The microstructure of the sleep in primary insomnia: An overview and extension

Bernd Feige *, Chiara Baglioni, Kai Spiegelhalder, Verena Hirscher, Christoph Nissen, Dieter Riemann

Department of Psychiatry and Psychotherapy, Freiburg University Medical Center, Hauptstraße 5, 79104 Freiburg, Germany

A R T I C L E   I N F O

Article history:
Received 23 October 2012
Received in revised form 2 April 2013
Accepted 4 April 2013
Available online 11 April 2013

Keywords:
Insomnia
Paradoxical insomnia
Polysomnography
Microstructure micro-arousal
Spectral analysis
CAP
Event-related potentials

A B S T R A C T

The present review was undertaken to summarize studies elucidating sleep microstructural differences in chronic insomnia. The etiology of insomnia is still unknown, whereas the hyperarousal concept has gained much attention with respect to pathophysiology. According to this model, insomnia is characterized by significant hyperarousal on an autonomous and central nervous level. Objective findings derived from polysomnography frequently show much less severe differences to good sleepers than subjective sleep complaints assessed by self-rating questionnaires. However, using more fine-grained methods to characterize the electrophysiology of sleep in insomnia, rather distinct differences between the sleep of good sleepers and patients with insomnia have been noted. These methods include the spectral analysis of the sleep EEG, micro-arousal and CAP (cyclic alternating pattern) analysis as well as the assessment of event-related potentials (ERPs) during night-sleep. The application of these methods shows stronger correlations with the subjective experience of disturbed sleep than standard sleep EEG scoring. An overview of the relevant empirical evidence is presented, previous investigations are extended and a theoretical synthesis within the framework of the hyperarousal concept of insomnia is attempted.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

With a prevalence of 10%, chronic insomnia is one of the most frequent health complaints in Europe and its impact will increase further with the aging of society (overview see Ohayon, 2002; Ohayon et al., 2012). Insomnia is defined by difficulties initiating or maintaining sleep or non-restorative sleep, accompanied by significant daytime impairments. The minimum duration of these symptoms for a diagnosis of chronic insomnia is currently fixed relatively arbitrarily at one month (DSM-IV, American Psychiatric Association, 2000). While studies have required a duration of up to six months to define chronicity (NIH, 2005), Ohayon et al. (2012) recommend three months. The disorder frequently occurs as a co-morbid condition in medical or mental disorders. Primary insomnia constitutes an exclusionary diagnosis of poor sleep, ruling out psychiatric, medical, substance, and other sleep-related pathology (overview see Morin and Benca, 2012). In the current version IV of the DSM (American Psychiatric Association, 2000), insomnia in the presence of major other diagnoses is interpreted as a secondary condition. However, in the DSM-V (to be published in 2013), the classification according to primary/secondary insomnias will be replaced by an overall category of “Insomnia Disorder”, which will allow to code co-morbid disorders (Reynolds, 2011; Riemann et al., 2011a).

Chronic insomnia is associated with diminished quality of life, increased fatigue, cognitive impairments, mood disturbances and physical complaints. It confers an increased risk for mental disorders, especially major depressive disorder (Riemann and Voderholzer, 2003; meta-analysis see Baglioni et al., 2011), and there is evidence that it is a risk factor for cardiovascular/metabolic disease and increased mortality (Kripke et al., 2002; Spiegelhalder et al., 2011; Sofi et al., in press). For the US (corresponding data for Europe are not available yet), the costs of insomnia due to low work performance and absenteeism have been estimated to exceed 60 billion $ per year (Kessler et al., 2011). The full socioeconomic burden is even higher (Ozminkowski et al., 2007; Fullerton, 2006; Leger et al., 2002). Despite this drastic socio-economic impact, insomnia’s etiology and pathophysiology are not well understood yet (Riemann et al., 2011b).

2. Insomnia — the hyperarousal concept

The model of Spielman et al. (1987) (the “3P” or behavioral model of insomnia) posits that there are predisposing, precipitating (triggering) and perpetuating factors to the disease. In particular, stressful life events are seen as main precipitating factors while insomnia is perpetuated by learned (negative) associations and maladaptive coping strategies. Closer to testable physiological hypotheses, Perlis et al.
(1997) later formulated the neurocognitive model of insomnia which suggests that chronic insomnia occurs in association with conditioned arousal. According to this model, the conditioned arousal leads to increased sensory and information processing and long term memory formation.

Current theories of primary insomnia (as a model of “pure” insomnia) consequently highlight the role of cognitive, emotional and physiological hyperarousal for its development and maintenance (see for example Harvey, 2002; Espie et al., 2006). Two summaries of the state-of-the-art of the hyperarousal concept (Bonnet and Arand, 2010; Riemann et al., 2010), listed substantial support for the assumption that hyperarousal processes from the cellular to the higher system level play a decisive role in the pathophysiology of insomnia.

Autonomous, neuroendocrine, neuroimmunological, electrophysiological, neuroimaging and psychological studies delivered converging evidence for increased levels of arousal in primary insomnia compared to good sleepers not only during the night but over the whole 24 h day. This corresponds to the subjective experience of patients with insomnia having difficulties to “shut down” or to disengage from wakefulness (when trying to sleep). One predisposing factor may be a genetic vulnerability to be unable to down-regulate arousal (e.g. Bastien and Morin, 2000).

A seminal study with FDG-PET (positron emission tomography; Nofzinger et al., 2004) disclosed that patients with primary insomnia show smaller declines in relative glucose metabolism from wakefulness to NREM sleep stage 2 in wake promoting regions (i.e. the ascending reticular activating system ARAS, the hypothalamus and thalamus). Comparable effects were seen in areas associated with cognition and emotion encompassing the amygdala, hippocampus, insular cortex, the anterior cingulate and prefrontal cortices. This study thus gave direct evidence for the hypothesis of a central nervous system hyperarousal in insomnia. Patients with primary insomnia in this investigation demonstrated patterns of brain activation indicative of the wake state in spite of showing obvious and clear-cut signs of electroencephalographically measured NREM stage 2 sleep. No PET data have yet been accumulated during REM sleep in primary insomnia patients. It has been suggested (see below) that REM sleep may be pathologically close to the waking state in insomnia patients, as evidenced by correlation with subjective wake time and increased frequency of micro-arousals during this sleep state compared to NREM sleep (Feige et al., 2008; Riemann et al., 2012).

3. Sleep misperception — paradoxical insomnia

A major enigma of insomnia research constitutes the frequently noted discrepancy between the subjective experience of sleep (measured by sleep questionnaires) and “objective”, i.e. polysomnographic findings. Polysomnographic (PSG) studies demonstrate that patients suffering from insomnia tend to underestimate their nocturnal sleep time (Carskadon et al., 1976; Frankel et al., 1976; Adam et al., 1986; Feige et al., 2008; Manconi et al., 2010). This sometimes striking discrepancy has led to the terms “sleep state misperception” or to put it into a more neutral phrasing “paradoxical insomnia” for patients with a relatively normal sleep continuity and architecture (Edinger and Krystal, 2003), in spite of massive subjective complaints of disturbed sleep.

A strong discrepancy between subjective and objective findings was also noted for the daytime cognitive performance as measured by neuropsychological testing in primary insomnia (overview Fortier-Brochu et al., 2012). Several hypotheses to explain this discrepancy were recently discussed in detail by Harvey and Tang (2012). These authors highlight the possible role of psychological and physiological factors involved. On the one hand, electrophysiologically defined sleep parameters, like for example sleep latency, might not be capable of capturing the subjective experience of falling asleep, because the parameter is operationalized in a way too crude (latency from lights out to the occurrence of first stage 2) to gauge the “on and off” of stages Wake, 1 and 2 at sleep onset in insomnia. On the other hand, primary insomnia patients may have an altered ability for time estimation or a bias towards perceiving wake states because they are highly focused on their symptoms. Harvey and Tang (2012) also list elevated cortical arousal, the presence of brief awakenings and a fault in neuronal circuity as possible explanations for paradoxical insomnia.

Feige et al. (2008) showed that the difference between subjective and objectively measured wake time correlated with the amount of REM sleep in insomnia patients — i.e. patients with higher amounts of REM sleep tended to report more minutes of subjective wakefulness. In addition, subjects suffering from insomnia displayed highly significantly increased numbers of micro-arousals and awakenings during REM sleep in comparison with healthy good sleepers — a finding which was evident in NREM sleep only to a lesser extent. This stimulated us to formulate the “REM sleep instability” hypothesis of insomnia (Riemann et al., 2012). This fits well with ideas of F. Snyder who proposed in the context of depression that “inner turmoil” specifically diminishes the ability to discriminate between REM sleep and waking, and that the development of a depressive episode was characterized by a succession of REM sleep deprivation and compensation (Snyder, 1965, 1972).

To our knowledge, no explanatory strategy concentrated its effort on “instability” of REM sleep as a possible reason for the discrepancy between objective and subjective sleep perceptions in primary insomnia. In that context it is important to look at investigations addressing the issue of sleep mentation/dreams in patients with insomnia. Schredl (2009), in a review article on the topic, concluded that insomnia patients experience themselves more negatively in their dream-life and that their acute wake-life concerns were reflected in their dream narratives — justifying the postulate that major waking-life concerns of primary insomnia patients, i.e. pre-sleep arousal related to the perceived inability to sleep properly — could be overrepresented in their dream experience. Mercer et al. (2002), by utilizing Stage 2 and REM sleep awakenings in a small sample of patients with insomnia and good sleepers, provided some preliminary evidence that indeed patients with primary insomnia, when aroused experimentally from both types of sleep, display significantly higher rates of judging their pre-awareness state as awake.

A major inroad to explain the discrepancy between subjective and objective sleep data may lie in the micro-analysis of sleep states instead of analyzing only the macrostructure as enshrined in the Rechtschaffen and Kales (1968) (R&K) manual of sleep staging.

4. Electrophysiology of insomnia — standard scoring versus fine-grained approaches

While R&K staging sets the standard definition of systemic sleep stages, the rule set is relatively insensitive in a number of areas. First, R&K scoring, by design, misses short transient phenomena by employing a “majority rule” (see below) for 30 second epochs. Second, a single quantitative rule exists for delta power; the quantitative composition of the EEG signal from other frequencies is disregarded. Third, R&K does not contain the notion of sub-states within a given sleep state. To compensate for these specific insensitivities, the assessment of sleep can sensibly be augmented by the analysis of transients (microarousals, spindles, K-complexes), spectral and CAP analysis.

A further limitation of conventional staging is the concept of a single sleep stage defining systemic brain state. In recent years, evidence for “local sleep” components has been found, first for local sleep homeostasis (Kattler et al., 1994; Huber et al., 2004) and more generally for sleep being orchestrated but not strictly synchronized across all parts of the brain (Rector et al., 2009; Nobili et al., 2011). This additional shortcoming of conventional sleep staging can be
addressed by topographical recordings or even high-density sleep EEG (cf. Manconi and Pisarenco in this issue) in combination with spectral power or coherence analysis.

R&K PSG alone had limited success in advancing hypotheses about insomnia pathophysiology: most of the investigations performed in the field demonstrated that PSG derived sleep variables produced far less pronounced differences to good sleepers than expected from the subjective estimates of the patients' sleep quality. Surprisingly, studies using the Multiple Sleep Latency Test (MSLT) in groups of patients with insomnia revealed that the data from patient samples were not characterized by indicators of preceding sleep loss (as should have been evidenced by reduced sleep latencies during daytime naps) but instead by normal or even increased sleep latencies during this daytime test.

In consequence of the insufficient explanation of subjective complaints in primary insomnia by conventional PSG data, the spectral analysis of EEG sleep, micro-arousal analysis (according to Sleep Disorders Atlas Task Force of the American Sleep Disorders Association, 1992), the application of the CAP (cyclic alternating pattern) method and the registration of event-related potentials prior to and during sleep became of high interest for the hyparousal concept. The advantages of these techniques compared to traditional epoch by epoch visual scoring according to Rechtschaffen and Kales (1968) criteria will be highlighted by presenting and discussing the accumulated empirical evidence.

4.1. Classical sleep EEG (R&K) and MSLT

The current consensus in sleep staging is described in the AASM scoring manual (Iber et al., 2007). It aims to reduce ambiguities in staging by more explicit rules and by observing additional EEG positions on the head, while keeping the central definitions of R&K staging. One exception is that R&K stages 3 and 4 are joined to stage N3. For brevity we still denote this modern concretized version as R&K staging.

Staging determines the systemic sleep state for successive time windows or “epochs” of typically 30 s duration. The majority rule determines that in case of signs of stage transitions an epoch is assigned the stage covering the largest fraction of the epoch. There also are “continuity” rules which cause epochs with unclear patterns being assigned the stage clearly discernible in epochs before or after. To determine stage changes, for example REM sleep onset, a common convention is that at least three consecutive minutes of the new stage must be observed. The, at least implicit, conception of systemic sleep state therefore is relatively slow in nature, reflecting the observation that sleep state generally changes on the basis of minutes rather than seconds. Short events or transients are described as occurring within or “on top of” the systemic sleep stage rather than modifying it. Often they even are telltale signs of the respective sleep stage – spindles and K complexes for stage 2, rapid eye movements for REM sleep. As noted above, while defining the general classification of systemic sleep state by common criteria, the definitions leave ample room for brain state differences within a given sleep stage.

It would be beyond the scope of this overview to review all the studies which have used PSG (registered and scored according to R&K) to characterize the sleep behavior of patients with insomnia. Alternatively, this review will cover published overviews and meta-analyses on the topic (Benza et al., 1992; Reite et al., 1995; Hudson et al., 1992; Perlis et al., 2010).

Summarizing from these publications, PSG derived sleep continuity in primary insomnia is compromised by increased sleep onset latency (SOL), enhanced wake time after sleep onset (WASO), decreased total sleep time (TST) and impaired sleep efficiency. The measured absolute differences are not impressive (ca.30–60 min with respect to total sleep time) but of statistical significance. As mentioned before, it is a very consistent finding that the PSG differences are of a much lesser magnitude than the subjectively estimated differences in sleep continuity between primary insomnia and good sleepers. Further evidence indicates that subtypes of primary insomnia (for example psychophysiological vs. paradoxical insomnia) are inhomogeneous regarding the PSG findings (Edinger et al., 2004). An additional complicating factor is the result of a marked night-to-night variability in some patients with respect to PSG data. Concerning sleep architecture, no clear picture emerges — only a minority of investigations described either SWS (slow wave sleep) or REM (rapid eye movement) sleep reductions in samples of patients with insomnia. REM latency (i.e. the interval between sleep onset and the occurrence of the first REM period), the target variable of sleep research in the field of depression (overview: Riemann et al., 2001), up to now has not been described to be shortened in insomnia patients.

A new meta-analysis of PSG studies in primary insomnia, summarizing the newer literature between 1994 and 2012 and following strict quality criteria for study inclusion, was performed by Baglioni et al. (in press). This meta-analysis by and large replicated the abovementioned findings with respect to sleep continuity, and, in addition, revealed that both SWS and REM sleep are significantly reduced in primary insomnia compared to good sleepers. The reduction amounts to only about 2% in both measures, again showing that macroscopic PSG differences stand in no relation to the severity of disruption in primary insomnia.

The Multiple Sleep Latency Test (MSLT) in its standard form (Carskadon et al., 1986) consists of 5 brief 20 min nap opportunities during the day (9 am, 11 am, 1 pm, 3 pm, 5 pm) with EEG sleep monitoring. As outcome variables, the latencies to stages 1, 2 and REM are measured to assess daytime sleep propensity and signs of narcolepsy.

Early work (Mendelson et al., 1984; Seidel et al., 1984; Stepanski et al., 1984; Sugerman et al., 1985) demonstrated that in contrast to conclusions drawn from research on the effect of experimental sleep loss on MSLT sleep latencies, patients with insomnia did not show reduced but prolonged sleep latencies. Bonnet and Arand (1995, 1998, 2010) confirmed increased sleep latencies during the day in spite of disturbed nocturnal sleep in several investigations. They reached the conclusion that the prolonged sleep latencies of patients with insomnia in the MSLT reflect a combination of sleep tendency and a high level of arousal during daytime, arguing strongly for a 24 h hyperarousal in primary insomnia.

It was also discussed that the demand characteristics of the MSLT (“try to sleep”) are involved in this prolongation of sleep latencies (Espie et al., 2006): the demand to “do something” that the patient with insomnia has difficulty doing would elicit performance anxiety, interfering with sleep initiation. This mechanism would therefore rather detect the specific fears regarding sleep than a general physiological hyperarousal.

4.2. Spectral analysis

This method is based on the fact that any given signal can be fully represented as a sum of harmonic waves (Fourier, 1822; FFT = Fast Fourier Transformation). Spectral power analysis quantifies the amplitude of each constituent wave (or its square: power) and is able to separate slower and faster parts of the EEG. In this way, the quantity of slow (= delta) waves can be independently determined. In R&K staging, the delta amplitude is used to discriminate slow-wave sleep from sleep stage 2 via a threshold (75 µV). Furthermore, spectral analysis can exactly measure other frequencies that are not exactly quantified by R&K criteria, such as alpha activity or fast (beta/gamma) EEG components in NREM sleep and REM sleep. The observation of large slow wave amplitudes during deep NREM sleep is one of the central findings at the core of the development of R&K sleep staging. In an early study using spectral analysis in healthy subjects (Johnson et al., 1968), the relative amount of higher frequency (beta and gamma) EEG power was found
to be increased in waking, stage 1 and REM sleep compared to stages 2, 3 and 4, thus establishing a phenomenological link to the degree of cortical arousal. Relative power is often used to homogenize spectral power data with respect to overall differences in EEG power occurring between subjects or sleep stages. As a cautionary note, spurious anticorrelations in relative power may be observed particularly between delta and higher frequency bands. This is because slow frequencies dominate the EEG spectrum (1/f spectrum, Robinson et al., 2001; Linkenkaer-Hansen et al., 2001) and strongly vary with sleep depth. In particular, a relative increase in high-frequency power may either reflect an increase of absolute high-frequency power or a reduction in absolute slow-wave power.

Only few studies have been performed to highlight differences in the spectral sleep EEG composition between primary insomnia and good sleeper controls. Freedman (1986) investigated 12 patients with sleep onset insomnia compared to 12 good sleepers. He found increased absolute EEG beta power during wake, stage 1 and REM sleep but not during NREM stages 2, 3 and 4 in patients with primary insomnia. In a study by Merica et al. (1998), 20 chronic sleep maintenance insomnia patients and 19 good sleepers were investigated. The authors described increased beta power throughout the night (maximally in REM sleep), in addition to reductions in slower EEG frequencies in both REM sleep and NREM sleep. Perlis et al. (2001b) compared relative spectral power in three groups (n = 9 subjects per sample): primary insomnia, insomnia secondary to major depressive disorder and good sleepers. Subjects with primary insomnia exhibited more average NREM activity for Beta-1 (14–20 Hz), Beta-2 (20–35 Hz) and Gamma activity (35–45 Hz) than the other two groups (p < .01). Additionally, a significant correlation between sleep misperception and the amount of fast frequencies was reported. Relative slow-wave power did not show significant differences in this study. In a second analysis of the same data set, Perlis et al. (2001a) showed that beta power and gamma power were increased in the patients with insomnia relative to good sleepers for several stages of NREM sleep, and that only Beta-2 was increased during REM sleep. They concluded that the finding of increased high frequency EEG activity in primary insomnia appears to be a NREM phenomenon.

A study by Krystal et al. (2002) analyzed absolute and relative spectral power in 12 “subjective” and 18 “objective” patients with insomnia and 20 good sleepers and reported reduced relative delta and increased relative alpha, sigma and beta power during NREM sleep in the patients with “subjective” insomnia but not in the “objective” insomnia group. No differences were found for REM sleep. Buyssse et al. (2008), who targeted absolute and relative NREM sleep EEG power in 48 primary insomnia patients and 25 good sleepers, were unable to confirm a general difference between the groups, but described heightened absolute delta and beta power specifically in the female primary insomnia patients only. An analysis of own data (Spiegelhalder et al., 2012) of 25 patients with primary insomnia and 29 good sleepers revealed increased absolute spectral power values in the beta and sigma frequency bands in the patients in stage 2 sleep but not in REM sleep.

In one of the scarce clinical studies considering EEG topography and therefore being able to address aspects of “local sleep”, St-Jean et al. (2012) focused on hemispheric asymmetry in 17 patients with psychophysiological insomnia, 14 with paradoxical insomnia and 19 good sleepers. Patients with paradoxical insomnia showed a trend for increased left frontal omega (60–125 Hz)-band activity while good sleepers had more omega-band power over right frontal sites. The psychophysiological insomnia group had more right parietal beta power than the paradoxical insomnia group but none differed from good sleepers in this measure. Interestingly, asymmetry was found to be independent of depressive symptoms, while particularly reduced frontal EEG asymmetry is reported in the context of depression during waking (cf. Coan and Allen, 2004). Summarizing, the evidence presented above points strongly towards increased fast frequencies (in the sigma and beta range) as characteristics of the sleep EEG of patients with primary insomnia. Cortical electrophysiological signals in the beta and especially gamma band have been assumed to reflect coherent cortical processing of sensory information (“feature binding”, Engel et al., 1991; Singer, 1993; and possibly of all cognitive activity: Galambos et al., 1981; Llinás and Ribary, 1993). Thus, an EEG power increase in these frequency bands during sleep can be interpreted as a marker of cortical hyperactivity or hyperarousal. First studies show topographical differences of these phenomena, indicating that more attention must be paid to “local sleep effects” and the subsystems involved.

Concerning sigma and beta activity of the sleep EEG an important caveat has to be discussed, as many patients with primary insomnia have experience with the intake of benzodiazepines (BZ) as sleep medication. BZ hypnotics lead to clear-cut modifications of the NREM sleep EEG spectrum, named “GABA-Benzodiazepine signature”: suppression of frequencies below 10 Hz and a clear increase of sigma band power (Borbély et al., 1985; Brunner et al., 1991); frequencies below 1 Hz are rather increased by acute intake (Monti et al., 2000).

In good sleepers, Feige et al. (1999) have shown that after a 4-week intake of different BZ-hypnotics (triazolam, zolpidem and zopiclone), mean NREM beta power tends to be increased, while a low-frequency reduction remains. Insofar, both the acute and continued use of benzodiazepine hypnotics result in increased relative amounts of beta power within the whole spectrum, while continued use may even lead to increased absolute beta power.

An important methodological study by Bastien et al. (2003) examined absolute NREM spectral EEG power in 46 older adults (>55 years) in three groups: insomnia with and without chronic use of benzodiazepines and good sleepers. No spectral power differences were found between insomnia patients without benzodiazepine use and the good sleeper group but only between chronic benzodiazepine users and non-users. Therefore, sleep EEG spectral power analysis in insomnia requires very strict criteria regarding the previous and current use of BZ hypnotics in order to exclude previous BZ use as a confounding factor. Commonly used wash-out times set target thresholds at least below 2% (5.6 half-lives, e.g. Martinot, 1992) or 1% (6.6 half-lives). To be on the safe side, at least with respect to modern short-acting hypnotics, two weeks of drug wash-out should be scheduled to eliminate possible confounding effects of BZ.

Given the strong, partially hereditary variability in spectral sleep EEG composition (Ambrosius et al., 2008; De Gennaro et al., 2008), large and well-controlled studies are needed to strengthen the picture — in particular, it is currently unclear whether EEG spectral power differences in insomnia are more pronounced within NREM or REM sleep, since several studies were restricted to NREM sleep and the remaining studies reported contradictory results for REM sleep. Future studies in larger samples with well-defined insomnia groups (using research diagnostic criteria and standard methods for sample description) and good sleepers are mandatory to determine the “real” difference concerning the spectral composition of the sleep EEG in contrast to the variance due to age, sex, severity of insomnia and prior intake of CNS active medications.

4.3. Micro-arousal analysis

This type of arousal analysis was first suggested by the American Sleep Disorders Association (Sleep Disorders Atlas Task Force of the American Sleep Disorders Association, 1992) as an appendix to the R&K manual, in order to delineate rules for scoring micro-events during sleep not captured by the original R&K criteria. In short, the ASDA criteria state “An EEG arousal is: An abrupt shift in EEG frequency, which may include theta, alpha and/or frequencies greater than 16 Hz but not spindles, subject to the following rules and conditions.” 11 specific criteria are given in the original text on how to score the
micro-arousals. These have to last at least for 3 s and are scored in REM sleep only when accompanied by concurrent increases in submental EMG amplitude, which is in contrast to NREM sleep where an increase in submental EMG amplitude is not necessary. These criteria have been used frequently in sleep research (e.g. Mathur and Douglas, 1995; De Gennaro et al., 2001); micro-arousals can be scored reliably (Loredo et al., 1999) and their validity has been proven (Bonnet and Arand, 2007).

With respect to insomnia, one study has been published (Feige et al., 2008). This retrospective data analysis focused on the discrepancy between subjective and polysomnographically determined total sleep time in patients with primary insomnia, comparing subjective sleep ratings and polysomnographic data from 100 drug-free patients with primary insomnia and 100 good sleepers. In line with previous studies, insomnia patients tended to underestimate their total sleep time when compared with good sleepers. Interestingly, a significant relationship between subjective wake time and the amount of REM sleep was found in insomnia patients. Furthermore, the frequency of micro-arousals (as measured by Sleep Disorders Atlas Task Force of the American Sleep Disorders Association, 1992 criteria) was significantly enhanced during both REM sleep and NREM sleep in primary insomnia compared to good sleepers, the REM sleep effect being more pronounced.

In the following we extend the analysis of micro-arousals and awakenings to NREM/REM periods across the night: Fig. 1 shows micro-arousals throughout the first four NREM and REM sleep cycles when comparing 100 primary insomnia and 100 good sleepers (all subjects free of psychoactive medication for at least 2 weeks). The group difference is expressed as effect size at the top of each plot. In accord with the nonparametric box plots, effect sizes were computed in a robust way based upon median and inter-quartile range (IQR): Equivalent standard deviations (SD) were computed as IQR/1.349 and pooled across the compared groups. Effect size d was then expressed as the difference in medians divided by the pooled SD.

Fig. 2 shows a composite index of micro-arousals plus full-blown awakenings for the same type of comparison.

This analysis shows that including full awakenings does not yield higher effect sizes; this means that the characteristic group difference really lies in the micro-arousals. As in the whole-night analysis (Feige et al., 2008; corrigendum Feige et al., 2012), increased arousal frequency is observed both in NREM and in REM sleep, with the REM sleep difference being 2–3 times larger. Interestingly, the very first NREM and REM periods both show a comparatively low effect size. This could hint towards a homeostatic unmasking of the hyperarousal at later phases of the night.

### 4.4. Other spontaneous sleep events: spindles and K-complexes

Sleep spindles (Jankel and Niedermeyer, 1985; De Gennaro and Ferrara, 2003) are typical transient phenomena (graphoelements) of stage 2 sleep. The main frequency is typically 12 Hz (frontal) to 14 Hz (parietal; cf. Jobert et al., 1992; Schabus et al., 2007) with a waxing and waning envelope that gives the oscillation a spindle-like appearance. Spindles occur spontaneously, i.e. not time-locked to stimulation, and have been shown to be grouped by slow oscillations (Amzica and Steriade, 1998; Mölle et al., 2002). Spindles are clearly not signs of sleep disruption but of ongoing stable sleep; on the other hand, a sleep-protecting role was suggested (Cote et al., 2000) but not conclusively shown (De Gennaro and Ferrara, 2003), possibly because it is hard to separate spindles from the slow waves and K complexes which typically precede them.

K complexes (Loomis et al., 1938) are graphoelements typical for sleep stage 2 consisting of a fast surface-positive peak followed by slow surface-negative and then surface-positive waves. They can occur either spontaneously or be evoked by stimulation. Cash et al. (2009) identified K complexes as isolated thalamocortical “down-states”. Down states are periods of neuronal silence normally associated with the surface-negative phase of slow oscillations. This discovery corroborated previous research showing K complexes to have a sleep-protecting property (Bastien et al., 2000). For evoked K complexes (those elicited by a stimulus) this means that the stimulus was not categorized as threat and the brain remains in its sleep state.

Regarding insomnia, Besset et al. (1998) examined homeostatic variation of slow-wave activity (SWA) and sleep spindles in 7 primary insomnia patients and 7 good sleepers. In the patients the sleep spindle index (SSI) was significantly lower than in good sleepers and did not decrease further during slow-wave sleep, which would correspond to the normal inverse relationship between SWA and spindles (Dijk et al., 1993). However, Bastien et al. (2009a) did not find a difference in the number and density of sleep spindles between 16 patients with chronic psychophysiological insomnia and 14 good sleepers. Furthermore, Bastien et al. (2009b) did not find a difference in spontaneous K-complexes between the same groups. They concluded that no deficiency exists in the sleep-protecting mechanism in chronic psychophysiological insomnia.
4.5. CAP (cycling alternating pattern)

The cyclic alternating pattern (CAP; Terzano et al., 1985) is predominantly a NREM sleep phenomenon and has some overlaps with R&K events. CAP is based on the observation that certain EEG elements, distinct from background EEG, recur with a periodicity of 20–40 s (Terzano et al., 2001). CAP marks an arousal fluctuation, containing a succession of unstable or aroused (phase A) and electrophysiologically quiet sleep (phase B). Therefore, Halasz (1998) described CAP as a larger-scale measure of arousal instability than individual micro-arousals. Recently, Parrino et al. (2012) reviewed CAP as a marker of sleep instability.

A large CAP rate marks disturbed sleep with poor sleep quality (Terzano et al., 1990). CAP rate decreases as a consequence of total sleep deprivation and correlates with EEG arousals (de Gennaro et al., 2002); furthermore, de Carli et al. (2004) showed a strict relationship between ASDA arousals and the CAP phase A subtypes A2 and A3.

In insomnia, Terzano et al. (2003) showed an increased number of arousals and an increased CAP rate, as well as increased nocturnal wake time, in the primary insomnia group when comparing 47 primary insomnia patients to 25 good sleepers. Inclusion of patient subgroups on active hypnotic medication is a limitation of this study. Parrino et al. (2004) however showed that hypnotic medication reduced CAP rate in patients with insomnia and concluded that CAP rate is a measure of the instability of sleep with increased CAP rates indicating a “destabilization” of sleep in insomnia.

Examining the effects of zolpidem vs. placebo on CAP in 17 patients with psychophysiological insomnia, Ozone et al. (2008) found a reduction in CAP rate (A1 but A2 + A3 as well). Similarly, in eight sleep maintenance insomnia patients, Parrino et al. (2008) found a reduction in CAP subtype A2 as well as arousals and arousal index in all three active treatment nights within a 7–night study protocol with zolpidem. Parrino et al. (2009) further examined the important question of the link between CAP and sleep misperception in primary insomnia and found a link between CAP phase A2 subtypes and sleep state misperception. Further investigations are necessary since an advantage of CAP analysis lies in the inclusion of phasic events different from arousals (A1 subtype) and some studies found the largest effect of hypnotics and largest correlation with subjective sleep quality with this subtype (Ozone et al., 2008; Parrino et al., 1997).

Recently, Chouvarda et al. (2011, 2012) defined new methodologies for the quantitative characterization of insomnia using CAP. Comparing 10 primary insomnia patients to 11 good sleepers, they confirmed the increased CAP rate in primary insomnia and added that CAP cycles were more irregular and desynchronisation phases more extended in primary insomnia.

4.6. ERPs (event-related potentials) in sleep

To keep this section short, we forego a detailed methodological introduction. In short, ERPs consist of typical EEG waves following a stimulus event or preceding an event, for example voluntary movements. The peaks are labeled by latency of occurrence and index different phases of event processing. Amplitudes correspond to the amount of neuronal recruiting into the respective process. Generally, more intense stimulation causes higher amplitudes (e.g. Phillips et al., 2011). Increased amplitudes to identical stimuli, on the other hand, reflect higher recruiting through higher neuronal availability, increased attention or alertness, or reduced inhibition of sensory inputs (Perlis et al., 2010). A good overview about auditory ERPs in sleep is given by Coenen (2012).

Yang and Lo (2007) examined auditory ERPs of 15 primary insomnia patients and 15 good sleepers across the whole night and found a larger N100 as well as a smaller P220 amplitude to rare tones and a smaller N350 to standard tones during the first 5 min of continuous stage 2. No differences across the whole night were found. The results were interpreted as an enhancement of attention and a reduction in the inhibitory process at sleep onset in insomnia.

Bastien et al. (2008) studied waking ERP amplitudes (evening and morning) as well as ERP during sleep onset in 15 primary insomnia patients and 16 good sleepers. During wakefulness, primary insomnia patients showed increased N100 amplitudes; during sleep onset, P220 was increased and N350 reduced. The authors interpreted their data...
more in favor of “inhibition deficits” (to disengage from waking processes) of the insomnia sample than as direct evidence for hyperarousal. Turcotte and Bastien (2009) reported, in the same subjects, an association between waking ERPs and sleep quality indicating that the observed hyperactivation is directly related to the impaired sleep quality in primary insomnia.

More recently, Turcotte et al. (2011) compared ERPs in wakefulness, sleep onset and early sleep stage 2 in 26 patients with psychophysiological, 26 with paradoxical insomnia and 26 good sleepers. The two insomnia groups exhibited qualitative differences during the sleep onset period which pointed to a more super physiological, 26 with paradoxical insomnia and 26 good sleepers. The mid-sleep latency component P50 is typically employed for this assessment.

Kertesz and Cote (2011) examined an auditory oddball task during the transition to sleep (wake to stage 2) in 12 good sleepers and 13 subjects with sleep-onset insomnia. Smaller P2 amplitudes were found in the subjects with insomnia for the frequent to-be-ignored stimuli during pre-sleep wake, which was interpreted as failure to inhibit stimuli during the attempt to fall asleep. No differences were found during sleep.

Two studies aimed at directly assessing sensory gating in insomnia using paired-click stimuli. Two short auditory stimuli presented in rapid succession (paired clicks) usually show reduced ERP response to the second stimulus, indicating intact brain mechanisms for suppressing irrelevant stimuli (sensory gating, Freedman et al., 1983). The mid-latency component P50 is typically employed for this assessment.

Milner et al. (2009) examined P50 suppression during pre-sleep wakefulness, stage 2 and REM sleep in 13 good and 13 poor sleepers. The poor sleepers were not diagnosed with insomnia but similar criteria of subjectively poor sleep and related daytime complaints were used. In the poor sleepers, reduced gating was seen only in the pre-sleep waking period but neither during REM (where both groups showed clear gating) nor during stage 2 (where both groups did not show significant gating). This result is interpreted as showing heightened sensory processing in the pre-sleep period in poor sleepers.

Comparing 18 patients with primary insomnia to 20 good sleepers, Hairston et al. (2010) examined ERP amplitude differences between paired clicks in sleep stage 2 and wakefulness as well as the number of evoked K complexes in stage 2. When analyzing sleep stage transitions in stimulation relative to preceding non-stimulation blocks, patients with insomnia transitioned significantly less often to sleep stage 2 than good sleepers when coming from REM sleep and more often to REM sleep when coming from stage 2. Sensory gating was examined using P300 and later components in this study, which is unusual because these components do not reflect early sensory processing. The gating effect was significant and not different between groups during wake for both P300 and N350 components; during sleep, significant gating remained for N350 and P450 components in the good sleepers but the effect was reduced in patients with insomnia. In addition, waking but not sleep component amplitudes were reduced in insomnia. K complex numbers during non-stimulation blocks were tendentially higher in patients with insomnia but significantly lower than in good sleepers in stimulation blocks. Spindle counts were higher in patients with insomnia both with and without stimulation. The arousal index was not increased in the stimulation night relative to an acclimation night, with no group difference. The results are interpreted to show impaired stimulus suppression in insomnia, since gating was reduced and fewer sleep-protecting K complexes were elicited. The particular findings of this study within sleep are, however, in contrast to several other studies noted above which found no gating deficiency within sleep and no differences in sleep spindles or K complexes.

To summarize, ERP studies in insomnia are scarce, particularly those probing consolidated sleep stages. The existing studies appear to show an increased sensitivity to auditory stimuli in primary insomnia patients during both wake and sleep onset. An increased N100 amplitude is also observed intraindividually when slow negative potentials such as the contingent negative variation (CNV) are modulated, reflecting increased expectation and lowered response threshold (i.e., increased cortical excitability; Rockstroh et al., 1993; Wagner et al., 1996; Olbrich et al., 2002). Alternatively, the described results may also be viewed as a consequence of a compromised ability of patients with insomnia to inhibit alerting sensory inputs when trying to fall asleep.

4.7. Summary

Summarizing, quantitative electrophysiological methods have already highlighted a number of neurophysiological differences in patients with primary insomnia that extend well beyond the possibilities of normal sleep staging, which considers the whole brain to be in one of few discrete states in every single moment. These studies hold promise of explaining why the macro-analysis of EEG sleep according to Rechtschaffen and Kales criteria often fails to fully reflect the subjective complaints of disrupted sleep in patients with insomnia. Most of the electrophysiological studies noted above are generally limited by small sample sizes as well as inhomogeneity in selected patient subgroups and concomitant intake of medication. There is, however, a common trend towards signs of increased arousal or cortical excitability, which was recently also shown directly using transcranial magnetic stimulation (van der Werf et al., 2010).

Since the cortical state is part of a complex feedback system, a classical “chicken and egg” dilemma arises: it may be that the ongoing attentive cortical activity during sleep in primary insomnia causes an aroused state of subcortical sensory systems, with increased sensory sensitivity. Conversely, due to inevitably impinging sensory inputs during the night, reduced sensory thresholds lead to increased activation of the ascending reticular activation system and therefore to cortical arousal and reduced sleep quality. In consequence this may be the physiological basis of the succession of precipitating and perpetuating factors in insomnia (cf. Perlis et al., 1997) – the negative thoughts about being unable to sleep creating an internal “threat” similar to subjects trying to sleep while expecting danger, and the experience of this non-recuperative sleep leading to perpetuation by associating the negative expectations with sleep itself.

A principal lack of data can be concluded regarding the question in which sleep stage the neurophysiological differences are most pronounced, and possibly be most closely related to the pathophysiology of insomnia. The reason is methodological in nature – phasic events are fundamentally different between REM sleep and NREM sleep, perhaps with the exception of arousals. “Perhaps” because the different spectral composition of REM sleep and NREM sleep poses different difficulties for the staging of arousals in both sleep states which have to be borne in mind. Spindles, K-complexes and CAP are NREM sleep phenomena; by these measures, primary insomnia patients can only be deviant in NREM sleep. Also, spectral power modulating factors influence REM sleep differently from NREM sleep, so that a formal comparison is hardly possible. The search for differences is therefore an ongoing effort.

5. Discussion and conclusion

Conventional sleep staging uses a stereotyped and low-dimensional model to describe sleep. It is therefore well defined and comparable, but risks to neglect patterns outside its dimensions, which may well be the correlates of highly relevant properties of disordered sleep. This overview focused on the important restorative quality of sleep that can, to the present day, be only captured subjectively. Its disturbance is the hallmark of primary insomnia. R&K scoring alone has shown a relatively weak sensitivity to capture this important aspect of sleep. Quantitative analyses of sleep microstructure hold promise to assess differences which cause sleep
disorders with practically identical R&K properties to have a wide range of restorative qualities across subjects.

The current review summarized promising steps elucidating sleep features that were usually not quantitatively analyzed in insomnia PSG studies, mostly because special detection steps are required (spectral power, EEG graphoelements, CAP) or because fine-grained analysis is not usually applied (micro-arousals). These methods can give indications regarding cortical activation (spectral power, ERPs), stimulus-related deactivation (ERPs) as well as the dynamic properties of consolidated sleep (instability or resilience). In group comparisons, several approaches point towards sleep instability as an important difference (CAP, micro-arousals); together with further differences in spectral power and ERPs this can be reconciled into a broader picture of hyperarousal (or inhibition deficits) characterizing insomnia.

While suggested by a broad array of studies and methods, the concept of hyperarousal is relatively unspecific in itself. For maximal impact for optimizing therapeutic approaches, it is necessary to better understand the pathways and mechanisms involved (for an overview see Broese et al., 2012). If lowered thresholds to stimuli are central, management of such stimuli, desensitization or focused manipulation of sensory pathways might help. On the other hand, sleep mentation might be central, in which case approaches focused on sensory pathways would not be the best option. For further clarification it is important to characterize whether the wake state, NREM or REM sleep is most affected, but as outlined above a common methodology is lacking for this comparison.

A promising but underused approach is the intra-group correlation with quantitative measures of (subjective) sleep quality (Perlis et al., 2001b; Parrino et al., 2009). This is the most direct approach to the enigma of sleep quality and should be targeted in both longitudinal and transversal studies. Based on an analysis of the relationship between subjective and objective sleep parameters, Feige et al. (2008) found indications for an involvement of REM sleep in subjective wake duration. Furthermore, REM sleep percentage is decreased and the micro-arousal frequency increased within REM sleep, which led us to conclude that REM sleep might be qualitatively altered in insomnia. REM sleep in insomnia may be especially prone to disturbance, leading to the “REM sleep instability” hypothesis of insomnia (Rieman et al., 2012).

On the other hand, NREM sleep instability has more frequently been implicated, as evidenced by increased EEG beta power, arousal and CAP rates and the direct link to misperception (Parrino et al., 2009).

Furthermore, while NREM sleep has been found to be more sensitive to external stimulation in patients with insomnia in some of the ERP studies (Bastien et al., 2008 at least in the sleep onset period; Hairston et al., 2010), REM sensory sensitivity or gating was generally not found different. The stage transition data of Hairston et al. (2010) even suggest a preferential occurrence of REM sleep under auditory stimulation in insomnia. In contrast, an early study on auditory arousal thresholds in insomnia found reduced thresholds and more subjective reports of having been awake in both REM sleep and NREM sleep, with REM sleep being rather more affected (Mendelson et al., 1986).

According to the REM sleep line of thinking, this most highly activated brain state during sleep (Maquet et al., 1996), requiring a very delicate balance of arousing/de-arousing brain activity, may be frequently interrupted and curtailed by a permanently increased arousal level and thus experienced subjectively more like waking than dreaming or sleeping in insomnia. Interestingly, the continuity hypothesis of dream production suggests that pre-sleep concerns of patients with insomnia, i.e. worries about poor sleep and its consequences, should dominate their dream content, which is in line with up to date (however only scarce) evidence on dream content in insomnia (Schredl et al., 1998). Enhanced arousal (especially evidenced by the increased frequency of micro-arousals) during REM sleep may render these wake-like cognitions more accessible to conscious perception, memory storage and morning recall. This may result in the subjective overestimation of nocturnal waking time and the experience of disrupted and nonrestorative sleep. Speculatively, the chronic fragmentation of REM sleep might lead to dysfunction in a ventral emotional neural network, including limbic and paralimbic areas that are specifically activated during REM sleep. This dysfunction might contribute to emotional and cognitive alterations and an elevated risk of developing depression (Nissen and Nofzinger, 2006; Baglioni et al., 2010, 2011; Baglioni and Riemann, 2012).

In summary, current research on sleep microstructure in primary insomnia indicates increased instability of both NREM sleep and REM sleep. In addition, spectral power analysis indicates a higher cortical arousal level and event-related potentials show increased responses to sensory stimuli during wake and sleep onset. These observations might have a common cause in a subcortical dysregulation towards higher alertness; it is of utmost interest both for treatment options and basic research to further elucidate how the observed physiological changes in primary insomnia relate to the primary complaints of the patients.

References


