fig 2 Header record of a 24 h recording of EEG and body temperature sampled at 500 Hz and 0 1 Hz, respectively Note that the offsets of EEG and body temperature are 35 μV and 37 3°C, while the gains are 4 31/μV and 706 2/°C, respectively In this example, each 30 sec data record contains 15,000 samples of the EEG followed by 3 samples of the body temperature signal

Gains, electrode montages and filters should remain fixed during the recording Of course, these may all be digitally modified during replay of the digitized recording

Results

The format has been implemented in the last year in practical exchanges of 8–24 h polygraphic sleep-wake recordings between laboratories in Tampere (Finland), Aalborg (Denmark), Marburg (Germany), Lisbon (Portugal), Oxford (U K), Berlin (Germany) and Leiden (The Netherlands) Each laboratory developed software for the conversion of the local data format to and from the standard format The time required for this software development varied from several hours to several days The format had been sufficiently specified to make further consultation unnecessary All laboratories successfully analyzed non-local recordings using local software

The format is independent of exchange hardware, operating systems, and software environment Some exchanges were realized using a network connection between an MS-DOS system (either version 3 or 4) and RMX or UNIX systems In other cases, MS-DOS files were transported using diskettes, tape streamers and optical disks In most cases, a particular type of 940MB WORM laser disk was used to exchange MS-DOS files in which integer samples were based on the usual Intel low-high-byte format (right justified with sign extension)

Exchanged data contained EEG, EMG, EOG, AOG, respiration, body temperature, ECG, SaO_2, event marker, etc Within and between recordings, various sampling frequencies were used for the different signals

In one particular application of the format, a laser disk with full-night polysomnographic recordings from 4 laboratories was sent to 6 laboratories and analyzed with the locally developed software In this way, different methods for the quantification of EEG rhythms were compared Although the format allows for any EEG sampling frequency, both manual and automatic analyses of EEG signals strongly depend on the chosen sampling frequency Therefore, the involved laboratories agreed to sample the EEG at 100 Hz in this particular application Some results of these analyses are shown in Fig 3 In a similar manner both automatic and manual detection of K-complexes and eye movements are now being evaluated

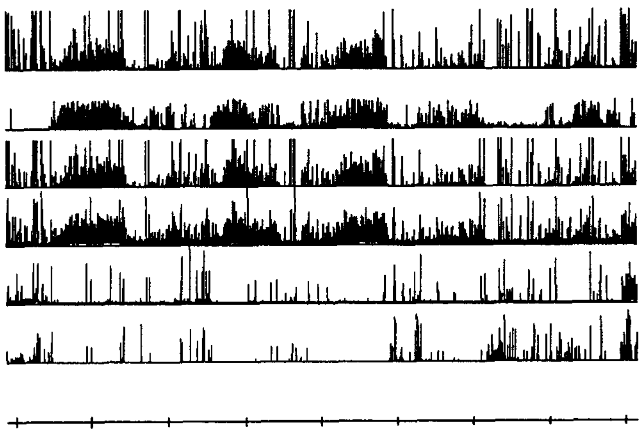

Fig 3 Delta power plots from a full-night (8 5 h) Tampere C3A2 recording as obtained by algorithms based on (from the top downwards) the Aalborg autoregressive model, Leiden synchronization model, Marburg Fourier analysis, and Tampere bandpass filter The two lower traces show Tampere-detected EMG artifacts and Leiden-detected rapid eye movements, respectively Vertical scalings were tuned for comparable amplitudes Horizontal calibration 1 h/div

Discussion

We intentionally designed a "core" format that encompasses only the minimally required information for the technical aspects of practical exchanges A tremendous effort would be required in order to formalize additional information about recording circumstances, patient characteristics, etc , and such an effort would still fail to completely cover all information that is relevant to particular medical and/or research applications In addition, privacy regulations can influence which information can be exchanged Therefore, all aspects of practical exchanges must be thoroughly and comprehensively discussed prior to each exchange, with agreement being reached over relevant technical choices such as prefiltering, sampling frequency, exchange media and file names

The core format is simple and easy to implement in different systems Because all signals can be specified separately, the format allows great flexibility in selection of signals, sampling frequencies, gains, prefiltering, etc As a result of the exchanges, we have concluded that the current format is adequate, we do not expect that it will be necessary to update the format for a considerable time It can, therefore, also be adopted by laboratories that have home-made analysis systems but only limited engineering facilities for the purpose of such exchange applications

This work was coordinated by the Task Group on Signal Analysis (leader B Kemp), which operates within the European Concerted Action on Methodology for the Analysis of the Sleep-Wakefulness Continuum (leader A N Nicholson) Erik Simonsen and Jan M Franzen helped to design the 1988 prototype of the format Laurel Beecher edited the language and style of this article