

## **4. Diskussion**

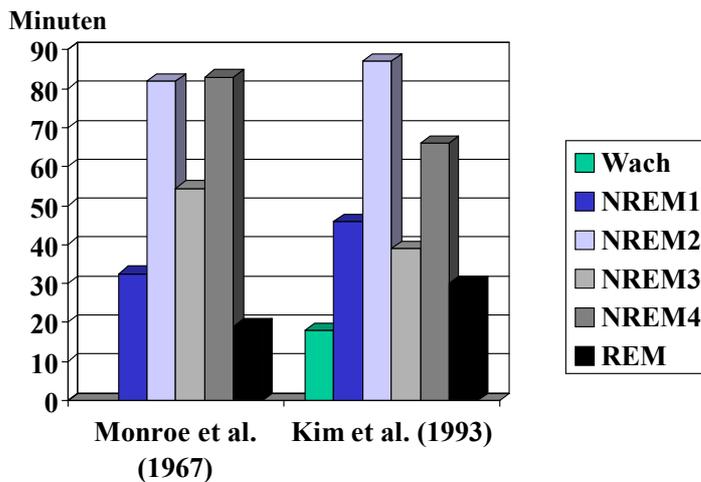
### 4.1 Interrater-Reliabilität

In der vorliegenden Arbeit wurde zunächst untersucht, inwieweit Abweichungen in der Verteilung der quantitativen Schlafparameter zwischen dem Auswertenden der Somnopolygraphie und den Auswertenden des fronto-polaren Signals bestehen. Darüber hinaus wurden die Verteilungen der paarweise für individuelle Nächte berechneten Differenzen zwischen den einzelnen Experten bestimmt. Die Auswertung ergab signifikante Verteilungsunterschiede für die Parameter NREM1 und NREM2 zwischen dem zweiten Auswertenden des fronto-polaren Signals (QUISI 2) und den beiden anderen Experten. Für die sieben weiteren berücksichtigten Schlafparameter ließen sich keine signifikanten Verteilungsunterschiede feststellen.

Bei der Berechnung der Paardifferenzen der einzelnen Schlafparameter fanden sich in der vorliegenden Arbeit keine signifikanten Unterschiede zwischen allen drei Auswertenden lediglich für den Parameter Einschlafzeit. Der erste Auswertende des fronto-polaren Signals (QUISI 1) erzielte bei den Parametern Schlafperiodenzeit, effektive Schlafzeit, Schlaf Effizienz Index und NREM2 im Mittel höhere Werte als der Auswertende der Somnopolygraphie (SPG-Experte); bei den Parametern Wach und SWS fanden sich im Mittel niedrigere Werte. Beim zweiten Auswertenden des fronto-polaren Signals (QUISI 2) fanden sich gegenüber dem SPG-Experten durchschnittlich höhere Werte für die Parameter Wach, NREM2 und REM bzw. niedrigere Werte bei den Parametern Effektive Schlafzeit, Schlaf Effizienz Index sowie NREM1. Zwischen den beiden Auswertenden des fronto-polaren Signals ließen sich bei den Parametern Einschlafzeit und REM keine signifikanten Unterschiede feststellen. QUISI 1 erreichte bei den Parametern Schlafperiodenzeit, Effektive Schlafzeit, Schlaf-Effizienz Index und NREM1 im Mittel höhere, für die Parameter Wach, NREM2 und SWS niedrigere Werte als der Auswertende QUISI 2.

Andere Arbeiten, die sich mit der Übereinstimmung von Auswertungen des menschlichen Schlafs befassten, fanden zum Teil ausgeprägte Unterschiede zwischen den einzelnen Experten. In einer Studie (Monroe, 1968) wurden 398 Minuten des Schlafs eines jungen gesunden Probanden durch 28 Experte bewertet. Hier variierten die Angaben für das Stadium

NREM1 im Bereich von 32,5 Minuten, bei Stadium NREM2 82,0 Minuten, bei den Stadien NREM3 und NREM4 54,5 bzw. 83,0 Minuten; für Stadium REM wurde eine Streubreite der Werte von 19,0 Minuten berechnet (Abb. 29). Ähnlich große Streubreiten fanden sich bei einer neueren Studie (Kim et al., 1993), bei der zehn verschiedene Auswertende den Schlaf von drei jungen gesunden männlichen Probanden bewerteten. Die Angaben für die einzelnen Schlafstadien variierten bei einzelnen Scorings im Bereich von 46 Minuten (NREM1), 87 Minuten (NREM2), 39 Minuten (NREM3) und 66 Minuten (NREM 4). Für die Stadien Wach und REM wurden hier Streubreiten von 18 bzw. 30 Minuten ermittelt (Abb. 29). Dagegen bei einer Untersuchung von zehn Patienten mit Obstruktivem Schlafapnoe Syndrom durch drei Experten kamen die Autoren auf der Basis einer deskriptiven vergleichenden Darstellung der expertenspezifischen Mittelwerte der quantitativen Schlafparameter zu dem Schluss, dass für fast alle Variablen (Ausnahme: effektive Schlafzeit, Streuung 20 Minuten) ein hohes Maß an Übereinstimmung zu beobachten war (McDannold et al., 1990).



**Abbildung 29** – maximale Abweichung zwischen den Auswertungen (in Minuten) bezogen auf die einzelnen Schlafstadien

Die oben genannten Studien, bei denen Auswertungen schlafgestörter Probanden (Obstruktives Schlafapnoe Syndrom) besser miteinander übereinstimmen als

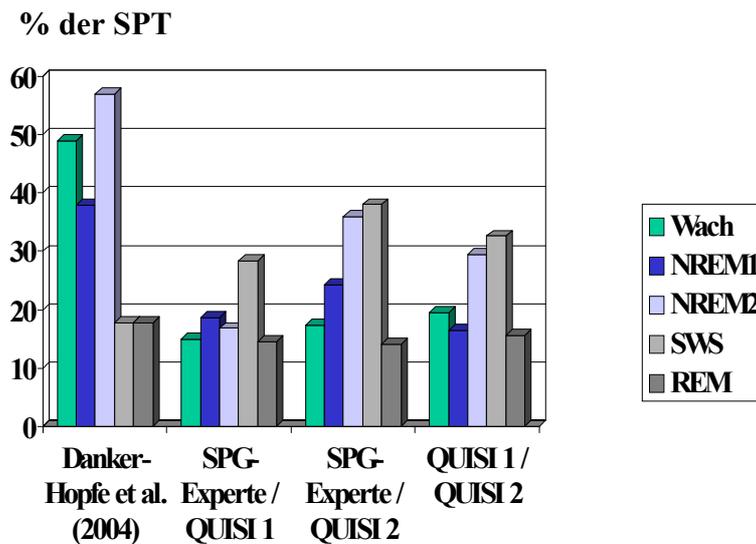
die von gesunden Individuen, stehen im Gegensatz zu den Ergebnissen einer Reihe neuerer Untersuchungen: Gemessen an absoluten Differenzen quantitativer Schlafparameter, sowie Intraklassenkorrelationen fanden Kunz et al. (2000) in einer Studie von 172 Aufzeichnungen gesunder Probanden insgesamt bessere Übereinstimmungen zwischen den Auswertungen als Danker-Hopfe et al. (2004) bei der Untersuchung von 196 Aufzeichnungen schlafgestörter Individuen. Bei Danker-Hopfe et al. (2004) wurden für die Parameter Schlafperiodenzeit und effektive Schlafzeit mittlere Abweichungen zwischen den verschiedenen Auswertungen von

2,0 bzw. 13,0 Minuten sowie maximale Abweichungen von 166,5 bzw. 264,0 Minuten berechnet. Die Mediane der Parameter Einschlafzeit und Schlaf Effizienz Index betragen 1,5 Minuten bzw. 2,8 % (Max 82,5 Minuten bzw. 54,2 %). Für die Mediane der Abweichungen die prozentualen Anteile der Schlafstadien betreffend ermittelten Danker-Hopfe et al. (2004) folgende Werte: 2,7 % (Wach), 4,7 % (NREM1), 5,9 % (NREM2), 3,2 % (SWS) und 2,0 % (REM).

Die vergleichende Betrachtung der Ergebnisse der vorliegenden Arbeit mit denen anderer Untersuchungen zum Thema Interrater-Reliabilität ist durch die Anwendung teils unterschiedlicher statistischer Methoden nur eingeschränkt möglich. In Übereinstimmung mit den oben genannten Studien lässt sich feststellen, dass bei der Beurteilung von Schlaf durch zwei oder mehrere Experten ein gewisses Maß an Abweichung zwischen den Auswertungen die quantitativen Schlafparameter betreffend nicht zu vermeiden ist. Das von McDannold et al. (1990) gefundene hohe Maß an Übereinstimmung konnte in der vorliegenden Arbeit für den Parameter Einschlafzeit bei allen drei Paarvergleichen bestätigt werden. Dies gilt auch für den Parameter NREM1 beim Paarvergleich SPG-Experte - QUISI 1, für die Parameter Schlafperiodenzeit und SWS beim Paarvergleich SPG-Experte - QUISI 2 sowie für den Parameter REM beim Paarvergleich QUISI 1 - QUISI 2. Außerdem fanden sich in der vorliegenden Untersuchung für die Beträge der Differenzen der quantitativen Schlafparameter (Anhang: Tab.9a-i) bis auf einige Ausnahmen bei allen drei Paarvergleichen im Mittel durchweg geringere oder gleiche Abweichungen zwischen den jeweiligen Auswertungen wie bei Danker-Hopfe et al. (2004): Beim Paarvergleich SPG-Experte - QUISI 1 betragen die Mediane der Abweichungen der Parameter Schlafperiodenzeit und Einschlafzeit ebenfalls 2,0 Minuten. Beim Paarvergleich SPG-Experte - QUISI 2 stimmten die Mediane der Abweichungen der Parameter Schlafperiodenzeit und prozentualer Wachanteil mit denen von Danker-Hopfe et al. (2004) überein.

Im Gegensatz zu Danker-Hopfe et al. (2004) zeigten sich in der vorliegenden Studie beim Paarvergleich SPG-Experte - QUISI 1 größere Werte für den Median (3,7 %) und das Maximum (28,3 %) des Anteils an Stadium SWS. Beim Paarvergleich SPG-Experte - QUISI 2 fanden sich für den Median der Abweichungen das Stadium NREM1 betreffend (5,3 %) und den Median und das Maximum der Stadium SWS-Abweichungen (3,4 bzw. 38,0 %) höhere

Werte. Der Paarvergleich QUISI 1 - QUISI 2 ergab höhere Mediane der Abweichungen für die Parameter Schlafperiodenzeit (2,25 Minuten) und prozentualer Anteil an Stadium NREM1 (5,4 %). Darüber hinaus betrug das Maximum, das für den Anteil an Stadium SWS ermittelt wurde, mit 32,6 % ebenfalls mehr als bei Danker-Hopfe et al. ( Maximum dort: 17,7 %) – siehe Abb. 30.



**Abbildung 30** – maximale Abweichung zwischen den Auswertungen (in % der SPT) bezogen auf die einzelnen Schlafstadien

Zu begründen sind die geringen Diskrepanzen zur oben genannten Studie ebenfalls am ehesten dadurch, dass sich aufgrund

der Ungenauigkeit der Regeln von Rechtschaffen und Kales (1968) ein gewisses Maß an Abweichung zwischen Auswertenden von Schlaf-EEGs nicht vermeiden lässt. Die teils starken maximalen Abweichungen, die für das Stadium SWS nachgewiesen wurden, sind am ehesten durch die unterschiedliche morphologische Ausprägung der Delta-Aktivität in den verwendeten EEG-Spuren zu erklären (siehe Kapitel 4.2.3), die das Auswerten des Schlafs bestimmter Individuen erschwert.

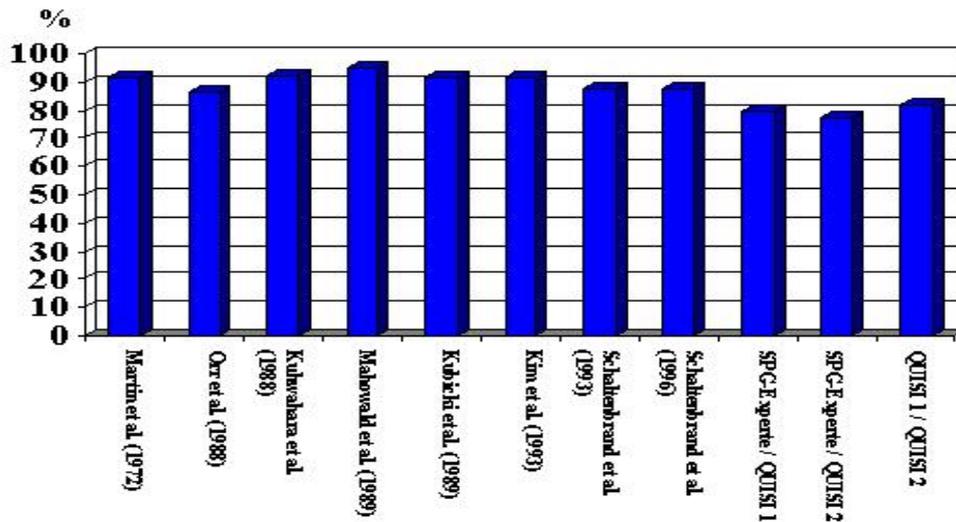
Die im Vergleich zu den oben genannten Studien größtenteils geringeren Abweichungen zwischen den Experten der vorliegenden Arbeit legen nahe, dass eine Beurteilung des menschlichen Schlafs anhand eines einzelnen von der Stirn abgeleiteten EEGs möglich ist, wobei akzeptable Übereinstimmungen mit der jeweiligen Referenz-Auswertung (SPG-Experte) sowie zwischen den beiden Auswertenden des Fp1-Fp2-Signals untereinander erreicht werden. Zu berücksichtigen ist hierbei jedoch, dass im Gegensatz zu der von Danker-Hopfe et al. (2004) durchgeführten Untersuchung in der vorliegenden Arbeit auch ein Anzahl

von Personen ( $n=19$ ) miteinbezogen wurde, bei denen man keine klassische Schlafstörung feststellen konnte, was sich eventuell in einer etwas höheren Interrater Reliabilität abzeichnen könnte (Danker-Hopfe und Herrmann, 2001). Dass dennoch bei den meisten berücksichtigten Parametern statistisch signifikante Unterschiede zwischen den einzelnen Experten nachgewiesen werden konnten ist am wahrscheinlichsten auf eine generelle Unzulänglichkeit der Regeln für das visuelle Bewerten von Schlaf (Rechtschaffen und Kales, 1968) zurückzuführen (siehe unten).

Die vergleichende Betrachtung der Übereinstimmung zweier oder mehrer Auswertender auf der Basis von Epochen stellt einen weiteren hier verwendeten methodischen Ansatz zur Bestimmung der Interrater Reliabilität dar. Um auszuschließen, dass zufällige Übereinstimmungen zwischen den Auswertenden Einfluss auf die Interrater-Reliabilität nehmen können, wurde in der vorliegenden Arbeit der Kappa-Koeffizient nach Cohen (für zwei Beobachter) verwendet. Die hier berechneten Kappa-Werte für die Paarvergleiche SPG-Experte - QUISI 1 ( $\kappa=0,72$ ) bzw. SPG-Experte - QUISI 2 ( $\kappa=0,70$ ) ergaben ebenfalls jeweils eine nach den Kriterien zur Evaluation von Kappa von Landis und Koch (1977) „starke“ Gesamtübereinstimmung zwischen den Auswertenden. Auch der Wert für die Interrater-Reliabilität zwischen den beiden Auswertenden des fronto-polaren Signals (QUISI 1 - QUISI 2) lag mit  $\kappa=0,75$  im Bereich für „starke“ Übereinstimmung. Weiterhin ergaben sich bei der hier vorliegenden Studie für die Gesamtübereinstimmung zwischen zwei Auswertenden auf der Basis von einzelnen Epochen prozentuale Werte von 79,5 % (SPG-Experte - QUISI 1), 77,7 % (SPG-Experte - QUISI 2) und 81,9 % (QUISI 1 - QUISI 2) von übereinstimmend gewerteten Epochen.

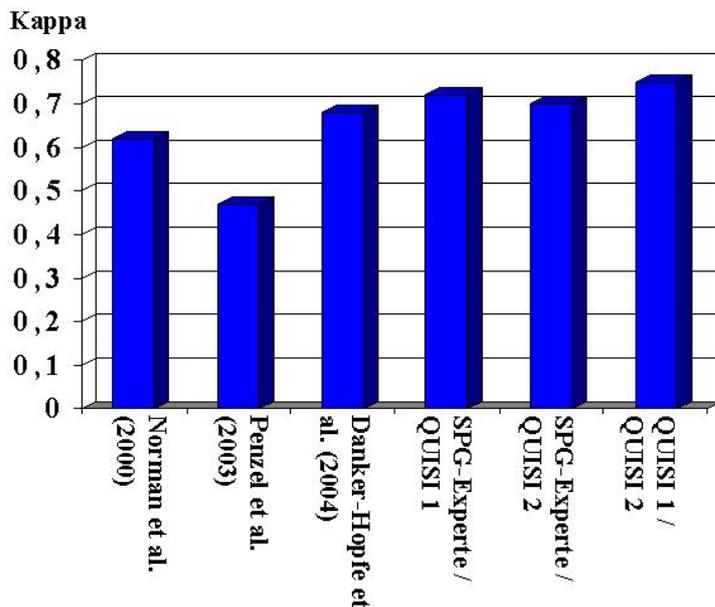
Die meisten Studien, die sich mit dieser Thematik beschäftigen, geben die ermittelten Werte in Prozent der miteinander in Einklang stehenden Epochen an (Martin et al, 1972; Orr et al., 1988; Kuhwahara et al., 1988; Mahowald et al., 1989; Kubicki et al., 1989; Kim et al., 1993; Schaltenbrand et al., 1993 und 1996) – siehe Abb. 31. Die Werte für die Gesamtübereinstimmungen für Auswertungen einzelner Somnopolygraphien variieren zwischen 60,9 und 96,0 %, wogegen bei Gruppenstatistiken die Angaben von 81,3 % bis 95,7 % reichen. Generell ist festzustellen, dass die Übereinstimmungen zwischen Auswertenden unter

Verwendung verschiedener statistischer Verfahren berechnet wird und deshalb nicht immer miteinander verglichen werden kann. Die Ergebnisse werden einerseits durch die Herkunft und Anzahl der Auswertenden, andererseits durch die Anzahl der unterschiedenen Schlafstadien, die Stichprobengröße sowie die Auswahl des statistischen Verfahrens beeinflusst.



**Abbildung 31** – Gesamtübereinstimmungen zwischen den Auswertungen (in %) bezogen auf alle Epochen

Einige Voruntersuchungen verwendeten ebenfalls den Kappa-Koeffizienten nach Cohen bzw. nach Fleiss (für mehrere Beobachter) (Norman et al., 2000; Kunz et al., 2000; Penzel et al., 2003; Danker-Hopfe et al., 2004) zur Berechnung der Interrater-Reliabilität. Norman et al. (2000) ermittelten in einer Untersuchung zur Auswertung des Schlafs von 62 Personen unter Verwendung des Fleiss-Kappa-Koeffizienten für mehrere Beobachter insgesamt eine „deutliche“ Übereinstimmung ( $\kappa=0,59$ ) bei Patienten mit Verdacht auf obstruktive schlafbezogene Atemstörung (OSAS) und eine „starke“ Übereinstimmung ( $\kappa=0,65$ ) bei gesunden Individuen. Eine nach den Kriterien zur Evaluation von Kappa von Landis und Koch (1977) im Wesentlichen „deutliche“ Übereinstimmung für drei Gesunde ( $\kappa=0,52$ ,  $\kappa=0,54$ ,  $\kappa=0,62$ ) und zwei Patienten mit PLMS-Syndrom ( $\kappa=0,58$ ,  $\kappa=0,59$ ), deren Schlaf von neun verschiedenen Experten begutachtet wurde, fand sich in einer Studie von Penzel et al. (2003). Bei drei Patienten mit OSAS wurde dabei im Wesentlichen eine nur „schwache“ Interrater-Reliabilität ( $\kappa=0,25$ ,  $\kappa=0,22$ ,  $\kappa=0,47$ ) errechnet – siehe Abb. 32.



**Abbildung 32** –  
Gesamtübereinstimmung  
aller Auswertungen auf  
der Basis einzelner  
Epochen

Die einzelnen Schlafstadien betreffend fanden sich bei Penzel et al. (2003) eine nach den Kriterien zur Evaluation von Kappa von Landis und Koch (1977) fast „vollständige“ Übereinstimmung für das Stadium REM, eine „deutliche“ bis „starke“ Übereinstimmung für SWS und NREM2 (bei gesunden und PLMS-Patienten). Für das Stadium Wach variierte die Interrater-Reliabilität zwischen „starke“ (PLMS-Patienten), „schwache“ (gesunde Probanden) und keiner Übereinstimmung (OSAS-Patienten). Beim Stadium NREM1 konnten Penzel et al. (2003) lediglich eine nach den Kriterien zur Evaluation von Kappa von Landis und Koch (1977) „schwache“ bzw. keine Übereinstimmung zwischen den Experten ermitteln. In der bereits erwähnten Untersuchung von Danker-Hopfe et al. (2004) ergab sich eine „starke“ Gesamtübereinstimmung ( $\kappa=0,68$ ) für 196 Aufzeichnungen schlafgestörter Patienten. Hier ließ sich für das Stadium REM eine „vollständige“ Übereinstimmung finden ( $\kappa=0,83$ ); die Interrater-Reliabilität der Stadien Wach ( $\kappa=0,78$ ), NREM2 ( $\kappa=0,67$ ) und SWS ( $\kappa=0,71$ ) lag im Bereich für „starke“ Übereinstimmung, wogegen die Experten bei NREM1 lediglich „schwach“ übereinstimmten. Diese Ergebnisse lagen durchweg etwas unter denen der Studie von Kunz et al. (2000), die 172 Nächte gesunder Individuen als Ausgangsdaten verwendeten.

Die in der vorliegenden Arbeit erzielten Ergebnisse liegen also im Bereich derer der oben genannten Studien bei denen Übereinstimmungen zwischen mehreren Auswertenden von

Somnopolygraphien untersucht wurden, bzw. übertreffen diese. Auch die Ergebnisse der schlafstadienspezifischen Übereinstimmungen zwischen den Auswertenden der vorliegenden Arbeit liegen durchweg im Bereich derer der oben genannten Studien, so dass es praktikabel erscheint, menschlichen Schlaf anhand eines einzigen von den Elektrodenpositionen Fp1 und Fp2 abgeleiteten EEG-Kanals visuell zu klassifizieren.

## 4.2 Die Schlafstadien

### 4.2.1 Stadium Wach und NREM1

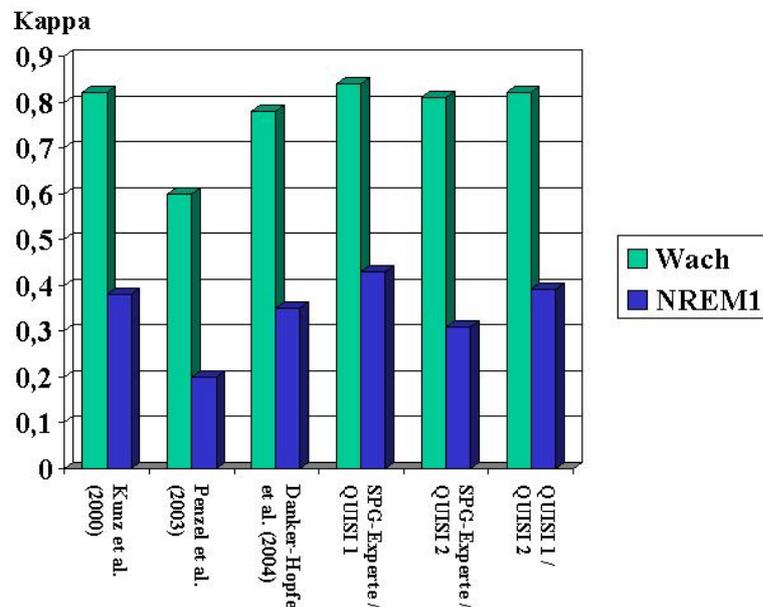
Für das Stadium Wach wurde bei allen Paarvergleichen eine gute Übereinstimmung erzielt ( $\kappa=0,84$ ;  $0,81$ ;  $0,82$ ). Die besten Übereinstimmungen fanden sich dabei für Schlafableitungen, die einen hohen Anteil an Stadium Wach aufwiesen und bei denen schnelle Augenbewegungen (Abb. 18) und Blinzeln (Abb. 19) eindeutig dieses Stadium kennzeichnen. Schwächen bei der Beurteilung von Stadium Wach mittels einer einzelnen fronto-polar abgeleiteten EEG-Spur lassen sich exemplarisch anhand von zwei Schlafableitungen aufzeigen: Eine für das Stadium Wach lediglich „schwache“ Übereinstimmung ( $\kappa=0,37$  resp.  $0,26$ ) zwischen SPG-Experte und QUISI 1 fand sich bei Ableitungen, in denen ein fragmentiertes Schlafprofil mit vielen Stadienwechseln und kurzen REM-Episoden zur Darstellung kamen (Anhang: Abb. 33a, b und Abb. 34). In diesen führte zudem das Auftreten von phasischer schneller Beta-Aktivität sowie schnellen Augenbewegungen fälschlicherweise zur Bewertung als Stadium Wach (Anhang: Abb. 35).

Für Stadium NREM1 fand sich in der vorliegenden Untersuchung mit Abstand die geringste Übereinstimmung bei allen Paarvergleichen. Während SPG-Experte und QUISI 1 noch eine „deutliche“ Übereinstimmung erzielten ( $\kappa=0,43$ ), kann die Interrater-Reliabilität zwischen SPG-Experte und QUISI 2 ( $\kappa=0,31$ ) bzw. QUISI 1 und QUISI 2 ( $\kappa=0,39$ ) nur als „schwach“ bezeichnet werden. Besonders schlechte Übereinstimmungen bezüglich Stadium NREM1 ( $\kappa<0,1$ ) wurden in der vorliegenden Arbeit bei Patienten mit Erkrankungen erzielt, die durch ein stark fragmentiertes Schlafprofil mit vielen Stadienübergängen charakterisiert sind (Beispiel Patient mit PLMS; Anhang: Abb. 36). Insbesondere stellten sich hier die durch Myoklonien ausgelösten Beschleunigungen der EEG-Aktivität (Arousal) im fronto-polar

abgeleiteten EEG anders dar und wurden so zum Teil nicht als solche erkannt (Anhang: Abb. 37).

Analog dazu konnte in Vorarbeiten von Kunz et al. (2000) und Danker-Hopfe et al. (2004) für schlafgestörte Patienten errechnete Interrater-Reliabilität ( $\kappa=0,78$ ) eine relativ hohe Übereinstimmung zwischen den Experten für das Stadium Wach sowie eine niedrige Übereinstimmung für das Stadium NREM1 nachgewiesen werden (Abb. 38). Penzel et al.

(2003) hingegen fanden für die beiden Parameter deutlich niedrigere Werte.



**Abbildung 38** – Gesamtübereinstimmung der Auswertungen auf einzelne Epochen bezogen für die Stadien Wach und NREM1

Unser Ergebnis einer niedrigen Interrater-Reliabilität in der Beurteilung des Schlaf-

stadiums NREM1 repliziert somit eine Reihe bereits publizierter Vorarbeiten, in denen postuliert wurde, dass Diskrepanzen in der Beurteilung Folge nicht eindeutiger Definitionskriterien des Klassifikationssystems von Rechtschaffen und Kales (1968) seien (Himanen und Hasan, 2000), die zudem von der Erfahrung des Auswerters abhängig seien (Monroe, 1967).

Während Rechtschaffen und Kales (1968) den Wach–Schlaf–Übergang mit der Beschreibung des Wegfalls der Alpha-Aktivität im Stadium NREM1 als hinreichend definiert sahen und das Auftreten von langsamen Augenbewegungen (slow eye movements- SEMs) auf das Stadium NREM1 beschränkten, wird in der aktuellen Literatur eine detailliertere Unterteilung dieses Abschnitts gefordert. Während Himanen (2000) eine Aufgliederung des Wach-Schlaf-Übergangs in acht unterschiedliche Stadien vorschlägt - wobei insbesondere auch das

Auftreten von schnellen und langsamen Augenbewegungen im Stadium Wach einbezogen wird – unterteilen Hori et al. (1991; 1994) diesen Zeitraum sogar in neun verschiedene Abschnitte (Tab. 10).

Himanen (2000)		Hori et al. (1991; 1994)	
Stadium	Merkmale	Stadium	Merkmale
Wake-low	EEG gemischter Frequenz und niedriger Spannung mit schnellen oder gar keinen Augenbewegungen	Alpha wave train	Alpha-Aktivität mit mind. 20 $\mu$ V Amplitude
Wake-alpha	posteriore Alpha-Aktivität mit schnellen oder gar keinen Augenbewegungen	Alpha wave intermitt (A)	intermittierende Alpha-Aktivität >50% der Epoche mit mind. 20 $\mu$ V Amplitude
Wake-alpha-F	diffuse oder fronto-zentrale Alpha-Aktivität mit schnellen oder gar keinen Augenbewegungen	Alpha wave intermitt (B)	intermittierende Alpha-Aktivität <50% der Epoche mit mind. 20 $\mu$ V Amplitude
Alpha-SEM	posteriore Alpha-Aktivität mit langsamen Augenbewegungen (SEMs)	EEG flattening	unterdrückte Alpha-Aktivität mit Amplitude <20 $\mu$ V
Alpha-SEM-F	diffuse oder fronto-zentrale Alpha-Aktivität mit langsamen Augenbewegungen (SEMs)	Ripples	Theta-Aktivität niedriger Amplitude (20-50 $\mu$ V), "burst suppression" ohne Vertex-Wellen
Drowsy-low	EEG gemischter Frequenz und niedriger Spannung mit langsamen Augenbewegungen	Vertex sharp wave solitary	Epoche mit einer gut definierten Vertex-Welle
S1	EEG mit zunehmender Theta-Aktivität	Vertex sharp wave train or burst	Epoche mit mind. zwei gut definierten Vertex-Wellen
		Vertex sharp wave and incomplete spindle	Epoche mit mind. einer gut definierten Vertex-Welle und einer inkompletten Spindel (<0,5s, 10-20 $\mu$ V)
S2	Spindeln und/oder K-Komplexe	Spindles	Epoche mit mind. einer gut definierten Spindel (>0,5s, >20 $\mu$ V)

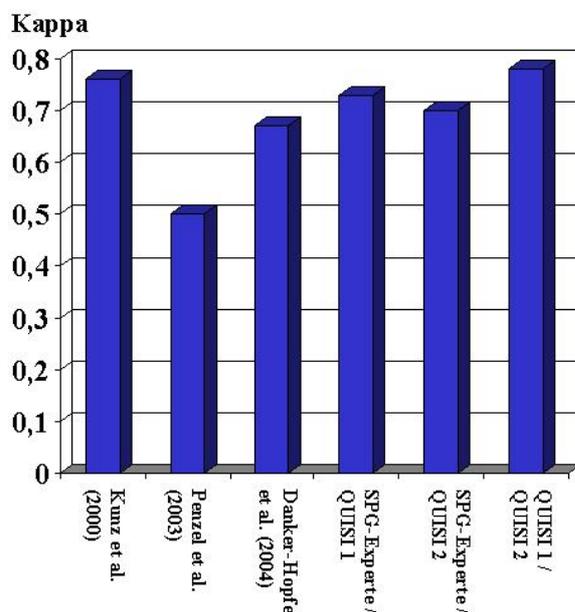
**Tabelle 10** – Unterschiedliche Einteilungen des Wach-Schlafüberganges

Dies verdeutlicht die Komplexität des Übergangs vom Wach zum Schlaf und die Schwierigkeit der Beurteilbarkeit mittels nur einer einzigen fronto-polaren EEG-Spur. Die unzureichende Einbindung der in Tabelle 10 genannten neurophysiologischen Parameter erklärt die schlechten Interrater-Reliabilitäten in der vorliegenden Untersuchung und den Voruntersuchungen. Ein zusätzlicher, für die fronto-polare EEG-Spur spezifischer Parameter ist die Schwierigkeit der Ableitbarkeit eines frontalen Alpha-EEGs. Hilfreiche Parameter

hingegen sind horizontale Augenbewegungen sowie durch Muskelaktivität hervorgerufene schnelle Betafrequenzen.

#### 4.2.2 Stadium NREM2

Die Gesamt-Kappa-Werte für das Schlafstadium NREM2 lagen den Kriterien zur Evaluation von Kappa von Landis und Koch (1977) folgend bei allen drei Paarvergleichen im Bereich für „starke“ Übereinstimmung ( $\kappa=0,73$ ;  $0,70$ ;  $0,78$ ), wobei beim Vergleich der beiden Auswertenden des fronto-polaren EEGs insgesamt fast eine „vollständige“ Interrater-Reliabilität erreicht wurde. Vergleicht man diese Ergebnisse mit denen anderer Untersuchungen (Penzel et al., 2003; Danker-Hopfe et al., 2004), lässt sich feststellen, dass auch für das Stadium NREM2 im Rahmen der Möglichkeiten, die eine visuelle Bewertung des Schlafs nach den international gültigen Regeln bietet, eine gute Interrater-Reliabilität erzielt worden ist, und es lässt sich schließen, dass die Regeln von Rechtschaffen und Kales (1968) auf das von der Stirn abgeleitete Signal in modifizierter Form anwendbar sind (Abb.39).



**Abbildung 39** – Gesamtübereinstimmung der Auswertungen auf einzelne Epochen bezogen für das Stadium NREM2

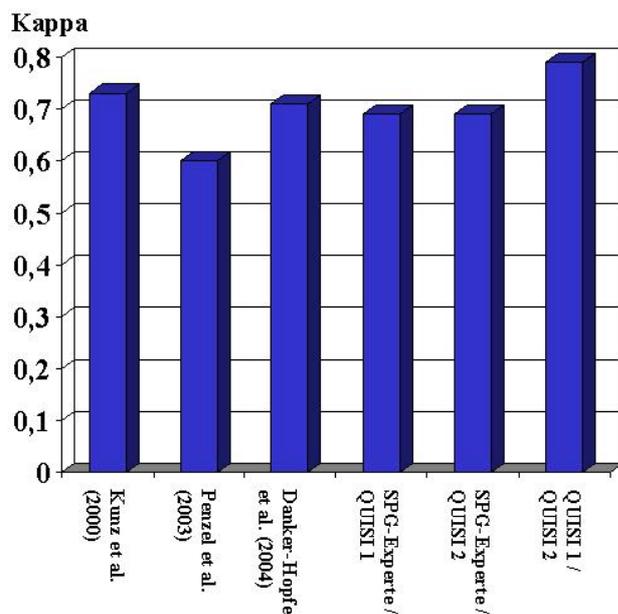
Die Ergebnisse der vorliegenden Arbeit zeigen eine besonders gute („vollständige“) Übereinstimmung zwischen zwei Auswertenden immer dann, wenn das EEG eines Individuums die für das Stadium NREM2 charakteristischen Graphoelemente in hoher Anzahl und deutlicher Ausprägung enthält. Exemplarische Ausschnitte mit gut zur Darstellung

kommenden Schlafspindeln und K-Komplexen finden sich in Abbildung 40 und 41 des Anhangs.

Schwierigkeiten bei der Beurteilung des Schlafs bezüglich NREM2 traten bei drei der einhundert ausgewerteten Nächte auf; hier wurden lediglich „schwache“ Übereinstimmungen erzielt. So betrug der Kappa-Wert für NREM2 bei der Auswertung des Schlafs einer 56-jährigen Patientin mit Syndrom der vorverlegten Schlafphase beim Paarvergleich SPG-Experte - QUI SI 1  $\kappa=0,23$  und beim Paarvergleich SPG-Experte - QUI SI 2  $\kappa=0,29$ . Die beiden Auswertenden des fronto-polaren Signals (QUI SI 1 - QUI SI 2) untereinander erreichten für NREM2 jedoch eine „starke Übereinstimmung“ ( $\kappa=0,75$ ). Die Diskrepanz zwischen Somnopolygraphie- und Fp1-Fp2-Auswertung von  $>14\%$  (der Schlafperiodenzeit) lässt sich durch eine systematische Fehlbeurteilung des EEG durch den SPG-Experten erklären, der trotz erfülltem Frequenz- und Amplitudenkriterium kein Stadium SWS, sondern NREM2 wertete (Anhang: Abb. 42). Ebenfalls eine nur „schwache“ Interrater-Reliabilität zwischen zentraler und fronto-polarer Auswertung wurde bei der Bewertung des Schlafs einer 30-jährigen Patientin mit REM-Schlafverhaltensstörung für das Stadium NREM2 erzielt ( $\kappa=0,36; 0,37$ ). Hier lag die Interrater-Reliabilität zwischen QUI SI 1 und QUI SI 2 mit  $\kappa=0,60$  ebenfalls deutlich höher. Eine Fehlklassifizierung des Stadium SWS als NREM2 durch den Auswertenden der Somnopolygraphie (SPG-Experte) ist auch bei dieser Auswertung als Ursache der Abweichung von  $>6\%$  (der Schlafperiodenzeit) für NREM2 zu sehen (Anhang: Abb. 43). Erschwerend zur korrekten Beurteilung des Schlafs als Stadium NREM2 kommt bei dem genannten EEG noch hinzu, dass äußerst wenig Schlafspindeln bei der Patientin zu identifizieren sind. Die dritte Aufzeichnung bei der, hier jedoch nur zwischen QUI SI 2 und SPG-Experte, lediglich eine „schwache“ Übereinstimmung für Stadium NREM2 errechnet wurde ( $\kappa=0,32$ ), stammt von einer 30-jährigen Patientin, die mit Verdacht auf schlafbezogene epileptische Anfälle im Schlaflabor untersucht wurde. Die besonders im fronto-polar abgeleiteten EEG auftretende Spike-Wave-Komplexe mit einer Frequenz von 1-1,5 Hz (Anhang: Abb. 44) wurden vom Auswertenden QUI SI 2 falsch als reguläre Deltawellen gewertet und die entsprechenden Abschnitte der Nacht als Stadium SWS anstatt NREM2 gewertet.

### 4.2.3 Stadium SWS

Die Gesamtübereinstimmung bezogen auf das Stadium SWS lag beim Vergleich der beiden Auswertenden des fronto-polaren EEGs mit dem Auswertenden der Somnopolygraphie ( $\kappa=0,69$ ) etwas unter der von Danker-Hopfe et al. (2004) für Patienten bestimmten Interrater-Reliabilität ( $\kappa=0,71$ ). Die beiden Auswertenden des fronto-polaren EEG-Kanals untereinander



erreichten mit einem Kappa-Werte von  $\kappa=0,79$  jedoch eine vergleichsweise gute Übereinstimmung für das Stadium SWS (Abb. 45).

**Abbildung 45** – Gesamtübereinstimmung der Auswertungen bezogen auf einzelne Epochen für das Stadium SWS

Da das Stadium SWS allein durch die Charakteristik des EEGs bestimmt wird, also keinerlei Zusatzinformationen wie EMG oder EOG benötigt werden, um SWS zu werten, lässt sich die Diskrepanz

zwischen den genannten Paarvergleichen am ehesten durch die unterschiedliche morphologische Ausprägung der Delta-Aktivität in den verwendeten EEG-Spuren erklären. Man geht davon aus, dass die Delta-Aktivität im frontalen Bereich ein eher monomorphes Bild abgibt und bisweilen sogar früher in Erscheinung tritt, als zentral abgeleitetes Delta (Naitoh et al., 1970; Beier und Kubicki, 1987).

Trotz der relativ guten Übereinstimmung der beiden Auswertenden des Stirn-Signals ließ sich beim Stadium SWS in der vorliegenden Untersuchung von allen Schlafstadien die größte Streubreite der Kappa-Werte finden. Die visuelle Analyse einzelner Aufzeichnungen, bei denen teilweise keine Übereinstimmung für das Stadium SWS erzielt werden konnte, ergab, dass Schwierigkeiten bei der korrekten Interpretation des Stadiums durch Fehler einzelner

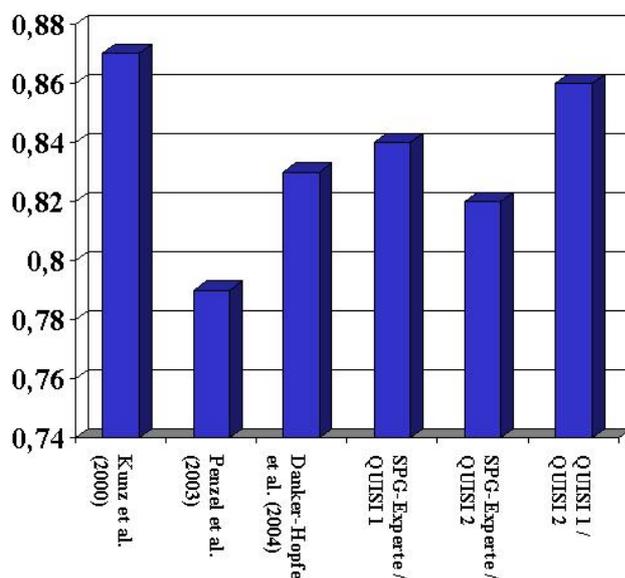
Auswertender in Bezug auf die Bewertung der Delta-Aktivität auszumachen waren (Anhang: Abb. 46)

Die vergleichende Betrachtung der drei Auswertungen des Schlafes eines an obstruktiver schlafbezogener Atemstörung (OSAS) leidenden Patienten zeigte hingegen gleich mehrere Probleme auf, die generell bei der Bewertung des Stadium SWS auftreten: Hier wurde bei keinem der Paarvergleiche eine Übereinstimmung bezüglich Stadium SWS gefunden ( $\kappa=0,001; 0,06; 0,07$ ). Da im Alter die Amplitude der Delta-Aktivität im Stadium SWS abnimmt (Webb und Dreblow, 1982; Feinberg et al., 1984) (beim männlichen Geschlecht im Durchschnitt bereits zehn Jahre früher), war es hier für alle drei Auswertenden schwer, das Stadium SWS korrekt zu bewerten, da die durchschnittliche Amplitude der Delta-Aktivität gerade an der Grenze zu  $75 \mu\text{V}$  im zentralen, bzw.  $30 \mu\text{V}$  im fronto-polaren EEG lag (Anhang: Abb. 47). Erschwerend hinzu kamen hierbei noch die nächtlichen Apnoen, die eine generelle Reduktion des Stadium SWS bewirken, so dass voneinander abweichend gewertete SWS-Epochen bei der Berechnung des Kappa-Koeffizienten noch stärker ins Gewicht fielen. Die fast „vollständige“ Übereinstimmung zwischen QUISI 1 und QUISI 2 legt also einerseits nahe, dass die Bewertung des Stadium SWS anhand des fronto-polar abgeleiteten EEGs mit einer relativ hohen Interrater-Reliabilität durchaus praktikabel ist. Andererseits deutet die große Streubreite der Kappa-Werte beim Stadium SWS und die geringere Interrater-Reliabilität der Paarvergleiche SPG-Experte - QUISI 1 bzw. 2 darauf hin, dass das Amplitudenkriterium der SWS eine potentielle Fehlerquelle darstellt und daher bei der Anpassung der Regeln von Rechtschaffen und Kales (1968) auf das von der Stirn abgeleitete Signal eher von dem Amplituden-Kriterium Abstand genommen werden sollte. Dies steht im Einklang mit Konklusionen anderer Autoren (Webb, 1986). Alternativ wurde aufgrund des monomorphen Erscheinungsbildes der fronto-polaren Delta-Aktivität empfohlen, statt eines bestimmten prozentualen Anteil an Delta-Wellen pro Epoche eine zusammenhängende Strecke von Delta-Aktivität bestimmter Länge als Bewertungskriterium für das Stadium SWS zu bestimmen.

Anhand der Beispiele wird deutlich, dass diskrepante Beurteilungen insbesondere bei geringer Dichte von NREM2-typischen Graphoelementen und dem Auftreten von Epilepsie-typischen oder –verdächtigen Potentialen wahrscheinlicher werden.

#### 4.2.4 Stadium REM

Für das Stadium REM wurde in der vorliegenden Arbeit bei allen Paarvergleichen insgesamt die beste Übereinstimmung zwischen zwei Auswertenden erreicht ( $\kappa=0,84, 0,82, 0,86$ ). Dies steht im Einklang mit einer Reihe von Vorarbeiten, in denen vergleichbar hohe Interrater-Reliabilitäten, insbesondere bei Patienten mit somnologischen Erkrankungen beschrieben worden sind (Kunz et al. 2000, Danker-Hopfe et al., 2001; (Abb. 48)).



**Abbildung 48** – Gesamtübereinstimmung der Auswertungen auf einzelne Epochen bezogen für das Stadium REM

Als Gründe für die hohe Interrater-Reliabilität des Stadium REM wird von einigen Autoren (Kim et al., 1993) in der besseren Charakterisierung dieses Schlafabschnitts durch zusätzliche Informationen wie Abfall des EMG-Tonus und schnelle Augenbewegungen angesehen. Da im fronto-polaren EEG lediglich teilweise die Muskelaktivität des Musculus frontalis als das EEG überlagernde schnelle Aktivität abgebildet ist bzw. sich aufgrund der Elektrodenpositionierung nur die horizontalen Augenbewegungen erkennen lassen, wird das Stadium REM wegen der Ähnlichkeit des EEGs bisweilen falsch als NREM1 bewertet. So fand sich bei einer der ausgewerteten Nächte zwischen den Scorings von QUISI 1 und QUISI 2 lediglich eine „schwache“ Übereinstimmung bezogen auf das Stadium REM ( $\kappa=0,24$ ). Da der Kappa-Wert für REM beim Paarvergleich SPG-Experte - QUISI 1 hier mit  $\kappa=0,62$  im Bereich für „starke“ Übereinstimmung liegt, der des Paarvergleichs SPG-Experte - QUISI 2 jedoch mit  $\kappa=0,28$  nur auf eine „schwache“ Übereinstimmung hindeutet, scheint es sich dabei

um eine systematische Abweichung in der Bewertung durch den Auswerter QUI SI 2 zu handeln. Eine visuelle Analyse dieser Aufzeichnung einer 77-jährigen Patientin, die an PLMS bzw. fragmentarischem Myoklonus leidet, deutet darauf hin, dass die hier vorgenommene abweichende Bewertung auf die Fehlinterpretation bestimmter elektrophysiologischer Erscheinungen als Augenbewegungen zurückzuführen ist. So wurden hier von QUI SI 2 durch Myoklonien ausgelöste K-Komplexe (Anhang: Abb. 49) und Mikroarousals (Anhang: Abb. 50) als rapid eye movements fehlgedeutet und die betreffenden Epochen als REM gewertet, so dass insgesamt gegenüber der Auswertung der Somnopolygraphie ein um 5,7 % höherer Anteil an Stadium REM erreicht wurde.

Ähnliche Fehleinschätzungen ließen sich bei der Auswertung des Schlags einer 56-jährigen Patientin mit einem Syndrom der vorverlegten Schlafphase finden (siehe auch: 4.2.2). Hier erreichten die beiden Auswertenden des fronto-polaren EEGs zwar eine „deutliche“ Übereinstimmung, der Kappa-Wert für REM lag mit  $\kappa=0,47$  (SPG-Experte - QUI SI 1:  $\kappa=0,52$ ), SPG-Experte - QUI SI 2:  $\kappa=0,59$ ) jedoch ebenfalls im untersten Bereich aller Werte für das Stadium REM. Die korrekte Beurteilung des Schlags anhand des von der Stirn abgeleiteten Signals war hier einerseits durch Myoklonien-begleitende spitze Ausschläge (Anhang: Abb. 51), andererseits durch in der gesamten Aufzeichnung vorkommende Artefakte (Anhang: Abb. 52) erschwert.

Bei der Aufzeichnung des Schlags eines 41-jährigen Patienten, der an einer saisonal affektiven Störung litt, wurde ebenfalls lediglich eine „deutliche“ Übereinstimmung für das Stadium REM erreicht. Die Kappa-Werte lagen hier bei  $\kappa=0,58$  (QUI SI 1 - QUI SI 2),  $\kappa=0,59$  (SPG-Experte - QUI SI 1) bzw.  $\kappa=0,77$  (SPG-Experte - QUI SI 2) und deuten so auf eine abweichende Einschätzung des Schlags durch QUI SI 1. Eine visuelle Analyse dieser Aufzeichnung legt nahe, dass QUI SI 1 tatsächlich stattfindende *rapid eye movements* als Bewegungsartefakte verkannte bzw. auf Beschleunigungen der EEG-Aktivität folgende Epochen eher als Stadium NREM1 beurteilte und somit insgesamt 14,5 % weniger Stadium REM wertete als der SPG-Experte.

Ebenfalls prozentuale Abweichungen von > 14% (QUI SI 1) bzw. > 12% (QUI SI 2) ergaben sich bei der Klassifikation des REM-Schlafes einer 30-jährigen Patientin mit einer REM-Schlaf Verhaltensstörung (siehe auch 4.2.2). Der gestörte REM-Schlaf wurde als NREM 1 verkannt. Dies bezüglich sind zwei Aspekte zu berücksichtigen: Erstens ist zu bedenken, dass

Schlafauswertungen oft in Kenntnis der Verdachtsdiagnose durchgeführt werden, und daher die Aufmerksamkeit auf die entsprechende Symptomatik gelenkt werden kann. Um das gebotene Maß an Objektivität zu wahren, wurden dagegen die Auswertungen des fronto-polaren EEGs in Unkenntnis der Verdachtsdiagnose vorgenommen. Zweitens wirkt die Auswertung des Schlafes aufgrund der Komplexität der Störung auch auf Basis einer PSG Probleme auf, zwischen REM-Schlaf und Stadium NREM 1 zu differenzieren, so dass unterschiedliche Auswerter zu divergierenden Ergebnissen kommen, d.h. die fehlende Übereinstimmung ist nicht auf die Verfügbarkeit nur eines EEG-Kanals zurückzuführen.

#### 4.3 Fazit

Die vorliegende Arbeit zeigt, dass die Anwendung von den von Rechtschaffen und Kales (1968) aufgestellten Regeln zur Bewertung des menschlichen Schlafs in modifizierter Form auf das fronto-polar abgeleitete Schlaf-EEG anwendbar sind. Die Auswertenden, die mit Hilfe dieses „Fronto-polaren Scoring Systems“, einen einzelnen von der Stirn abgeleiteten EEG-Kanal visuell analysierten und so ein Schlafprofil des zu Untersuchenden erstellten, erzielten im Durchschnitt hohe Übereinstimmungen mit dem Auswertenden der Somnopolygraphie, dem wesentlich mehr Informationen zur Verfügung standen.

Zusammenfassend lässt sich feststellen, dass die Beurteilung des Schlafs mittels eines einzelnen von der Stirn abgeleiteten EEG-Kanals immer dann gut durchführbar ist, wenn das EEG des untersuchten Individuum bestimmte charakteristische Graphoelemente (Spindeln, K-Komplexe und Deltaaktivität mit hoher Amplitude) in hoher Anzahl und deutlicher Ausprägung enthält, was vor allem bei jüngeren Patienten der Fall ist. Außerdem eignet sich die Methode insbesondere dafür, die Einschlafzeit eines Individuums zu bestimmen, dessen Schlafverhalten zudem durch das QUISI®-Gerät weit weniger beeinflusst wird als durch eine somnopolygraphische Aufzeichnung.

Probleme bei der Beurteilung von Schlaf mittels des fronto-polaren Schlaf-EEGs können dann auftreten, wenn das untersuchte Individuum an einer Schlafstörung leidet, deren Ursache primär nicht im EEG abgebildet wird, wie das z.B. bei Patienten mit obstruktiven schlafbezogenen Atemstörungen oder periodischen Beinbewegungen im Schlaf der Fall ist. Außerdem gestaltet sich die Beurteilung eines fronto-polar abgeleiteten EEGs bei Personen besonders schwer, deren EEG viele Artefakte durch einen dauerhaft erhöhten Muskeltonus aufweist (Patienten mit Angststörungen oder atypischem Stadium REM).

Die bei einzelnen Nächten erzielte geringe Interrater-Reliabilität lässt sich im wesentlichen durch drei Faktoren erklären:

- Erstens lässt sich bei der visuellen Bewertung des menschlichen Schlafs durch voneinander unabhängige Experten aufgrund individueller Unterschiede zwischen den Auswertenden niemals eine einhundertprozentige Übereinstimmung erzielen.
- Zweitens wurden die von Rechtschaffen und Kales (1968) erstellten Regeln für junge gesunde Probanden aufgestellt und weisen so einige Unzulänglichkeiten bei der Beurteilung der EEGs älterer und schlafgestörter Personen auf (Kubicki et al., 1982; Himanen und Hasan, 2000).
- Drittens erwies sich die Auswertung des fronto-polaren Signals dadurch erschwert, dass in diesem Signal kaum Alphaaktivität zu erkennen ist, die zur korrekten Beurteilung des Wach-Schlaf-Übergangs nach Rechtschaffen und Kales notwendig ist, so dass die Auswertenden ihre Aufmerksamkeit verstärkt auf im QUISI-Signal abgebildete Augenbewegungen richten mussten, um die Vigilanz des zu Untersuchenden einschätzen zu können.

Hervorzuheben ist jedoch die Tatsache, dass die beiden Auswertenden des fronto-polaren Signals untereinander eine deutlich höhere Übereinstimmung als mit dem SPG-Experten erzielten, was für die Verwendbarkeit dieses Signals zur Schlafbeurteilung spricht. Verbesserungen, die Interrater-Reliabilität mehrerer Auswerter des QUISI-Signals betreffend, ließen sich in Zukunft noch erzielen, wenn man einige der von Himanen und Hasan (2000) verfassten Vorschläge zur Beurteilung der Schlaftiefe bei der visuellen Analyse des von der Stirn abgeleiteten Signals berücksichtigt: Insbesondere die Interrater-Reliabilität für die Stadien NREM1 und SWS lässt sich durch eine noch genauere Beschreibung der Augenbewegungen bzw. durch Verzicht auf das Amplituden-Kriterium zur Bewertung der Deltaaktivität noch verbessern.

Das „Fronto-polare Scoring System“ ist also durchaus geeignet, die von der QUISI-Software durchgeführte automatische Klassifikation der Schlafstadien auf ihre Plausibilität hin zu überprüfen.