Stimulating Rapid Research Advances Via Focused Competition: The Computers in Cardiology Challenge 2000

GB Moody[1,2], RG Mark[1,2], AL Goldberger[1], T Penzel[3]

[1]Harvard-M.I.T. Division of Health Sciences and Technology, Cambridge, MA, USA
[2]Beth Israel Deaconess Medical Center, Boston, MA, USA
[3]Department of Internal Medicine, Philipps-University, Marburg, Germany

Abstract

Obstructive sleep apnea is associated with a variety of serious health risks. Standard methods for detecting and quantifying sleep apnea are based on respiration monitoring, but previous studies have suggested that apnea detection based on the ECG alone might be possible. We therefore offered a challenge to the research community, to demonstrate the efficacy of ECG-based methods for apnea detection using a large, well-characterized, and representative set of data made freely available via the Internet. The goal of the contest was to stimulate effort and advance the state of the art in this clinically significant problem, and to foster both friendly competition and wide-ranging collaborations. The event was an outstanding success, with most entrants achieving 90% to 100% accuracy in identifying subjects with significant apnea, and minute-by-minute apnea detection accuracy between 85% and 93%, comparable to the concurrence of human experts scoring full polysomnograms.

1. Introduction

Between February and September 2000, PhysioNet [1] hosted the Computers in Cardiology Challenge 2000, an open competition focusing on the problem of detecting and quantifying obstructive sleep apnea (intermittent cessation of breathing) using the ECG. We selected this topic for the first in a planned series of challenges for several reasons:

- Obstructive sleep apnea is a common problem associated with major health implications ranging from excessive daytime drowsiness to serious cardiac arrhythmias, high blood pressure, myocardial infarction, and stroke, and with increased mortality rates.
- Respiration monitoring, the standard method for detecting and quantifying sleep apnea, often disturbs or interferes with sleep and is generally expensive.
- Studies during the past 15 years [2–4] have hinted at the possibility of detecting sleep apnea using features of the ECG.

Apnea detection methods based on analyses of the ECG are minimally intrusive, inexpensive, and may be particularly well-suited for screening. These methods may exploit respiratory sinus arrhythmia, beat-to-beat variations in waveform morphology related to motion of the ECG electrodes relative to the heart, or both of these phenomena. The major obstacle to use of such methods has been that careful quantitative comparisons of their accuracy against that of conventional techniques for apnea detection have not been published.

2. The Apnea-ECG Database

The immediate goal of the challenge was to demonstrate that apnea can be detected reliably from the ECG in a large, representative, and well-characterized set of reference data. Data for this contest were provided by the fourth author, and are available at http://www.physionet.org/physiobank/database/apnea-ecg/.

The data used in the contest were divided into a learning set and a test set of equal size. Each set consists of 35 recordings, containing a single ECG signal digitized at 100 Hz with 12-bit resolution, continuously for approximately 8 hours (individual recordings vary in length from slightly less than 7 hours to nearly 10 hours). Each recording includes a set of reference annotations, one for each minute of the recording, that indicate the presence or absence of apnea during that minute. These reference annotations were made by human experts on the basis of simultaneously recorded respiration signals. The reference annotations for the test set were not made available before the conclusion of the contest. Eight of the recordings in the learning set include three respiration signals (oronasal airflow measured using nasal thermistors, and chest and abdominal respiratory effort measured using inductive plethysmography) each digitized at 20 Hz, and an oxygen saturation signal digitized at 1 Hz. These additional signals were made available as reference material to understand how the apnea annotations were made, and to study the relationships between the respiration and ECG signals. The database is further described in a companion article in these proceedings [5].

0276-6547/00 $10.00 © 2000 IEEE

The database contains neither pure central apnea nor Cheyne-Stokes respiration; all apneas in these recordings are either obstructive or mixed. Minutes containing hypopneas (defined as intermittent drops in respiratory flow below 50%, accompanied by drops in oxygen saturation of at least 4%, and followed by compensating hyperventilation) are also scored as minutes containing apnea. The subjects of these recordings are men and women between 27 and 60 years of age, with weights between 53 and 135 kg; AHI (apnea-hypopnea index) ranges from 0 to 93.5 in these recordings.

2.1. Sleep apnea definitions

Several definitions for clinically significant sleep apnea have been in clinical use since 1978, when Guilleminault [6] defined "sleep apnea syndrome" as more than 30 apneas per night. In 1981, Lavie [7] proposed a more selective criterion of 100 apneas per night. Later criteria were based on an "apnea index" (the number of apneas per hour, or the number of minutes containing apnea per hour). Most clinicians regard an apnea index below 5 as normal, and an apnea index of 10 or more as pathologic. In 1988, He et al. [8] found increased mortality in untreated patients with apnea indices of 20 or more, and such patients are now recognized as in need of treatment. Criteria used in current practice rely not only on an apnea index, but also on symptoms and cardiovascular sequelae.

2.2. Data classes

For the purposes of this challenge, based on these varied criteria, we have defined three classes of recordings:

- **Class A (Apnea)**: These meet all criteria. Recordings in class A contain at least one hour with an apnea index of 10 or more, and at least 100 minutes with apnea during the recording. The learning and test sets each contain 20 class A recordings.
- **Class B (Borderline)**: These meet some but not all of the criteria. Recordings in class B contain at least one hour with an apnea index of 5 or more, and between 5 and 99 minutes with apnea during the recording. The learning and test sets each contain 5 class B recordings.
- **Class C (Control)**: These meet none of the criteria, and may be considered normal. Recordings in class C contain fewer than 5 minutes with apnea during the recording. The learning and test sets each contain 10 class C recordings.

The use of a common data set consisting of ECG recordings that have been extracted from manually-annotated polysomnograms and divided into equal training and test sets, permits reproducible evaluation and fair comparison of different analysis methods. Apnea annotations for the training set were made available for study; those for the test set have not been released at the time of writing, pending the conclusion of follow-up studies by several of the entrants in the challenge.

It should be noted that the test set contains two nearly identical recordings, designated as x33 and x34, as a result of an error in the selection process. This error is not likely to have had any significant effect on the difficulty of the challenge, however.

Entrants competed in one or both of the following events:

- **Screening**: The object was to design software to classify the 35 test set recordings into class A (apnea) and class C (control or normal) groups, using the ECG signal to determine if significant sleep apnea was present. Classifications for the 5 class B (borderline) recordings did not influence scores in this event (but entrants had to classify them into either class A or class C, since the identity of the class B records was not known to them). The score for this event was simply the number of correct classifications; thus the maximum score possible was 30.
We chose to exclude the class B recordings from the calculation of the scores because the utility of a screening test depends primarily on the accuracy with which it classifies the unambiguous cases, both positive and negative (classes A and C respectively in this instance).
- **Quantification**: In this event, entrants designed software to generate a minute-by-minute annotation file for each recording, in the same format as those provided with the learning set, using the ECG signal to determine when sleep apnea occurs. The generated annotations were then compared with a set of reference annotations to determine the score. Each annotation that matched a reference annotation earned one point; thus the highest possible score for this event was 17268 (the total number of reference-annotated minutes in the 35 test set records). It is important to understand that scores approaching the maximum are very unlikely, since apnea assessment can be very difficult even for human experts. Nevertheless, the scores provide a reasonable ranking of the ability of the respective algorithms to mimic the decisions made by human experts.

Entrants submitted their classifications to an automatic scorer on PhysioNet, and received their scores by return e-mail. Only the scores themselves were returned; neither the identities of the misclassified records nor the types of errors were revealed to the entrants. Entrants were permitted to submit a limited number of additional entries in an attempt to improve their scores, and many did so.

The data were initially made available on PhysioNet on 10 February 2000. Scoring of entries began on 12 April, and revised entries were accepted until noon GMT on 22 September.

3. Results

Four entrants achieved perfect scores of 30/30 in the screening event, and the first place award went to Murray Jarvis [9], who was the first to obtain this result.

In the quantification event, competition among the three top finishers was intense during the final two weeks, and the winning entry, from James McNames [10] with a score of 15994/17268, or 92.62%, was submitted only 12 hours before the deadline. The final entry by Ben Raymond [11], only 65 minutes before the deadline, nearly matched the winning score. Both Raymond and McNames had previously received perfect scores in the screening event, as had Philip de Chazal [12], who was the only entrant to do so on the first attempt.

Only one entrant attempted to identify the 5 class B records in the screening event. Phyllis Stein [13] achieved a remarkable score of 33/35 correct classifications; the only errors were misclassification of a class A record into class B and vice versa.

Details of the methods used were presented by thirteen entrants at Computers in Cardiology 2000 and appear in these proceedings [9–21].

The more profound goal of this challenge was to leverage the opportunities presented by the PhysioNet resource to foster rapid progress by a widely distributed group of researchers on an important clinical problem. A major benefit of a research resource such as PhysioNet is to create a meeting place for new data and novel analytic methods. It is especially significant that several of the most successful entrants would not otherwise have had access to the data necessary for studying this topic. The spectacular success of this challenge is among the clearest demonstrations yet of how PhysioNet offers a new paradigm for catalyzing advances in subjects where access to data has traditionally been a barrier to entry.

Acknowledgements

GBM, RGM, and ALG thank our coauthor and esteemed colleague, Dr. Thomas Penzel, who made this challenge possible by his generous contributions of carefully selected data, reference annotations, expertise, and thoughtful advice and comments throughout the past months. We received the wholehearted and constant support of the board and members of Computers in Cardiology from the outset, for which we are grateful. The challenge was conducted using the facilities of PhysioNet, a public service of the Research Resource for Complex Physiologic Signals, which is supported by a grant from the National Center for Research Resources of the US National Institutes of Health (P41 RR13622). Awards were contributed by the Margret and H.A. Rey Laboratory for Nonlinear Dynamics in Medicine at Boston's Beth Israel Deaconess Medical Center.

Without the outstanding efforts of the entrants, we would have nothing to report here. Fifteen teams in nine nations took up the challenge and exceeded our expectations in every way in a span of seven months. Thirteen submitted abstracts for publication in these proceedings, and without exception they were judged of high or very high quality by the abstract review committee; the papers written by these entrants appear elsewhere in these proceedings. We thank them, and heartily congratulate them!

References

[1] Moody GB, Mark RG, Goldberger AL. PhysioNet: A research resource for studies of complex physiologic and biomedical signals. In Computers in Cardiology 2000; 2000. [this volume].

[2] Moody GB, Mark RG, Zoccola A, Mantero S. Derivation of respiratory signals from multi-lead ECGs. In Computers in Cardiology 1985. Washington, DC: IEEE Computer Society Press, 1986; 113–116. [http://www.physionet.org/physiotools/edr/cic85/].

[3] Moody GB, Mark RG, Bump MA, Weinstein JS, Berman AD, Mietus JE, Goldberger AL. Clinical validation of the ECG-derived respiration (EDR) technique. In Computers in Cardiology 1986. Washington, DC: IEEE Computer Society Press, 1987; 507–510. [http://www.physionet.org/physiotools/edr/cic86/].

[4] Pallas-Areny R, Canals Riera F. Recovering the respiratory rhythm out of the ECG. Medical and Biological Engineering and Computing 1985;23((Suppl. part 1)):338–339.

[5] Penzel T, Moody GB, Mark RG, Goldberger AL, Peter JH. The apnea-ecg database. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000. [this volume].

[6] Guilleminault C, van den Hoed J, Mitler MM. Clinical overview of the sleep apnea syndromes. In Guilleminault C, Dement WC (eds.), Sleep Apnea Syndromes. New York: Liss, 1978; 1–12.

[7] Lavie P, Halperin E, Zomer J, Alroy G. Across-night lengthening of sleep apneic episodes. Sleep 1981 (September); 4(3):279–282.

[8] He J, Kryger MH, Zorick FJ, Conway W. Mortality and apnea index in obstructive sleep apnea. Experience in 385 male patients. Chest 1988 (July);94(1):9–14.

[9] Jarvis MR, Mitra PP. Multitaper spectral estimation reveals excess power at 0.02 Hz which characterizes apnea patients. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[10] McNames J, Fraser A, Rechtsteiner A. Sleep apnea classification based on frequency of heart-rate variability. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[11] Raymond B, Cayton R, Bates R, Chappell M. Screening for obstructive sleep apnoea based on the ECG: The CinC challenge. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[12] de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automatic detection of apnea with the electrocardiogram. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[13] Stein PK, Domitrovich PP. Detecting OSAHS from patterns

seen on heart-rate tachograms. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[14] Drinnan MJ, Allen J, Langley P, Murray A. Detection of sleep apnoea from frequency analysis of heart-rate variability. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[15] Schrader M, Zywietz C, von Einem V, Widiger B, Joseph G. Detection of sleep apnea in single channel ECGs from the PhysioNet data base. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[16] Marchesi C, Paoletti M, Gaetano SD. Global waveform delineation for RR series estimation: Detecting the sleep apnea pattern. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[17] Ng F, Garcìa I, Gomis P, Passariello G, Mora F. Bayesian hierarchical model with wavelet transform coefficients of the ECG in the obstructive sleep apnea screening. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[18] Ballora M, Glass L, Pennycook B. Detection of obstructive sleep apnea through auditory display of heart-rate variability. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[19] Maier C, Bauch M, Dickhaus H. Recognition and quantification of sleep apnea by analysis of heart-rate variability parameters. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[20] Mietus JE, Peng CK, Goldberger AL. Detection of obstructive sleep apnea from cardiac interbeat interval time series. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

[21] Shinar Z, Baharav A, Akselrod S. Detection of obstructive sleep apnea from the ECG. In Computers in Cardiology 2000. Piscataway, New Jersey: IEEE Press; 2000.

Address for correspondence:

George B. Moody
MIT Room E25-505A, Cambridge, MA 02139 USA.
george@mit.edu