Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research

CHRISTIAN O'REILLY[1,2], NADIA GOSSELIN[1,3], JULIE CARRIER[1,3] and TORE NIELSEN[1,2]

[1]Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur de Montréal, Montreal, QC, Canada, [2]Département de Psychiatrie, Université de Montréal, Montreal, QC, Canada and [3]Département de Psychologie, Université de Montréal, Montreal, QC, Canada

Keywords
collaboration tool, methodology, open-data, reproducibility

Correspondence
Christian O'Reilly, Dream and Nightmare Laboratory, Center for Advanced Research in Sleep Medicine, Hôpital du Sacré-Coeur de Montréal, 5400 boulevard Gouin Ouest, Montréal, Québec H4J 1C5, Canada.
Tel.: 1-514-338-2222 ext 2738;
fax: 1- 514-338-2531;
e-mail: christian.oreilly@umontreal.ca

Accepted in revised form 4 May 2014;
received 4 March 2014

DOI: 10.1111/jsr.12169

SUMMARY

Manual processing of sleep recordings is extremely time-consuming. Efforts to automate this process have shown promising results, but automatic systems are generally evaluated on private databases, not allowing accurate cross-validation with other systems. In lacking a common benchmark, the relative performances of different systems are not compared easily and advances are compromised. To address this fundamental methodological impediment to sleep study, we propose an open-access database of polysomnographic biosignals. To build this database, whole-night recordings from 200 participants [97 males (aged 42.9 ± 19.8 years) and 103 females (aged 38.3 ± 18.9 years); age range: 18–76 years] were pooled from eight different research protocols performed in three different hospital-based sleep laboratories. All recordings feature a sampling frequency of 256 Hz and an electroencephalography (EEG) montage of 4–20 channels plus standard electro-oculography (EOG), electromyography (EMG), electrocardiography (ECG) and respiratory signals. Access to the database can be obtained through the Montreal Archive of Sleep Studies (MASS) website (http://www.ceams-carsm.ca/en/MASS), and requires only affiliation with a research institution and prior approval by the applicant's local ethical review board. Providing the research community with access to this free and open sleep database is expected to facilitate the development and cross-validation of sleep analysis automation systems. It is also expected that such a shared resource will be a catalyst for cross-centre collaborations on difficult topics such as improving inter-rater agreement on sleep stage scoring.

INTRODUCTION

The polysomnographic (PSG) study of sleep is notorious for requiring extensive and laborious signal processing. This ongoing problem stems from the need to assess multiple whole-night recordings usually composed of large numbers of sensors. A typical recording consists of biosignals of body functions as diverse as brain activity (electroencephalography, EEG), eye movements (electro-oculography, EOG), chin, face or limb muscle activity (electromyography, EMG), heart rhythm (electrocardiography, ECG), respiratory rhythms and oximetry. Typical analyses require artefact rejection, identification of sleep stages according to standard rules (Iber *et al.*, 2007) and the detection of a wide variety of behavioural events such as leg movements, apneas and micro-arousals, as well as transient EEG events such as sleep spindles, slow waves and K-complexes. Much effort has been expended to develop tools that automate these complex, error-prone and time-consuming tasks. For example, multiple systems have been proposed for the automated detection and classification of sleep stages (Anderer *et al.*, 2005; Berthomier *et al.*, 2007; Figueroa Helland *et al.*, 2010; Hang *et al.*, 2013; Koley and Dey, 2012; Park *et al.*, 2000; Tagluk *et al.*, 2010; Virkkala *et al.*, 2007), muscular, ocular and cardiac artefacts (Brunner *et al.*, 1996; Devuyst *et al.*, 2008; Durka *et al.*, 2003; Moretti *et al.*, 2003; Park *et al.*, 2002), sleep disorders

symptoms such as apneas and hypopneas (Al-Angari and Sahakian, 2012; Khandoker *et al.*, 2009; Koley and Dey, 2013; Liu *et al.*, 2008; Mendez *et al.*, 2009) and transient events such as
K-complexes (Bankman *et al.*, 1992; Da Rosa *et al.*, 1991; Erdamar *et al.*, 2012; Huy Quan *et al.*, 2012; Jansen and Desai, 1994; Richard and Lengelle, 1998; Tang and Ishii, 1995), vertex waves (Da Rosa *et al.*, 1991), sleep spindles (Acır and Güzeliş, 2004; Babadi *et al.*, 2012; Duman *et al.*, 2009; Huupponen *et al.*, 2007; Schimicek *et al.*, 1994; Schonwald *et al.*, 2006; Ventouras *et al.*, 2005) and others.

However, a key problem with these developments is that independent research teams often develop algorithms with closed-access implementation and assess their performance on private databases. Thus, it is not possible to replicate these works or compare different algorithms with a common set of benchmarks. Determining which algorithms are the most reliable and which are best suited for different types of applications is thus not feasible.

Therefore, sleep research is confronted with a multi-faced methodological problem. The first part is related to the widespread use of closed-source algorithms. Although these algorithms are often reported in scholarly publications using high-level pseudo-code, there is an intrinsic limitation—known as the 'curse of ambiguity' (Ince *et al.*, 2012)—in using natural languages for such descriptions. Two researchers attempting to replicate the same paper describing an algorithm and its implementation and testing may not obtain identical results because of unspecified parameters, fuzzy definitions, coding bugs, and so on. Such a situation obviously hinders the reproducibility of research outcomes and the identification and improvement of the best systems. To address this part of the problem, many researchers have proposed open-source versions of their algorithms on their websites or on public repositories (e.g. code.google.com, sourceforge.net, bitbucket.org, github.com). Some examples related to sleep research include the various software projects related to the processing of EDF files (edfplus.info/downloads) and other toolboxes proposed in the scientific literature (e.g. Leclercq *et al.*, 2011; O'Reilly and Nielsen, 2014).

The second part of the methodological problem is the question of the missing gold standard. Indeed, expert scoring has been considered *de facto* as a gold standard, although it has been known for many years that expert agreement is very low [e.g. interscorer reliability approximately 86% for spindle scoring (Campbell *et al.*, 1980) and 83% for sleep stage scoring (Shambroom *et al.*, 2012)]. Interesting work has been performed by Warby *et al.* (2014) towards the search for a better gold standard for the case of sleep spindle scoring.

The third part of the problem is an issue of rigour and completeness in reporting system evaluation metrics, such as the invalid use of specificity (instead of precision) for highly unbalanced detection problems (O'Reilly and Nielsen, 2013; O'Reilly and Nielsen, in preparation).

The fourth part of sleep research's methodological problem is that researchers and engineers do not have access to a high-quality open-access PSG database that may be used to benchmark and validate new, independently developed analytical tools. More limited databases have been made publically available, such as those associated with the DREAMS project (Devuyst *et al.*, 2011), but these are very restricted in size. The Sleep Heart Health Study also proposes a PSG database derived from a prospective cohort study (Quan *et al.*, 1997). However, this large collection of 9736 PSG recordings failed to become accepted as a standard resource by third parties. Potential causes for this limited use may be associated with the fact that the investigators proposing the database were not aiming to provide an open-access and general purpose archive. Indeed, it is oriented specifically towards the investigation of relationships between sleep-disordered breathing and heart diseases, and its recordings are available only upon special request and approval. Also, the EEG montage (C3-A2 and C4-A1 channels, sampled at 125 Hz) is relatively limited for general-purpose investigations in sleep research.

Some other databases can be found on http://physionet.org/, including as the CAP sleep database [108 nights with three (or more) EEG channels, from different populations (16 healthy and 92 pathological), targeting the study of cyclic alternating patterns], the archive from St Vincent's University Hospital (25 overnight PSGs with two EEG channels from a population of subjects with sleep apnea) and the database from Kemp *et al.* (2000) (61 PSGs with two EEG channels, from subjects with mild sleep onset difficulties, recorded at 100 Hz). However, these archives are limited in different aspects (number of channels, sampling frequency, number of records) and are not embedded within a collaborative project; for example, additional nights, annotations and comments cannot be added interactively to the database. Also, in our opinion, none of these databases has been widely accepted for benchmarking automated systems. To contribute to a solution for this part of the problem, we present a general-purpose, open-access database designed primarily for the study of sleep but also applicable to a wide range of physiological investigations: the Montreal Archive of Sleep Studies (MASS).

METHODS

Included recordings and variables

Two hundred PSG recordings gathered from eight research protocols conducted between 2001 and 2013 in three different laboratories of the Center for Advanced Research in Sleep Medicine (CARSM), Montreal, Canada, were pooled to form the archive. To provide homogeneous cohorts, these recordings were organized into five subsets (entitled SS1–SS5), according to the research protocols from which they were gathered, with the exception of recordings from four research protocols with similar properties that were merged into one subset (SS5). Specifications related to the acquisition of signals (amplifier system, software, digitization gain and filtering) vary across subsets and are reported in Table 1.

Table 1 Acquisition characteristics for subsets of the Montreal Archive of Sleep Studies cohort 1 (MASS-C1);* low and high filter cutoffs are first-order (−6 dB/octave)

	SS1	*SS2*	*SS3*	*SS4*	*SS5*
Amplifier system	Grass† Model 15	Grass† Model 12	Grass† Model 15	Grass† Model 12 or Grass† Model 15	Grass† Model 15
Filters low cutoff	0.10 Hz (EOG, ECG, thermistance, cannula, oximetry, respiratory belt), 0.30 Hz (EEG) or 10 Hz (EMG)	0.10 Hz (EOG, ECG, thermistance), 0.30 Hz (EEG) or 10 Hz (EMG)	0.10 Hz (EOG, ECG), 0.30 Hz (EEG) or 10 Hz (EMG)	0.10 Hz (EOG, ECG, thermistance), 0.30 Hz (EEG) or 10 Hz (EMG)	0.10 Hz (EOG, ECG), 0.30 Hz (EEG) or 10 Hz (EMG)
Filters high cutoff	30 Hz (EOG, ECG, thermistance, cannula, oximetry, respiratory belt) or 100 Hz (EMG, EEG)	30 Hz (ECG, thermistance) or 100 Hz (EEG, EOG, EMG)	30 Hz (ECG, EOG) or 100 Hz (EEG, EMG)	30 Hz (ECG, thermistance) or 100 Hz (EEG, EOG, EMG)	30 Hz (ECG, EOG) or 100 Hz (EEG, EMG); 60 Hz notch filter
60 Hz notch filter	On most recordings	No	Yes	No	On some recordings
Gain	10 000	EEG: 20 000 EOG: 10 000 Chin EMG: 200,000 Other EMG: 50,000 EKG: variable 2000–5000	10 000	EEG: 20 000 EOG: 10 000 EMG: 20 000 EKG: 5000	10 000
Acquisition software	Harmonie‡ 6.2b FR	Harmonie‡ 5.4	Harmonie‡ 6.2b FR	Harmonie‡ 6.0a EN-0	Harmonie‡ 6.2b FR

*C1 designates cohort 1, as explained further in Methods, Structure section.
†Grass Technologies, Astro-Med Inc., West Warwick, RI, USA.
‡Natus Medical Inc., San Carlos, USA.

Identification, types and characteristics of signals included in the PSGs for the different subsets are reported in Table 2. In every case, EEG electrodes were placed according to the standard 10–20 system and recordings were performed at 256 Hz. Acquisition frequencies for the non-EEG signals are also reported in Table 2. Except for the four-channel case, EEG montages always include the BASE 16-channel montage (i.e. C3, C4, Cz, F3, F4, F7, F8, O1, O2, P3, P4, Pz, T3, T4, T5, T6) plus additional channels (Fp1, Fp2, Fpz, Fz or Oz), depending on the subset. Although these specifications are generally homogeneous within subsets, there are some minor exceptions (e.g. SS1 contains some nights with 17 channels, others with 19; in SS5, some records were recorded with a 60 Hz notch filter, some without). These specifications can be obtained on a subject-by-subject basis on the MASS website.

Inclusion/exclusion criteria were also defined differently for each project and are reported in Table 3. All subjects included in this cohort were healthy controls (even though the protocol may have also included a patient group), except for 15 included in SS1, who were diagnosed with MCI according to a neuropsychological evaluation. No experimental manipulations were performed for any included nights except for the use of placebo in some subjects of SS5. For SS1, subjects with an apnea/hypopnea index up to 20 were included in order to allow the testing of tools developed for this population. Also, in the same subset, seven subjects with an index of periodic leg movements in sleep (PLMS) associated with a micro-arousal were included to provide opportunities for validating (PLMS) detectors.

Sleep was staged by experienced PSG technicians according to either American Academy of Sleep Medicine (AASM) guidelines (Iber *et al.*, 2007) or Rechtschaffen and Kales (1968) rules (RK), as reported in Table 4. Apnea/hypopnea, micro-arousals and PLMS were also scored by experts in SS1. Muscular artefact rejection was performed using automatic annotation based on the technique of Brunner *et al.* (1996) and followed-up by manual correction by an expert scorer.

Apart from biosignals and annotations listed in Table 4, several supplementary variables are available for each PSG recording. These are provided in two supplementary files referred to as the 'open-access descriptor file' and the 'restricted-access descriptor file'. Although this set of variables may change over time, Table 5 provides a list of variables included in the former file at time of publication.

The restricted-access descriptor file contains supplementary information for each subject, including age, sex, education and neuropsychological testing scores. The detailed list is available on the MASS website.

File format of PSG records

MASS records are provided in the European Data Format (EDF)++ format. This is a substandard of the EDF+ format (much as EDF+ is a substandard of the EDF format), meaning that these files are compatible with the EDF+

Table 2 Properties of polysomnography (PSG) recordings for subsets of the Montreal Archive of Sleep Studies cohort 1 (MASS-C1)

	SS1	*SS2*	*SS3*	*SS4*	*SS5*
EEG montage size	17 or 19 (BASE, Fz, and in some cases Fp1, Fp2)	19 (BASE, Fp1, Fp2, Fpz)	20 (BASE, Fp1, Fp2, Fz, Oz)	4 (C3, C4, O1, O2)	20 (BASE, Fp1, Fp2, Fz, Oz)
Reference	Most CLE, some LER	CLE	LER	CLE	LER
EOG	2 channels (left, right) (256 Hz)	4 channels (left, right, up, down) (256 Hz)	2 channels (left, right) (256 Hz)	4 channels (left, right, up, down) (256 Hz)	2 channels (left, right) (256 Hz)
EMG	5 bipolar channels (anterior tibialis right, anterior tibialis left, 3 chin channels) (256 Hz)	1 bipolar channel (chin) (256 Hz)	3 bipolar channels (chin) (256 Hz)	1 bipolar channel (chin) (256 Hz)	3 channels (chin) (256 Hz)
ECG	1 channel (D-I) (512 Hz)	1 channel (D-I) (256 Hz)	1 channel (D-I) (512 Hz)	3 channels (D-I, D-II, D-III) (512 Hz)	1 channel (D-I) (256 Hz)
Respiratory thermistance	1 channel (32 or 128 Hz)	1 channel (64 Hz)	NA	1 channel (64 Hz)	NA
Air flow cannula	1 channel (128 Hz)	NA	NA	NA	NA
Oximetry	1 channel (256 Hz)	NA	NA	1 channel (64 Hz)	NA
Belt	1 or 2 channels (abdominal and sometimes thoracic) (128 Hz)	NA	NA	NA	NA
Skin conductance	NA	NA	NA	1 channel (finger) (64 Hz)	NA

LER, linked-ear reference with a 10 kΩ resistance; CLE, computed linked-ear.
In the case of montages with a CLE reference, electroencephalography (EEG) channels A1 or A2 are implicitly included among the available channels, such that the montage can be reformatted if desired. For SS3 and SS5, all recording channels were referential. Bi-polar channels provided in the MASS [e.g. for electromyography (EMG)] were obtained through offline reformatting. EOG, electro-oculography; ECG, electrocardiography; NA, not available.

Table 3 Exclusion criteria for different subsets of the Montreal Archive of Sleep Studies cohort 1 (MASS-C1)

	SS1	*SS2*	*SS3*	*SS4*	*SS5*
BDI	<22	<13	<13	<13	<13
BAI	<30	–	<7	–	<7
Apnea and hypopnea index (h^{-1})	≤ 20	–	<10	–	<10
PLMS-A	–	–	<10	–	<10
No MCI	Verified but not used for rejection	–	Verified	–	Verified
No mental disorders*	Verified	Verified	Verified	Verified	Verified
Others	See†	At least one dream recalled per week	See†	At least one dream recalled per week	See†

BDI, Beck Depression Inventory score; BAI, Beck Anxiety Inventory score; PLMS-A: number of periodic leg movements associated with arousal (per hour of sleep); MCI, mild cognitive impairment according to neuropsychological evaluation following Gagnon *et al.* (2009) criteria.
*Excluded were subjects taking antidepressant drugs, currently diagnosed with depression or other major mental disorders in the 10 last years, or with a personality disorder.
†Excluded were subjects with a history of concussion or traumatic brain injury or having been diagnosed with brain stroke, pulmonary disease, neurological disorder or sleep problems (except for mild obstructive sleep apnea).

format. However, a richer annotation scheme was judged more useful than the one incorporated readily into the EDF+ format. We used a distinct name for this substandard (EDF++) to be able to indentify clearly the recordings following this annotation scheme. In EDF+, an arbitrary character string can be used to encode annotations (Kemp and Olivan, 2003). To allow event properties to be recorded alongside events themselves, an XML description is used in the EDF++. For example, the following string would describe a spindle event detected automatically on channel Cz-A1 during sleep stage 2 of the second rapid eye movement–non-REM (REM–NREM) cycle with a root mean square (RMS) amplitude of 12.07 μV: *<Event channel = "Cz-A1" name = "SpyndleRMS" scoringType = "automatic" groupName = "spindle" rmsAmp = "12.07" stage = "Sleep stage 2" frequency = "12.98" cycle = "2"/>*. More generally, the

Table 4 Details about available annotations for Montreal Archive of Sleep Studies cohort 1 (MASS-C1) (at publication time)

	SS1	*SS2*	*SS3*	*SS4*	*SS5*
Sleep stage scoring rules	AASM	RK	AASM	RK	RK
Page size (s)	30	20	30	20	20
Sleep spindles	N/A	Expert (in progress)	NA	NA	NA
Muscular artefacts	Automatic	NA	Automatic	NA	Automatic
Apnea/hypopnea	Expert	NA	NA	NA	NA
Micro arousal	Expert	NA	NA	NA	NA
PLMS	Automatic	NA	NA	NA	NA

AASM, American Academy of Sleep Medicine; RK, Rechtschaffen and Kales; PLMS, periodic leg movements in sleep; NA, not available.

Table 5 Definition of variables in the open-access descriptor file (at publication time)

Variable	*Definition*
Night code	Unique identifier for each PSG recording following the format CC-SS-XXXX where CC is the cohort number, SS is a subset number (unique within the cohort) and XXXX is a record number (unique within the subset)
Reference	Reference used in the EEG montage
Notch filter	Whether or not a 60 Hz notch filter was applied during acquisition
AHI	Apnea/hypopnea index (available in SS1)
MA	Micro arousal (available in SS1)
PLMS	Index of periodic leg movements in sleep (available in SS1)
PLMS-A	Index of periodic leg movements in sleep associated with a micro arousal (available in SS1)
Scoring rules	Either RK or AASM, depending on which scoring rules were applied
Page duration	Duration of the scoring pages; either 20 or 30 s

AASM, American Academy of Sleep Medicine; EEG, electroencephalography; PSG, polysomnography; RK, Rechtschaffen and Kales.

format is *<Event X/>*, where *X* is a space-separated list of fields encoded as *Name = "Description"* where *Name* is the name of the attribute and *Description* is a string representation of its value. To allow these annotations to be truly interoperable, a more precise and complete description of the EDF++ format is given on the MASS web page. For a given recording, biosignals and annotations are separated into two files, both having the same name except for the addition of the suffix *PSG* and *Annotations*, respectively. This separation has the advantages of (i) lowering the risk of accidental file corruption (the *PSG* file containing the raw recordings never needs to be modified), (ii) improving efficiency of data manipulation (e.g. the *PSG* file does not need to be rewritten when annotations are modified) and (iii) enhancing the sharing of findings (*PSG* files are downloaded once, *Annotations* files can be updated as frequently as necessary). Manual sleep scoring is included in the *Annotations* file and is in agreement with standard EDF+ annotation texts. For more flexibility in their distribution, PSG recordings and annotations are shared as two separate bundles.

Availability

To access the MASS database, interested users should consult the documentation available at http://www.ceams-carsm.ca/en/MASS. In agreement with the requirements of our local ethics board, the database will initially be made available for 25 years, with possible renewal if it still presents scientific value. Depending upon the type of data to be accessed, downloads can be made directly or an application can be completed. Because the MASS database contains lengthy biophysiological recordings, ethical concerns related to subject privacy were rigorously respected, including complete anonymization and control over the distribution and use of the database. Nonetheless, access was designed to be as simple and straightforward as possible. Restrictions on access to the different types of files are described in Table 6.

As shown in Table 6, annotations and basic descriptors are open-access; no restrictions to access are applied. For biosignals and restricted-access descriptors, a request must be submitted from a researcher affiliated with a hospital, university or other research institution. In that process, a document certifying that the project has been approved by the researcher's local Ethics Review Board (ERB) must be submitted such that the ERB of the MASS host [i.e. the Research Center of the Hôpital Sacré-Coeur de Montréal (RC-HSCM)] can keep a record of it. Note that no access restrictions are applied to the use of these data (the host institution will deny no request) as long as there are no obvious ethical problems.

In most cases, the use of restricted-access descriptors should not be necessary, but when they are, the research project will also be evaluated by the host sleep centre to ensure that the proposed request does not conflict with research hypothesis currently evaluated in other projects using the database (Table 6). Note that this restriction only applies to research hypotheses (e.g. a project testing whether sleep spindle density is different between men and women) and not to system and algorithm development (e.g. more than one

Table 6 Access restrictions for the different kinds of information available in the Montreal Archive of Sleep Studies (MASS) database

	Affiliation with a research institution required	*Local ethics review board accreditation required*	*Project validated for conflict**
Annotations (.edfa files)	No	No	No
Open-access descriptors	No	No	No
Biosignals (.edf files)	Yes	Yes	No
Restricted-access descriptors	Yes	Yes	Yes

*In this case, a request must be submitted stating the research hypotheses that are to be investigated with these data. The database manager will verify that these hypotheses are not already being investigated before granting access. See the text for more details.

research team can use the database simultaneously to design K-complex detectors for elderly subjects).

Structure

To allow future expansions of the database without compromising the comparison of results from different studies, it is organized in cohorts; the 200 night recordings described in the present paper constitute cohort 1 (C1). Within a cohort, recordings sharing similar characteristics are grouped into subsets identified by an ID, in order to facilitate selecting sets of homogeneous records. To facilitate interstudy comparisons and meta-analyses over time, authors using the archive should report in their methods sections which cohort(s) and subset(s) were used (e.g. MASS-C1 for the whole of cohort 1, MASS-C1/SS1-SS3 for subsets 1, 2 and 3 of cohort 1). Future cohorts gathered from internal and/or external contributors will be added, depending on interests demonstrated by collaborators in the research community.

Principal characteristics of the archive

Table 7 reports the distribution of sex across the different subsets of the MASS first cohort, whereas Table 8 gives descriptive statistics for age [mean, standard deviation (SD), minimum (min) and maximum (max)] for the whole cohort, for each subset and for women, men and both sexes combined.

Table 7 Sex distributions for Montreal Archive of Sleep Studies cohort 1 (MASS-C1) subsets

	SS1	*SS2*	*SS3*	*SS4*	*SS5*	*Total*
Male	34	8	28	14	13	97
Female	19	11	34	26	13	103
Total	53	19	62	40	26	200

Table 8 Age characteristics for Montreal Archive of Sleep Studies cohort 1 (MASS-C1) subsets

		Mean	*SD*	*Min*	*Max*
C1/SS1	Total	63.6	5.3	55	76
	Men	63.5	5.6	55	76
	Women	63.7	4.9	56	72
C1/SS2	Total	23.6	3.7	18	33
	Men	24.3	4.2	19	33
	Women	23.2	3.5	18	30
C1/SS3	Total	42.5	18.9	20	69
	Men	40.4	19.4	20	69
	Women	44.2	18.6	20	69
C1/SS4	Total	25.3	4.3	18	35
	Men	27.4	4.5	21	35
	Women	24.1	3.9	18	33
C1/SS5	Total	25.0	7.4	20	59
	Men	23.0	2.2	20	27
	Women	26.9	10.0	20	59
C1	Total	40.6	19.4	18	76
	Men	42.9	19.8	19	76
	Women	38.3	18.9	18	72

DISCUSSION

MASS: an ongoing exploratory collaboration

Because the value of the proposed database depends largely upon the willingness of the research community to use it, we deeply appreciate comments from researchers regarding possible improvements, limitations, desired expansion to future cohorts, and so on. These can be submitted on the MASS website or by e-mail to the first author.

Collaboration in the form of additions to the annotation files for individual recordings is also most welcome, and can also be submitted through the MASS website. References to papers describing additions to annotation files can be provided and will be noted on the MASS website so that future works using these annotations can refer to the appropriate methodological descriptions. Collaborations are also invited regarding possible improvements to the EDF++ format.

External validity

The first cohort of this database was selected to be representative of a generally healthy adult population, apart from subset SS1, which includes some MCI subjects. Researchers are invited to choose only subsets that will have good external validity for their specific topic of investigation. Subset characteristics described in this paper should facilitate their choices. Clearly, future inclusion of clinical cohorts will be necessary to make the MASS database more useful for developing tools to be used with these more specialized clinical populations.

Usefulness of a shared archive of sleep studies

The MASS project is designed to further support the many outstanding efforts that have already been made to automate tedious and time-consuming signal processing tasks. In general, it is hoped that an easily accessible, shared database might foster more and better collaborations between research teams, help in standardizing the domain and promote overall advancement in the sleep field. The easy availability of open-access tools could go far in facilitating synergy between independent sleep researchers and in enhancing the reliability of studies. For example, a collective effort by different laboratories to score the MASS recordings could facilitate a thorough assessment of inter-rater agreement for events that are frequently scored in sleep studies.

It is also hoped that availability of the MASS database will provide users who do not have access to sleep laboratories or large cohorts of PSG recordings with the basic sleep recordings they might need for exploratory analyses, or for confirmation of previously obtained results. A particularly valuable application of the database would be for rapid replication of important findings using a subject cohort with known attributes. Some of the priority clinical and research applications of the MASS database might include the following:

- Helping in the establishment of a commonly agreed upon benchmarking system. The development of such a benchmarking system as a part of the MASS website constitutes a future project.
- Giving opportunities for the development of PSG analysis automation systems by providing an easily accessible set of recordings for engineering teams who are willing to test pattern recognition techniques on typical problems of the sleep field but who may be lacking the facilities and expertise to perform PSG recordings.
- Testing automated detection/classification systems for various micro- and macro-structural sleep events (artefacts, transient events, sleep stages, symptoms of sleep disorders) on a common archive, thus facilitating interstudy comparisons.
- Evaluating inter-rater agreement with standard or novel manual detection/classification approaches.
- Conducting interinstitutional assessments of human judge scoring.
- Benchmarking the scoring of expert judges and increasing interinstitution scoring concordance.
- Performing exploratory analyses of PSG properties for normal sleep.
- Validating average statistics of particular sleep events with a standard reference.
- Allowing rapid replication of important or controversial findings.

The Montreal Archive of Sleep Studies has the potential to improve manual and automatic sleep analysis and to inspire multiple interinstitutional and interdisciplinary collaborations. We invite scholars of all backgrounds to participate in the deployment of this new open initiative.

ACKNOWLEDGEMENTS

This work was supported by the NSERC PGS BP scholarship to Christian O'Reilly, CIHR and NSERC grants to Tore Nielsen, Julie Carrier and Nadia Gosselin and FRQS grants to Nadia Gosselin. The authors wish to thank Sonia Frenette, Tyna Paquette and Hélène Blais for their valuable help in retrieving and preparing PSG recordings and related information from the in-house archives.

AUTHOR CONTRIBUTIONS

CO'R contributed to the study design, data analysis and preparation of the manuscript. NG contributed to the study design, data collection and revision of the manuscript. JC contributed to the study design, data collection and revision of the manuscript. TN contributed to the study design, data collection and revision of the manuscript.

CONFLICT OF INTEREST

No conflicts of interest declared.

REFERENCES

Acır, N. and Güzeliş, C. Automatic recognition of sleep spindles in EEG by using artificial neural networks. *Exp. Syst. Appl.*, 2004, 27: 451–458.

Al-Angari, H. M. and Sahakian, A. V. Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *IEEE Trans. Inf Technol. Biomed.*, 2012, 16: 463–468.

Anderer, P., Gruber, G., Parapatics, S. *et al.* An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 $\times$ 7 utilizing the Siesta database. *Neuropsychobiology*, 2005, 51: 115–133.

Babadi, B., Mckinney, S. M., Tarokh, V. and Ellenbogen, J. M. DiBa: a data-driven Bayesian algorithm for sleep spindle detection. *IEEE Trans. Biomed. Eng.*, 2012, 59: 483–493.

Bankman, I. N., Sigillito, V. G., Wise, R. A. and Smith, P. L. Feature-based detection of the K-complex wave in the human electroencephalogram using neural networks. *IEEE Trans. Biomed. Eng.*, 1992, 39: 1305–1310.

Berthomier, C., Drouot, X., Herman-Stoica, M. *et al.* Automatic analysis of single-channel sleep EEG: validation in healthy individuals. *Sleep*, 2007, 30: 1587–1595.

Brunner, D. P., Vasko, R. C., Detka, C. S., Monahan, J. P., Reynolds, C. F. III and Kupfer, D. J. Muscle artifacts in the sleep EEG: automated detection and effect on all-night EEG power spectra. *J. Sleep Res.*, 1996, 5: 155–164.

Campbell, K., Kumar, A. and Hofman, W. Human and automatic validation of a phase-locked loop spindle detection system. *Electroencephalogr. Clin. Neurophysiol.*, 1980, 48: 602–605.

Da Rosa, A. C., Kemp, B., Paiva, T., Lopes Da Silva, F. H. and Kamphuisen, H. A. A model-based detector of vertex waves and K complexes in sleep electroencephalogram. *Electroencephalogr. Clin. Neurophysiol.*, 1991, 78: 71–79.

Devuyst, S., Dutoit, T., Stenuit, P., Kerkhofs, M. and Stanus, E. Cancelling ECG artifacts in EEG using a modified independent component analysis approach. *EURASIP J. Adv. Signal Process.*, 2008, 2008: 747325.

Devuyst, S., Dutoit, T., Stenuit, P. and Kerkhofs, M. Automatic sleep spindles detection—overview and development of a standard proposal assessment method. *Proc. Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, 2011: 1713–1716.

Duman, F., Erdamar, A., Erogul, O., Telatar, Z. and Yetkin, S. Efficient sleep spindle detection algorithm with decision tree. *Exp. Syst. Appl.*, 2009, 36: 9980–9985.

Durka, P. J., Klekowicz, H., Blinowska, K. J., Szelenberger, W. and Niemcewicz, S. A simple system for detection of EEG artifacts in polysomnographic recordings. *IEEE Trans. Biomed. Eng.*, 2003, 50: 526–528.

Erdamar, A., Duman, F. and Yetkin, S. A wavelet and teager energy operator based method for automatic detection of K-Complex in sleep EEG. *Exp. Syst. Appl.*, 2012, 39: 1284–1290.

Figueroa Helland, V. C., Gapelyuk, A., Suhrbier, A. *et al.* Investigation of an automatic sleep stage classification by means of multiscorer hypnogram. *Methods Inf. Med.*, 2010, 49: 467–472.

Gagnon, J. F., Vendette, M., Postuma, R. B. *et al.* Mild cognitive impairment in rapid eye movement sleep behavior disorder and Parkinson's disease. *Ann. Neurol.*, 2009, 66: 39–47.

Hang, L.-W., Su, B.-L. and Yen, C.-W. Detecting slow wave sleep via one or two channels of EEG/EOG signals. *Int. J. Signal Proces. Syst.*, 2013, 1: 84–88.

Huupponen, E., Gomez-Herrero, G., Saastamoinen, A., Varri, A., Hasan, J. and Himanen, S. L. Development and comparison of four sleep spindle detection methods. *Artif. Intell. Med.*, 2007, 40: 157–170.

Huy Quan, V., Gang, L., Sukhorukova, N. S. *et al.* K-complex detection using a hybrid-synergic machine learning method. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, 2012, 42: 1478–1490.

Iber, C., Ancoli-Israel, S., Chesson, A. and Quan, S. F. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. American Academy of Sleep Medicine, Westchester, IL, 2007.

Ince, D. C., Hatton, L. and Graham-Cumming, J. The case for open computer programs. *Nature*, 2012, 482: 485–488.

Jansen, B. H. and Desai, P. R. K-complex detection using multi-layer perceptrons and recurrent networks. *Int. J. Biomed. Comput.*, 1994, 37: 249–257.

Kemp, B. and Olivan, J. European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. *Clin. Neurophysiol.*, 2003, 114: 1755–1761.

Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A. and Oberye, J. J. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.*, 2000, 47: 1185–1194.

Khandoker, A. H., Gubbi, J. and Palaniswami, M. Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings. *IEEE Trans. Inf Technol. Biomed.*, 2009, 13: 1057–1067.

Koley, B. and Dey, D. An ensemble system for automatic sleep stage classification using single channel EEG signal. *Comput. Biol. Med.*, 2012, 42: 1186–1195.

Koley, B. L. and Dey, D. Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers. *Measurement*, 2013, 46: 2082–2092.

Leclercq, Y., Schrouff, J., Noirhomme, Q., Maquet, P. and Phillips, C. fMRI artefact rejection and sleep scoring toolbox. *Comput. Intell. Neurosci.*, 2011, 2011: 598206.

Liu, D., Pang, Z. and Lloyd, S. R. A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG. *IEEE Trans. Neural Netw.*, 2008, 19: 308–318.

Mendez, M. O., Bianchi, A. M., Matteucci, M., Cerutti, S. and Penzel, T. Sleep apnea screening by autoregressive models from a single ECG lead. *IEEE Trans. Biomed. Eng.*, 2009, 56: 228838–228850.

Moretti, D. V., Babiloni, F., Carducci, F. *et al.* Computerized processing of EEG–EOG–EMG artifacts for multi-centric studies in EEG oscillations and event-related potentials. *Int. J. Psychophysiol.*, 2003, 47: 199–216.

O'Reilly, C. and Nielsen, T. Revisiting the ROC curve for diagnostic applications with an unbalanced class distribution. Proceedings of 8th International Workshop on Systems, Signal Processing and their Applications. Algiers, Algeria. 2013: 363–370.

O'Reilly, C. and Nielsen, T. The Spyndle toolbox: an open source software package for the investigation of EEG transient events. *Rev. Int. Technol. Avancées*, 2014. (in press).

Park, H.-J., Oh, J.-S., Jeong, D.-U. and Park, K.-S. Automated sleep stage scoring using hybrid rule- and case-based reasoning. *Comput. Biomed. Res.*, 2000, 33: 330–349.

Park, H. J., Jeong, D. U. and Park, K. S. Automated detection and elimination of periodic ECG artifacts in EEG using the energy interval histogram method. *IEEE Trans. Biomed. Eng.*, 2002, 49: 1526–1533.

Quan, S. F., Howard, B. V., Iber, C. *et al.* The Sleep Heart Health Study: design, rationale, and methods. *Sleep*, 1997, 20: 1077–1085.

Rechtschaffen, A. and Kales, A. Eds. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. US National Institute of Neurological Diseases and Blindness, Neurological Information Network, Bethesda, MD, 1968.

Richard, C. and Lengelle, R. Joint time and time–frequency optimal detection of K-complexes in sleep EEG. *Comput. Biomed. Res.*, 1998, 31: 209–229.

Schimicek, P., Zeitlhofer, J., Anderer, P. and Saletu, B. Automatic sleep-spindle detection procedure: aspects of reliability and validity. *Clin. Electroencephalogr.*, 1994, 25: 26–29.

Schonwald, S. V., De Santa-Helena, E. L., Rossatto, R., Chaves, M. L. and Gerhardt, G. J. Benchmarking matching pursuit to find sleep spindles. *J. Neurosci. Methods*, 2006, 156: 314–321.

Shambroom, J. R., Fabregas, S. E. and Johnstone, J. Validation of an automated wireless system to monitor sleep in healthy adults. *J. Sleep Res.*, 2012, 21: 221–230.

Tagluk, M. E., Sezgin, N. and Akin, M. Estimation of sleep stages by an artificial neural network employing EEG, EMG and EOG. *J. Med. Syst.*, 2010, 34: 717–725.

Tang, Z. and Ishii, N.. Detection of the K-complex using a new method of recognizing waveform based on the discrete wavelet transform. *IEICE Trans. Inf. Syst.*, 1995, E78-D: 77–85.

Ventouras, E. M., Monoyiou, E. A., Ktonas, P. Y. *et al.* Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: a feasibility study. *Comput. Methods Programs Biomed.*, 2005, 78: 191–207.

Virkkala, J., Hasan, J., Varri, A., Himanen, S. L. and Muller, K. Automatic sleep stage classification using two-channel electro-oculography. *J. Neurosci. Methods*, 2007, 166: 109–115.

Warby, S. C., Wendt, S. L., Welinder, P. *et al.* Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nat. Methods*, 2014, 11: 385–392.