Journal of Clinical Monitoring and Computing (2009) 23:1–10
DOI: 10.1007/s10877-008-9154-8

VALIDATION OF FOREHEAD VENOUS PRESSURE AS A MEASURE OF RESPIRATORY EFFORT FOR THE DIAGNOSIS OF SLEEP APNEA

Djordje Popovic, MD[1], Christopher King, MD[2], Melanie Guerrero, MD[2], Daniel J. Levendowski, MBA[3], Delmer Henninger, MD[4] and Philip R. Westbrook, MD[3]

Popovic D, King C, Guerrero M, Levendowski DJ, Henninger D, Westbrook PR. Validation of forehead venous pressure as a measure of respiratory effort for the diagnosis of sleep apnea.
J Clin Monit Comput 2009; 23:1–10

ABSTRACT. Objectives. The aim of the study was to validate the measurement of Forehead Venous Pressure derived from a single site on the forehead as an alternative to esophageal manometry and respiratory effort bands in the differential diagnosis of sleep apnea. **Methods.** Fourteen subjects underwent a laboratory polysomnography concurrently with ARES Unicorder at Walter Reed Army Medical Center. Two-hundred respiratory events were selected by a scorer boarded in sleep medicine and classified into six event categories used in the differential diagnosis of sleep disordered breathing. Four sets of events were prepared, each containing airflow and one of four measures of respiratory effort (i.e., esophageal manometer, chest and abdomen bands, and forehead venous pressure). A second board-certified scorer scored each set of events twice while blinded to the type of the effort signal. **Results.** The inter-rater Kappa scores across all event types indicated all four effort signals provided moderate agreement (κ = 0.43–0.47). When comparing the intra-rater Kappa scores, the chest belt was superior (κ = 0.88) to the esophageal manometry, FVP and abdomen belt (κ = 0.78–0.82). The Kappa scores for the intra-rater comparison with the esophageal serving as the gold standard, FVP abdomen and chest all showed near perfect agreement (κ = 0.81–0.86). The esophageal manometer and FVP provided slightly better inter-rater agreement in the detection of both obstructive hypopneas and apneas as compared to the chest and abdomen belts. There was a 20–30% drop in inter-rater reliability in the detection of flow-limitation and ventilation-change events compared to obstructive events, and all effort signals showed poor inter-rater agreement for central and mixed events. **Conclusions.** The results of the study suggest that the FVP can serve as an alternative to respiratory bands in the differential diagnosis of sleep disordered breathing, and in the recognition of patients appropriate for bilevel continuous positive airway pressure devices.

KEY WORDS. respiratory effort, sleep disordered breathing, central sleep apnea, OSA, effort bands, venous pressure, photoplethysmography.

From the [1]University of Southern California, Los Angeles, CA, USA; [2]Walter Reed Army Medical Center, Washington, DC, USA; [3]Advanced Brain Monitoring, Inc, 2237 Faraday Avenue, Suite 100, Carlsbad, CA 92008, USA; [4]Complete Sleep Solutions, Inc, Murrieta, CA, USA.

Received 16 September 2008. Accepted for publication 25 November 2008.

Address correspondence to D. J. Levendowski, Advanced Brain Monitoring, Inc, 2237 Faraday Avenue, Suite 100, Carlsbad, CA 92008, USA.
E-mail: Dan@b-alert.com

INTRODUCTION

Respiratory effort is routinely used in combination with airflow to diagnose patients with sleep-disordered breathing. The effort measurement clarifies mechanisms that underlie the observed changes in airflow thereby enabling an accurate distinction among physiological and various pathological respiratory patterns. A sustained or increasing effort to breathe during a significant reduction

or complete cessation of airflow points to obstructive sleep apnea whereas its absence, at least at the beginning of the respiratory event, is characteristic of central or complex apneas. More importantly, frequent episodes of an increase in respiratory effort terminated with an EEG arousal in the absence of any significant change in airflow or arterial oxygen saturation are the only indicator of upper airway resistance syndrome. Therefore, measurement of respiratory effort is an important component in providing a differential diagnosis of sleep disordered breathing.

Esophageal manometry is generally accepted as the gold standard for measuring respiratory effort. The method is rarely used in clinical practice due to its invasiveness and discomfort to the patient, and technical expertise needed to acquire high quality signals. Reports suggest that esophageal catheterization modifies the pharyngeal dynamics [1] and impairs the quality of sleep [2, 3]. Therefore, most sleep studies are conducted with respiratory bands placed around the chest and abdomen to monitor respiratory effort. Inductive plethysmography, piezo electric crystals, conductive elastomeres, magnetometers and strain gauges have all been used to measure thoracic cage expansion. The bands are generally comfortable and do not require special training to apply but these techniques are not without drawbacks. The quality of recorded signal depends on the sensitivity of transducers used, and these devices tend to fail in obese subjects in whom respiratory excursions are small compared to the dimensions of the thorax and abdomen. A sub-set of these devices (i.e., Piezo-belts and strain gauges but not the RIP bands) are prone to trapping artifact as a subject turns from one side to another which may significantly affect the recorded amplitude of respiratory effort. When used in an in-home environment, the ease of self-application should be considered since it can influence failure rates. Two studies found the failure rate for effort bands ranged between 7% and 21% even though the effort bands were either applied by a technician in the home [4] or pre-fitted by a technician in the lab [5]. There is little to no information about the failure rates of effort bands self-applied by the patient and used in an unattended setting.

A number of alternative methods have been studied to obtain a qualitative measurement of respiratory effort although none are widely used in the clinical practice [6]. Pulse transit time has shown a high sensitivity and specificity in differentiating obstructive and central apneas [7] and in the detection of non-apneic respiratory events [8] but it is not ideal for unattended recordings as it requires the training for application of EKG electrodes (which would be as difficult to apply as respiratory effort belts). The diaphragmatic electromyogram measured transcutaneously correlates well with esophageal pressure but it shares the same drawback as the pulse transit time with respect to the need for self-application of electrodes. Although the presence of flow limitation in nasal flow signal has been shown to be useful in identifying upper airway resistance [9–11] and the absence of flow limitation in combination with a reduction in tidal volume indicates a reduction in effort, it can be compromised in the presence of significant oral breathing [12].

More recently, respiratory induced intensity variations of the photoplethysmographic (RIIV-PPG) signal have been used to monitor respiratory rate [13–16]. Although the physiological background of the RIIV-PPG is not fully understood, variations in venous return to the heart caused by the alterations in intrathoracic pressure are believed to be a dominant mechanism [17–19]. The peak-to-peak amplitude of RIIV-PPG is closely correlated with tidal volume [19, 20] and respiration-related changes in peripheral venous pressure [21] and arterial blood pressure [22]. The latter correlation is important since drop in systolic blood pressure that occurs with inspiration (pulsus paradoxus) has been shown to correlate well with the degree of inspiratory effort in simulated obstructive sleep apnea [23]. Taken together, these findings suggest that the amplitude of RIIV-PPG may correlate with changes in intrathoracic pressure, and therefore could be used at least as a qualitative measure of respiratory effort.

The goal of this study was to assess whether changes in RIIV-PPG in combination with additional physiological signals can provide physiologically meaningful and clinically relevant information about changes in respiratory effort accompanying physiological and pathological respiratory events during sleep equivalent to esophageal manometry and respiratory effort bands.

METHODS AND MATERIALS

Forehead venous pressure signal

The forehead venous pressure (FVP) signal used in this study to measure respiratory effort is derived from a combination of physiological signals obtained from a recorder affixed to the forehead (ARES^{TM} Unicorder, Advanced Brain Monitoring, Carlsbad, CA [24]). The respiratory induced intensity variations of the photoplethysmographic (RIIV-PPG) signal are obtained using red and infrared light emitting diodes and a photodiode encapsulated in medical grade silicone and sampled at 100 Hz. A piezoresistive silicone absolute pressure sensing chip embedded in the reflectance sensor is sampled at 10 Hz to measure changes in surface pressure. A 3-axis MEMS accelerometer mounted horizontal to the

forehead is sampled at 10 Hz to measure subtle motions associated with respiration. The signals are recorded and saved to a flash card for subsequent off-line processing. Acquisition of the FVP signal requires no additional training on the part of the patient. The optical signals are monitored during acquisition and audio feedback is provided to the patient when required to ensure high quality signals are recorded.

The FVP is a composite signal consisting of the components derived from the photoplethysmographic sensor, accelerometers and pressure transducer by means of adaptive digital filtering (Figure 1). Reflectance infrared and red PPG signal (''optical signals''), three signals coming from the three axis of the accelerometer (''motion'') and the signal from the forehead pressure transducer (''pressure'')are subject to a zero-phase band-pass filtering. The low and high pass filter are selected by the algorithm from a bank of pre-designed filters based on median value and standard deviation of breathing rate across the whole record. Breathing rate is determined from the airflow signal recorded simultaneously with the composite signals. The filter bank contains high- and low-pass infinite-impulse response (IIR) filters with cutoff frequencies from 0.05 to 0.5 Hz in steps of 0.0167 Hz which corresponds to breathing rates of 3–30 breaths per minute in steps of 1 breath per minute. Such a wide range of available filters was deemed necessary to account for the full spectrum of (instantaneous) breathing rates that can substantially vary across types of respiratory events, subjects and age groups ranging from near-zero during long apneas to almost thirty in children and young adolescents.

Prior to combining the signals into the composite measurement, the band-pass filtered signals are re-sampled (10 samples/s) and checked for the presence of artifacts. Artifacts are identified as segments in which the envelope of any component exceeds a threshold that is based on the average variance of the component across the whole record, and are subsequently removed by setting the signal values to zero across the identified segments. These segments are labeled as containing artifact in the software which also makes the later review and interpretation of data easier and less prone to errors. Finally, the artifact-corrected components are multiplied with weighting coefficients that determine their contribution to the resultant composite and combined into the FVP. The strongest contributor is the infrared optical signal (IR PPG) that accounts for at least 80% of the FVP amplitude and shape. The contribution of the other components is significant only in the presence of artifacts in the IR PPG signal, and depends also on the head position. In the supine position the RED PPG contributes most, whereas when the head is in the lateral position the forehead pressure transducer provides the most valuable information. The motion signals become most significant towards the end of long obstructive events while the head is lying supine when an increase in the amplitude of the respiratory-synchronous head movements can be observed just prior to the event termination.

Study design

Seventeen adult subjects (9 males) who had been referred to the Walter Reed Army Medical Center (Washington, DC) for an overnight polysomnography (PSG) because of excessive daytime somnolence or suspected obstructive sleep apnea were recruited for the study. Each subject underwent an attended overnight polysomnography (Alpha Somnostar system, Sensormedics, Yorba Linda, CA) and a concurrent recording with the ARES Unicorder. The PSG consisted of continuous recordings of central and occipital electroencephalograms (EEG), bilateral electro-oculograms (EOG), submental and bilateral tibial electromyograms (EMG) and electrocardiogram (EKG). Nasal and oral airflow are measured by both thermistor and a nasal cannula connected to a pressure transducer. A single cannula was placed on the patient, and the tubing was split to

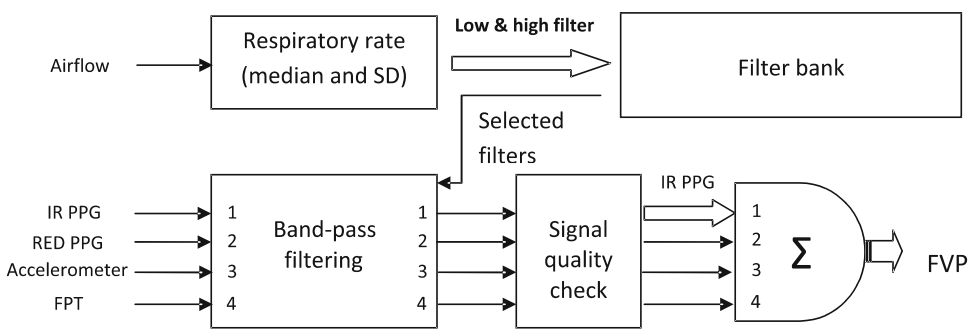

Fig. 1. Diagram for acquisition of the forehead venous pressure with the increased contribution of the infrared PPG signal emphasized by the thicker white arrow and the other composite signals represented with thin black arrows.

Table 1. Demographic and PSG data (N = 14)

Variable	Mean ± SD
Age (years)	48.0 ± 7.2
BMI (kg/m^2)	29.3 ± 4.1
Epworth sleepiness scale	10.8 ± 5.3
RDI (events/hour)	11.8 ± 13.2
Total sleep time-TST (min)	266 ± 88

provide pressure inputs to both the pressure transducer of the PSG system and the ARES Unicorder. Respiratory effort was directly measured with a multi-port esophageal catheter (Gaeltec Ltd, Hackensack, NJ) whereas thoracic and abdominal excursions were recorded with piezo-electric belts of several different brands. Continuous oxygen saturation is assessed using non-invasive pulse oximetry. Body positioning is verified by infrared video recording. Prior to lights out, the esophageal transducer was calibrated at various pressures using a syringe attached to a manometer and by comparing the computer digital reading to the manometer reading with −50 cm H_2O the lowest and +50 cm H_2O the highest values.

The majority of the PSG recordings lasted for at least 6 h, however, four studies were interrupted after two to three hours to perform a CPAP titration. The data from three subjects were not used in the analysis due to: an improperly placed nasal cannula resulting in less than 2 h of valid airflow signal, a lost PSG record due to hard disk failure; and failure of an ARES Unicorder. Basic demographic data and sleep and respiratory indices for the remaining 14 subjects (8 males and 6 females) whose recordings were used in this report are shown in Table 1. Five subjects had mild to moderate sleep apnea syndrome (RDI between 5 and 30), one had severe OSA with an RDI of 34, two subjects had upper airway resistance syndrome, two were diagnosed with primary snoring and four had a RDI < 5. Only seven of the fourteen subjects exhibited central or mixed events, all with less than five events/hour. None of the subjects reported a significant cardiac disease other than controlled hypertension which might otherwise have affected central venous pressure changes.

Respiratory event selection

The ARES and PSG recordings were synchronized off-line using the airflow signal from the split nasal cannula. Three hundred candidate events were chosen (20–25 events per subject) based only on the amplitude and shape of the airflow signal by a clinician who was not board certified in sleep medicine. Care was taken to equally distribute selected events across the diagnostic portion of the recording period (i.e. the selected events were uniformly distributed across the night) whenever possible. A board certified sleep medicine specialist with over 30 years of experience (Rater 1) examined the airflow signal, esophageal pressure, respiratory belts and FVP in each of the candidate events by and classified them into one of the six categories as described in Table 2. Rater 1 discarded events from further analysis if they did not clearly belong to one of the six categories, or if the signal quality of any of the four effort measures (esophageal pressure, thoracic and abdominal belts, FVP) was deemed low. A total of 100 events were discarded: 33 could not be classified with certainty into one of the six categories whereas in 67 events there were signal quality problems with one or more effort measures (i.e., 27 because of the FVP, 14 because of the esophageal catheter, seven because of the thoracic and six because of the abdomen belt, and 13 because of the problems in two or more signals). The 200 acceptable events consisted of 104 obstructive hypopneas, 19 obstructive apneas, 18 central events (5 apneas and 13 hypopneas), 6 complex apneas, 15 periods of flow limitation and 38 events representing physiological variability in ventilation.

Blinded validation

The two hundred events accepted by Rater 1 were organized into four sets. Each set contained the airflow signal and one effort measure: esophageal manometry (esophageal), chest band (chest), abdomen band (abdomen) or forehead venous pressure (FVP). For the four sets of data, events were presented on a 120 s screen with marks identifying the beginning and end of the event (Figure 2). A second board certified sleep medicine specialist with over 15 years of experience (Rater 2) applied the rules presented in Table 2 to classify the events in each set while blinded to the type of effort signal. To reduce the likelihood of errors, only one set was scored per day and each set was scored twice with at least 1 week between the rounds.

Data analysis and interpretation

Inter-rater agreement was assessed for each effort measure separately by comparing the categories assigned to events by Rater 2 to those of Rater 1. Agreement percentages were calculated separately for Round 1 and Round 2 and then averaged. Intra-rater agreement was calculated similarly for each effort measure by comparing the categories assigned to events by Rater 2 on Round 2 to those from Round 1. To further assess the discriminatory power of the various effort measures, we eliminated the influence of

Table 2. Definition of sleep disordered breathing event types

Event category	Airflow pattern	Expected effort pattern
Obstructive hypopnea	>50% decrease in amplitude lasting >10 s with clear termination (strong breath and/or movement)	Increase in effort starts 2 or more breaths prior to event termination
Flow limitation	>30 s of flow limitation without a clear termination	Increasing
Obstructive apnea	>10 s of a complete cessation of airflow	Increase in effort begins ≥2 breaths prior to resumption of flow, and peaks before the peak in airflow
Change in ventilation	Variation of amplitude and/or frequency of airflow such as those seen on sleep onset or in REM sleep	Effort signal should mirror the changes in airflow
Central apnea or hypopnea	Same as for obstructive apnea or hypopnea except that there should be no flow limitation	Changes in effort synchronous with changes in flow, or the increase in effort starts 1 breath prior to resumption of flow
Mixed apnea or hypopnea	Same as for obstructive apnea/hypopnea	Effort signal decreases like a central and increases like an obstructive event

all inter- and intra-rater variability by only analyzing those events which Rater 2 had classified into the same category on Round 1 and 2 (i.e. for each effort measure there could be no difference between Round 1 and Round 2). Event-by-event comparisons were then performed comparing the FVP, chest and abdominal bands against the event classification selected using the esophageal pressure signal (i.e., gold standard). Overall inter- and intra-rater agreements as well as agreements per event category were reported along with the corresponding kappa statistics. Kappa values >0.80 were interpreted to indicate near perfect, between 0.61 and 0.80 indicated substantial agreement, and between 0.41 and 0.60 moderate agreement [25]. For the inter- and intra-rater comparisons, percentage agreements ≥95% were considered near perfect, agreements between 80 and 94% were considered substantial, between 60 and 79% moderate agreement, between 40 and 59% modest agreement and <40% were considered poor agreement. Comparisons were made to identify the superior effort measure(s), i.e., those which provided consistently higher percentages of agreement within and between the raters across most or all event types.

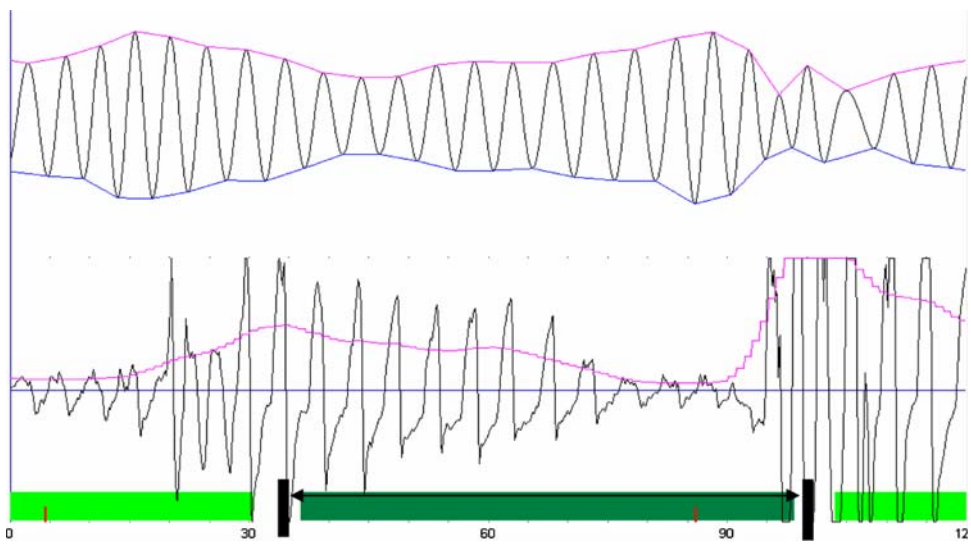

Fig. 2. Example of an event which was provided to Rater 2 for classification, the upper tracing is the effort measure (i.e., FVP) and the lower tracing is the airflow signal. The black marks and line identify the event region that requires classification.

Table 3. Inter-rater event comparison based on the percent agreement by event type

	# Events	Obstructive hypopnea 104 (%)	Flow limitation 15 (%)	Obstructive apnea 19 (%)	Ventilation change 38 (%)	Central events 18 (%)	Mixed events 6 (%)	Overall agreement 200
Effort signal used by Rater 2	Esophageal	80	58	80	50	39	15	65% $\kappa = 0.47$
	FVP	83	68	80	46	18	0	66% $\kappa = 0.47$
	Chest	80	57	77	29	25	15	61% $\kappa = 0.43$
	Abdomen	76	58	70	63	33	0	65% $\kappa = 0.47$

RESULTS

Inter-rater comparison

The overall agreement between Raters 1 and 2 across all event types was moderate for each of the effort signals (Table 3). Given that obstructive events made up to 61% of all events, the absence of more prominent differences in overall agreement mostly reflects the fact that the agreement on the obstructive events (apneas or hypopneas) was high and was not affected by the effort signal type. The raters agreed on the flow limitation and ventilation changes in about half of the cases with the FVP being somewhat superior to the other effort signals in the detection of flow limitations, and the chest being inferior in the detection of ventilation changes. The agreement between the raters on the central events was low with the esophageal showing the best and FVP the least agreement. Most events classified as central by Rater 1 were regarded as obstructive hypopneas or ventilation changes by Rater 2. Mixed events were mostly misclassified as obstructive (due to the small number of events, a 15% agreement corresponds to only 1 event).

Intra-rater comparison

The intra-rater agreement (Round 1 vs. Round 2) was substantially higher than the inter-rater agreement (Table 4). There was very high agreement for obstructive apneas and hypopneas, and changes in ventilation with no important differences across effort types. Less consistent identification of flow limitation and central events was apparent with the FVP signal, with misclassifications typically called central in one round and obstructive hypopneas or ventilation changes the other round. Only the esophageal signal provided consistent identification of mixed events. None of the mixed events were consistently recognized using the abdomen or FVP signals.

Table 5 provides results based on the 141 respiratory events similarly classified by Rater 2 in both rounds (Table 5), and compares classifications made with esophageal pressure and airflow as the gold standard against the other three effort measures. The patterns were similar to those reported for inter- and intra-rater reliability. The agreement was again lower for central events, with the FVP providing the poorest result. The chest signal had the poorest agreement for ventilation changes and inconsistent identification of central events. There were no mixed events in this subset due to inconsistent identification in the previous round.

DISCUSSION

The findings from this study provide face validity to support the use of forehead venous pressure (FVP) to distinguish clinically relevant classes of respiratory events associated with sleep disordered breathing. The two raters exhibited substantial agreement in detecting obstructive hypopneas using all effort measures except the abdomen.

Table 4. Intra-rater event comparison based on the percent agreement by event type

	Obstructive hypopnea (%)	Flow limitation (%)	Obstructive apnea (%)	Ventilation change (%)	Central events (%)	Mixed events (%)	Overall agreement
Esophageal	89	65	93	100	94	100	89% $k = 0.82$
FVP	94	50	92	100	50	N/A	86% $k = 0.79$
Chest	95	82	93	91	96	33	93% $k = 0.88$
Abdomen	84	73	94	100	68	N/A	85% $k = 0.78$

Table 5. Percent agreement with esophageal pressure as the gold standard versus the other effort measures

# Events	Obstructive Hypopnea 86 (%)	Flow limitation 15 (%)	Obstructive apnea 19 (%)	Ventilation change 13 (%)	Central 8 (%)	Overall agreement 141
FVP	96	100	100	89	38	92% $k = 0.86$
Chest	94	100	89	69	62	89% $k = 0.81$
Abdomen	91	100	95	93	75	91% $k = 0.85$

The esophageal manometer and FVP provided substantial agreement in detection of obstructive apneas. There was at least substantial intra-rater agreement across all effort measures for both obstructive apneas and hypopneas. We found the abdomen belt provided the lowest intra- and inter-rater reliability for obstructive hypopneas.

On average there was a 20% drop in agreement between Raters 1 and 2 across the effort measures from obstructive apneas and hypopnea to flow limitation events, and a 30% drop in consistent recognition of ventilation change events. When detecting flow limitation events, the FVP provided moderate agreement while the other three measures provided modest agreement. For changes in ventilation, the abdomen band provided moderate agreement, the esophageal manometer and FVP showed modest agreement, and the chest belt poor agreement.

Across all effort measures, there was poor inter-rater agreement in the detection of central and mixed events. This finding should be interpreted with caution because there were very few events available for selection, most of the central events were 10–20 s hypopneas and only 5 were true central apneas. It is likely that lower reliability for these events was attributed to the difficulty in assessing the timing of the decrease and increase in effort and flow during the short duration central hypopneas (as opposed to recognizing the presence or absence of effort in a central apnea event). It would be expected that distinguishing short pathologic hypopneas of central origin from physiologic respiratory variations such as those seen in REM or at sleep onset would be even more difficult without access to the hypnograms which were not available to the raters. This point is illustrated by the fact that the two raters disagreed in 8 out of 13 central hypopneas, while both raters agreed in 4 out of 5 central apnea events while Rater 2 was using esophageal signal.

Not withstanding the limitations of the data set, the FVP provided the lowest inter- and intra-observer agreement among the available effort signals. We partially attribute the lower sensitivity of FVP to the peculiar behavior of the signal during central events. Namely, the amplitude of the FVP does not completely diminish during clear central apnea, and even if it does, the amplitude increases prior to the resumption of flow (Figure 3). We found that this signal pattern was present only in the RIIV-PPG signal and was not apparent in the motion or pressure signals were also used to derive the FVP. Additionally, we found that the magnitude of this pattern and frequency of its occurrence varied greatly among individuals. We believe this phenomenon is the main reason that only two out of five central apneas were correctly classified with FVP, and that in a paradigm in which the effect of subjective factors was minimal (Table 5) there were almost two times fewer central events detected with FVP as with chest or abdomen band. While it is not yet clear what mechanisms are causing the unique signal pattern during central events, we suspect that the interplay between other components of the PPG and the applied signal processing may be partly responsible.

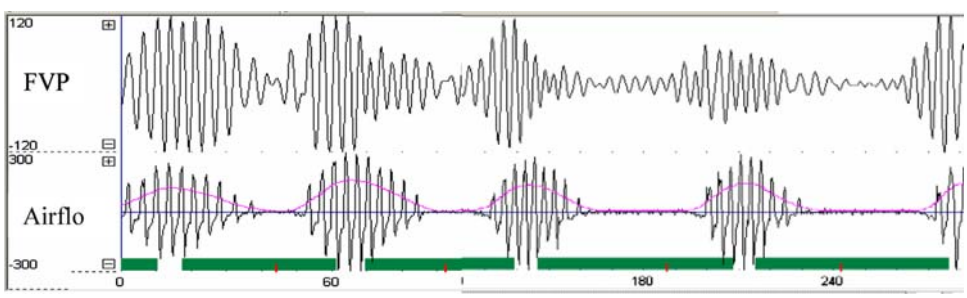

Fig. 3. A sequence of clear central apneas during which FVP seems to show some persisting respiratory effort. Note the difference in frequencies of the FVP between the portions with and without breathing (particularly during the third and fourth apnea in the row), which suggests that the 'effort' during the apneas is coming from a slower PPG component not directly related to changes intrathoracic pressure.

Besides variations synchronous with cardiac and respiratory rhythms, PPG contains low-frequency variations often associated with the baroreflex loop and thermoregulation, probably mediated through the sympathetic nervous system [20, 26]. Since respiration modulates sympathetic activity [27], and the cutaneous vessels are under sympathetic control, a sympathetic modulation of the RIIV-PPG signal is possible. If the frequency at which this modulation occurs falls within or even somewhat outside of the pass band of the applied digital filter (as long as the attenuation at that frequency is less than 5 times, or −10 dB), the sympathetic component will be present in the extracted RIIV-PPG signal, and may in fact dominate in the absence of a usually much stronger respiratory component (Figure 3). This would also explain why the same phenomenon is not observable during obstructive events even though the sympathetic modulation could still be taking place—as long as the amplitude of the respiration-synchronous component is much higher than the vasomotor PPG components it will dominate the picture. We also need to investigate whether these patterns and/or the individual differences are influenced by variability in venous volume resulting from the position of the head position relative to the sternal angle/right atrium. A study with a greater number of central apneas and hypopneas will be needed to determine if adjustments in the signal processing routines can limit this phenomenon or alternative scoring patterns can be defined to improve reliability.

The subjectivity of human scoring tended to suppress the inter- and intra-rater reliability in spite the fact that we: (a) defined a stringent set of classification rules, (b) discarding the events that did not clearly belong to one of the defined categories, and (c) used two very experienced raters. Some of the differences could be attributed to the fact that Rater 1 used five channels (all four effort measures and airflow) and only classified the events once whereas Rater 2 was provided the airflow plus only one effort measure for each of four sets of event classifications. It should be noted that the inter- and intra-rater variability in this study was comparable to that reported to validate pulse transit time as the measure of respiratory effort [7, 8].

One of the limitations of our methodology is that we relied on the kappa statistics to assess the overall agreement although an important assumption for its use—that of statistical independence of raters—was violated in a test-retest paradigm, and there is no agreed-upon model of human decision making that could be used to estimate the probability of obtaining the same classification outcome by pure chance. However, kappa has become a standard approach in sleep medicine for assessing inter- and intra-rater reliability and for comparing human and automated scoring in cases where the scoring is done on a nominal scale, such as for respiratory events [28], EEG arousals [4, 28] and sleep stages [4, 29]. Our use of kappa in addition to reporting overall agreements was motivated by the characteristics of our dataset in which three quarters of all events were obstructive, and more than a half of all events belong to a single category (obstructive hypopneas). While such dataset represents fairly well the samples of patients and/or respiratory problems commonly seen in sleep laboratories, it may give rise to high overall agreement percentages even if the agreement on non-obstructive categories between the raters is poor. This may be additionally aggravated by the fact that sleep clinicians (including our raters), seeing dominantly patients with obstructive sleep apnea, can through their everyday practice develop an unconscious bias towards recognizing obstructive events more easily, and scoring them more frequently. Kappa is able to account for such situation because its chance term will be dominated by the observed frequency of the most frequent event category, in this case obstructive hypopneas, and was therefore deemed a more appropriate measure of overall agreement on our dataset than simple agreement percentages. As one can see from Tables 3, 4, 5, the raters indeed agreed better on obstructive events, and the kappa is always lower than the corresponding overall agreement percentage. We should additionally emphasize that our primary goal was not to compare the raters nor analyze strategies humans use while scoring, but to rather use human raters as methodological tools to compare the four signals that measure respiratory effort in a clinical context. Thus, the reported agreement percentages and kappa values should be only compared against one another, and should by no means be thought of as absolute measures of 'goodness' of the investigated effort signals.

By providing Rater 2 with only one effort band at the time we deviated from the common clinical practice of inspecting both respiratory bands (often with addition of thee sum of chest and abdomen). This approach was selected to provide a blinded validation of the FVP signal and there was little evidence that it significantly affected the results of this study. Phase differences in chest and abdomen movements are primarily used to detect increased upper airway resistance. If the presentation of only one effort band indeed impaired the classification accuracy one would expect to have seen more disagreement between the raters on obstructive hypopneas or flow limitations with the chest or abdomen than with the FVP or esophageal catheter. An additional benefit of this approach was to assess the sensitivity using only a chest belt, a configuration used in several commercially available Level III devices [5].

The goal of this study was to establish the face validity of the FVP against the established measures of respiratory effort. Thus we chose to limit the analysis to differences in

signal patterns attributed to purely anatomical and physiological factors and not those attributed to technical considerations. It could be argued that this approach may have biased the results, given more than 20% of the initially chosen respiratory events were discarded and artifacts occurred in the FVP more frequently than in the other effort measures. The reasons for discarding low signal quality events included movement artifacts (which affected all effort measures), strong cardiac artifact and/or tonic spasms that obscured the respiratory component in the esophageal pressure signal and amplifier saturation (when the amplifier gain of the respiratory band was improperly adjusted). The majority of FVP events were discarded as a result of adjustments to the red and infrared light intensity triggered by a gross movement or position change occurring at the end of a sleep disordered breathing event. The optical signal light adjustment is triggered because either of these events can change the pressure applied to the forehead sensor against the skin and could compromise the accuracy of the SpO2 signal. Subsequent to this study, engineering changes were made to reduce the frequency of light adjustments especially when these would have otherwise occurred during sequences of respiratory events. While the percentage of artifact in the FVP in this study likely reflected the amount of artifact that would be present if the ARES Unicorder was self-applied, we were unable to assess the amount of effort band artifact resulting from self-application because the effort bands were applied by a technician.

The findings from this study should be interpreted within the context of the targeted application i.e., accurate, reliable, in-home sleep studies. Most patients referred for in-home sleep studies will likely have a high pretest probability of having obstructive sleep apnea and no known congestive heart failure. Viewing central sleep apnea within the context of the entire study as opposed to an event by event basis, we found that FVP provided the clinically relevant information needed to correctly identify all patients who had central or mixed events. Thus the primary reason for including respiratory effort in an in-home study, i.e., ensuring that patients with central sleep apnea are not placed on auto-titrating continuous positive airway pressure, would have been achieved. In an appropriately designed in-home sleep study program, these patients would be referred to a sleep laboratory for a follow-up polysomnogram.

In conclusion, FVP was found to provide clinical relevant information needed for the differential diagnoses of sleep disordered breathing. FVP was superior to effort belts in the detection of obstructive apneas and hypopneas, similar in the detection of persistent flow limitation and physiological changes in ventilation and inferior in the detection of central events. Although the behavior of FVP is different than effort belts during central events, with proper orientation as to the behavior of this signal coupled with visual inspection of the entire study, patients with central or mixed apnea will likely be correctly triaged.

Grant Support Dr. Popovic was a consultant to and now is an employee of Advanced Brain Monitoring, Inc. Mr. Levendowski and Dr. Westbrook are employees and shareholders in Advanced Brain Monitoring, Inc. This study was partially funded by NIH Grants 5R44HL068463-05, 1R43HL088781-01, R44DE016772. The authors wish to thank Teotimo Andrada of Walter Reed Army Medical Center for assistance in acquiring the PSG and ARES data, and Daniela Scarfeo and Timothy Zavora of Advanced Brain Monitoring for assistance in processing and reviewing the ARES data.

REFERENCES

1. Woodson BT, Wooten MR. A multisensory solid-state pressure manometer to identify the level of collapse in obstructive sleep apnea. Otolaryngol Head Neck Surg 1992; 107: 651–656.
2. Chediak AD, Demirozu MC, Nay KN. Alpha EEG sleep produced by balloon catheterization of the esophagus. Sleep 1990; 13: 369–370.
3. Chervin RD, Aldrich MS. Effect of esophageal monitoring on sleep architecture. Am J Respir Crit Care Med 1997; 156: 881–885.
4. Redline S, Sanders MH, Lind BK, et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep 1998; 21(7): 759–767.
5. White DP, Gibb TJ, Wall JM, Westbrook PR. Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing at home. Sleep 1995; 18(2): 115–126.
6. Farre R, Montserrat JM, Navajas D. Noninvasive monitoring of respiratory mechanics during sleep. Eur Respir J 2004; 24: 1052–1060.
7. Argod J, Pepin JL, Levy P. Differentiating obstructive and central sleep respiratory events using pulse transit time (PPT). Am J Respir Crit Care Med 1998; 158: 1778–1783.
8. Argod J, Pepin JL, Smith RP, Levy P. Comparison of esophageal pressure with pulse transit time (PPT) as a measure of respiratory effort for scoring obstructive non-apneic respiratory events. Am J Respir Crit Care Med 2000; 162: 87–93.
9. Condos R, Norman RG, Krishnasamy I, et al. Flow limitation as noninvasive assessment of residual upper-airway resistance during continuous positive airway pressure therapy of obstructive sleep apnea. Am J Respir Crit Care Med 1994; 150: 475–480.
10. Series F, Marc I. Accuracy of breath-by-breath analysis of flow-volume loop in identifying sleep-induced flow-limited breathing cycles in sleep apnea-hypopnea syndrome. Clin Sci 1995; 88: 707–712.
11. Hosselet JJ, Norman RG, Ayappa I, Rapoport DM. Detection of flow limitation with a nasal canula/pressure transducer system. Am J Respir Crit Care Med 1998; 157: 1461–1467.

12. Montserrat JM, Farré R, Ballester E, et al. Evaluation of nasal prongs for estimating nasal flow. Am J Respir Crit Care Med 1997; 155: 211–215.
13. Vegfors M, Ugnell H, Hok B, et al. Experimental evaluation of two new sensors for respiratory rate monitoring. Physiol Meas 1993; 14: 171–181.
14. Nilsson L, Johansson A, Kalman S. Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique. J Clin Monit 2000; 16: 309–315.
15. Johansson A, Oberg PA, Sedin G. Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique. J Clin Monit 1999; 15: 461–467.
16. Nakajima K, Tamura J, Miike H. Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. Med Eng Phys 1996; 5: 365–372.
17. Dorlas JC, Nijboer JA. Photo-electric plethysmography as a monitoring device in anaesthesia: application and interpretation. Br J Anaesthesiol 1985; 57: 524–530.
18. Bernardi L, Radaelli A, Solda PL, et al. Autonomic control of skin microvessels: assessment by power spectrum of photoplethysmographic waves. Clin Sci 1996; 90: 345–355.
19. Johansson A, Oberg PA. Estimation of respiratory volumes from the photoplethysmographic signal. Part 2: a model study. Med Biol Eng Comput 1999; 37: 48–53.
20. Johansson A, Oberg PA. Estimation of respiratory volumes from the photoplethysmographic signal. Part 1: experimental results. Med Biol Eng Comput 1999; 37: 42–47.
21. Nilsson L, Johansson A, Kalman S. Respiratory variations in the reflection mode photoplethysmographic signal: relations to peripheral venous pressure. Med Biol Eng Comput 2003; 41: 249–254.
22. Nilsson L, Johansson A, Kalman S. Macrocirculation is not the sole determinant of the respiratory induced variations in the reflection mode photoplethysmographic signal. Physiol Meas 2003; 24: 925–937.
23. Lea S, Ali NJ, Godman M, et al. Systolic blood pressure swings reflect inspiratory effort during simulated obstructive sleep apnea. In: Horne, J, ed., Sleep. Pontanagel Press, Bochum, 1990: 178–181.
24. Ayappa I, Norman RG, Seelall V, Rapoport DM. Validation of a self-applied unattended monitor for sleep disordered breathing. J Clin Sleep Med 2007; 4(1): 26–37.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–174.
26. Lindberg LG, Ugnell H, Oberg PA. Monitoring of heart and respiratory rates using a fiber-optic sensor. Med Biol Eng Comput 1992; 30: 533–537.
27. Macefield VG, Wallin BG. Modulation of muscle sympathetic activity during spontaneous and artificial ventilation and apnoea in humans. J Auton Nerv Syst 1995; 53: 137–147.
28. Thomas RJ. Arousals in sleep-disordered breathing: patterns and implications. Sleep 2003; 26(8): 1042–1047.
29. Norman RG, Pal I, Stewart C, et al. Inter-observer agreement among sleep scorers from five different centers on a large dataset. Sleep 2003; 23(7): 901–908.