Detection of K-Complexes based on the Wavelet Transform

Laerke K. Krohne[1], Rie B. Hansen[1], Julie A.E. Christensen[1,2], Helge B.D. Sorensen[1] and Poul Jennum[1,2]

***Abstract*— Sleep scoring needs computational assistance to reduce execution time and to assure high quality. In this pilot study a semi-automatic K-Complex detection algorithm was developed using wavelet transformation to identify pseudo-K-Complexes and various feature thresholds to reject false positives. The algorithm was trained and tested on sleep EEG from two databases to enhance its general applicability. When testing on data from subjects from the DREAMS© database, a mean true positive rate of 74 % and a positive predictive value of 65 % were achieved. After adjusting a few thresholds to adapt to the second database, the Danish Center for Sleep Medicine, a similar performance was achieved. The algorithm performs at the level of the State of the Art and surpasses the inter-rater agreement rate.**

I. INTRODUCTION

Analysis of biomedical signals for monitoring and diagnosis requires identification of both macro- and microstructural signal components. In sleep medicine several hours of polysomnographic (PSG) data are analyzed to perform sleep stage scoring and to track micro-sleep-events as a diagnostic aid for sleep disorders and certain diseases.

Visual analysis of all-night PSG electroencephalogram (EEG) is a time-consuming and non-consistent task. Furthermore, subjective results lead to a very low agreement rate between different sleep scorers [1]. This calls for advanced signal processing techniques to perform reliable and automatic segmentation and detection of micro-events.

The K-Complex (KC) is a micro-sleep-event that normally occurs during non-REM sleep stage 2 (N2) and assists the scoring of this stage [2]. According to the newest standard for scoring sleep, the American Standard for Sleep Scoring (AASM), a KC is a negative sharp wave immediately followed by a positive component standing out from the background EEG, with total duration $\geq$ 0.5 sec [2]. KCs are ideally observed clearly delineated, but in practice they are often difficult to distinguish from delta and vertex sharp waves [1].

Different KC detection algorithms have been proposed to enable functional and more reliable methods for KC detection [1][3-7]. They suggest different methods and approaches such as wavelets, thresholding techniques, and machine learning, but due to the wide diversity in EEG and KC appearances in between subjects, it is very challenging to obtain satisfactory performance and reliability.

In this pilot study a semi-automatic detection algorithm was developed (KCWavelet) with the aim to determine KC density in N2-sleep. The algorithm is a two-fold process as it uses wavelet transformation to identify possible KCs (pseudo-KCs) in N2-sleep, and various feature thresholds to reject the false-positive pseudo-KCs.

[1]Technical University of Denmark, Department of Electrical Engineering, Building 349, 2800 Kgs. Lyngby, Denmark

[2]Danish Center for Sleep Medicine, Glostrup University Hospital, Glostrup, Denmark

II. DATA

The algorithm is trained and tested on data from both the DREAMS© database [8] and data obtained from the Danish Center for Sleep Medicine at Glostrup Hospital [9] (DCSM database). All in all, the data contain a total of around three hours of sleep EEG from 8 healthy subjects.

A. The DREAMS© K-Complex database

The DREAMS© K-Complex database contains 30 minutes of data from ten subjects [8]. Five of these subjects were scored independently by two sleep experts (V1 and V2), and therefore the EEG CZ-A1 channel in N2-sleep from these subjects was used in this study. The data were extracted from all-night PSG recordings with a sampling frequency of 200 Hz. Manually scored hypnograms with sleep stage representations were used to extract N2-sleep. Table I summarizes demographics and sleep facts, and the number of KCs annotated by V1, V2 and both, respectively.

B. Data from the Danish Center for Sleep Medicine

To test the consistency of the approach, 30 minutes of EEG from three subjects collected at DCSM were included in this study. The data include all sleep stages, and not only N2-sleep. The PSGs were recorded and scored manually according to the AASM standard [2], and the channel C3-A1 was extracted with a sampling frequency of 256 Hz [10]. The KC annotations were made by the main and second author of this paper after training session on sleep scoring.

III. METHODS

This study took its starting point from the "Wavelet and Teager energy operator" algorithm presented by Erdamar et al. [1]. However, the final algorithm differs significantly from its source of inspiration. By optimizing the

TABLE I

DEMOGRAPHICS AND SLEEP FACTS.

	No. of subjects (♂,♀)	**Age** ($\mu \pm \sigma$)	**Total duration in N2**
DCSM	3 (3,0)	58.7 ± 12.9	1h 30 min[a]
DREAMS©	5 (4,1)	27.4 ± 11.1	1h 26 min
	No. of KCs pr. min (Visual annotation) ($\mu \pm \sigma$)		
	V1	**V2**	**V1∩V2**
	1.7 ± 0.70	0.41 ± 0.28	0.34 ± 0.19

a. All sleep stages, not only N2.

wavelet transformation step, the Teager Energy Operator used by Erdamar et al. [1] became unnecessary, and thus the detection efficiency of the algorithm was improved by omitting this step. Furthermore a "rejection step", including four new features, was added to improve the positive predictive value. The overall methodology of the KCWavelet algorithm is provided in Fig. 1.

The final KCWavelet algorithm thus consists of two main steps: the Extraction step and the Rejection step. In the Extraction step pseudo-KCs are identified by use of wavelet transformation with a KC-similar mother wavelet. In the Rejection step, pseudo-KCs are rejected, if they do not meet certain feature thresholds. The algorithm was trained and tested using cross-validation, and the best feature and threshold combination was found based on performance measures weighting both True Positive Rate (*TPR*) and Positive Predictive Value (*PPV*). The following sections provide more detailed explanations of the methodologies used.

A. Algorithm Structure

The KCWavelet algorithm is a semi-automatic KC detection algorithm implemented using MATLAB (R2013a, 64-bit, MathWorks, Natick, MA, USA). The required inputs are a single-channel EEG signal and a hypnogram. The hypnogram is scored by a sleep expert and used to extract N2-sleep. Future work includes automation of N2-sleep estimations to make the algorithm fully automatic. The output is a binary array where 1 indicates a KC and 0 indicates background activity for each sample of data.

B. Extraction Step

Wavelet transformation is a local time-frequency analysis that breaks a signal into shifted and scaled versions of a mother wavelet [11]. In wavelet terms, a function $f(t)$ can be described by

$$f(t) = \sum_{j,k} b_{jk} w_{jk}(t) \quad (1)$$

where $w_{jk}(t)$ are the wavelet basis functions constructed from the mother wavelet $w(t)$, b_{jk} are the coefficients, and j and k denotes the translation and scaling factors, respectively [11]. In this study the discrete wavelet transform (DWT) was used to decompose the EEG signal into multi-resolution subsets of coefficients. The aim of this step is to approximate a signal only containing the KCs, and therefore the approximation coefficient is used for reconstruction. Various combinations of mother wavelets (Daubechies 4 and 5, and Symlets 4 and 7) and levels of decomposition (levels 4 and 5) were investigated. The best combination was found to be the Daubechies 4 mother wavelet at a 5 level decomposition (approximation frequency sub-band 0-3.12 Hz).

The Extraction step finds all possible pseudo-KCs in the N2 epochs, and thereby aims for a high *TPR*. Initially, KC-similar waveforms are enhanced by means of the wavelet transformation, and then the duration, amplitude, and slope of the waveforms are examined in comparison to certain thresholds. The feature, *Slope*, was implemented as one of the main features of the KC, and its threshold is determined from the training process of the algorithm. All pseudo-KCs found in this step are stored in a logical KC array, *KW_Total*, which is given as input for the Rejection step.

C. Rejection Step

The Rejection step tests all pseudo-KCs identified in the Extraction step in regard to specific control-features. If a pseudo-KC obeys certain thresholds of all the control-features, it is accepted as a true KC, if not, it is rejected as background activity. This step thus aims for a high *PPV*. The control features ensure a background amplitude check, a peak amplitude check and a relative power check, respectively:

1) *Background_Amp*: Tests if the peak-to-peak amplitude (*pp amplitude*) of the pseudo-KC is minimum twice as large (as suggested in [6]) as that of its background amplitude. The background amplitude is calculated in two ranges (*Background_1* and *Background_2*) to ensure that two consecutive KCs can still be detected. Furthermore a threshold *MeanTHR* removes the influence of small fluctuations at baseline-level.
2) *Range*: Defines the KC background by determining the duration (in seconds) of the two ranges *Background_1* and *Background_2*.
3) *PeakAmp*: Ensures that KCs with too low positive amplitude are rejected, while ensuring that the negative peak has at least half the amplitude of the positive peak.
4) *RelDelta*: Rejects pseudo-KCs if the relative power in the surrounding signal is mostly delta activity (0.75-4 Hz), and thereby minimizes the number of false-positives originating from delta waves.

Fig. 2 is an illustration of some parameters of the features described above. The implementation of the features and their thresholds can be seen in the pseudo-algorithm in Fig. 3.

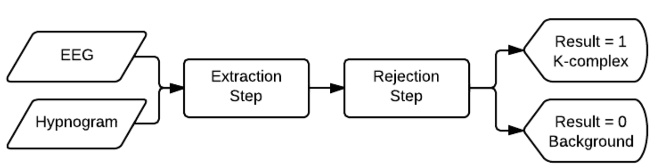

Fig. 1. Overall structure of the KCWavelet algorithm.

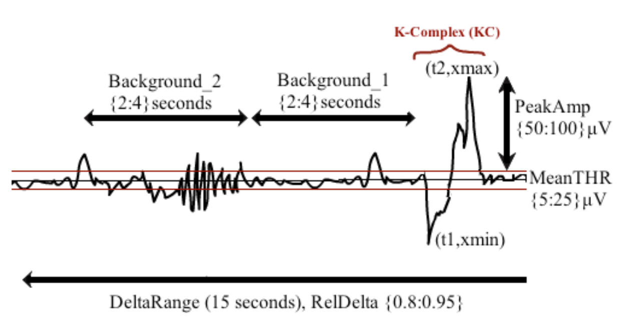

Fig. 2. K-Complex and features used in Rejection step.

Input: 1) *KW_Total* from the Extraction step, 2) the raw EEG data, and 3) feature thresholds
Output: A logical KC array *K_Total*

Set duration of *Background_1* and *Background_2* to *Range*
foreach *pseudo-KC* **do**
    if *pp amplitude* $<2\cdot$ *pp amplitude of Background_1* **and** *pp amplitude* $<2\cdot$ *pp amplitude of Background_2* **then**
        reject;
    end
    if *xmax* $<$ *PeakAmp* **or** *xmin* $<$ *PeakAmp*$/2$ **then**
        reject;
    end
    if *relative power of delta band in the DeltaRange* $>$ *RelDelta* **then**
        reject;
    end
end
K_Total $\leftarrow$ KC array with ones at samples corresponding to non-rejected KC instances and zeros else;

Fig. 3. Pseudo-Algorithm for the Rejection step.

D. Optimization and Cross-validation

To find the best feature combination, all combinations were investigated for the DREAMS© dataset in given feature ranges indicated in Fig. 2. To reduce the risk of overfitting a leave-one-subject-out cross-validation procedure was used. Because the duration of N2-sleep in each subject was not fixed, the amount of data held out in each fold differed.

The overall estimated performance $\hat{P}$ for the model trained using all the data was given as the average performance across all the K folds [12]:

$$\hat{P}=\frac{1}{K}\sum_{i=1}^{K}\hat{P}(i) \tag{2}$$

The model was optimized using the Matthews Correlation Coefficient (*MCC*) as the performance measure. The *MCC* was chosen, as it is a single value performance measure including all four confusion matrix instances (True Positive (*TP*), True Negative (*TN*), False Positive (*FP*), and False Negative (*FN*)). Furthermore, *MCC* was chosen because it is regarded as a balanced measure, which can be used even if classes are of unequal sizes. The *MCC* is expressed as,

$$MCC=\frac{TP\cdot TN-FP\cdot FN}{(TP+FP)\,(TP+FN)\,(TN+FP)\,(TN+FN)} \tag{3}$$

and returns a value between -1 (worst) and 1 (best), where 0 indicates a result no better than a random prediction [12].

This study also considered a case where the True Positive Rate ($TPR=TP/(TP{+}FN)$) is weighted slightly higher than the Positive Predictive Value ($PPV=TP/(TP+FP)$), using the F_2 measure to optimize the model. The F_2 is given as,

$$F_2=(1+2^2)\cdot\frac{PPV\cdot TPR}{(2^2\cdot PPV)+TPR} \tag{4}$$

and returns a value between 1 (best) and 0 (worst).

IV. RESULTS AND DISCUSSION

The KCWavelet method proposed in this study was applied on data from two databases, DREAMS© and DCSM. Furthermore, the inter-rater agreement in DREAMS© was evaluated to compare the performance of the algorithm with the agreement rate between different sleep scorers.

A. Inter-rater agreement

Visual recognition of KCs presents large viability in terms of definition and classification [1][3-7]. The annotations made by the visual scorers V1 and V2 in DREAMS©, showed to have an agreement rate of only 33 % [6]. Similarly, the three scorers of the data presented by Erdamar et al. had an agreement rate of 59 % [1]. These results show that it is very difficult to develop and evaluate a KC detection algorithm based on the Gold Standard.

V1 was chosen as the Gold Standard in this study after clinical judgment made by a sleep-scoring expert from DCSM and an evaluation of the KC density estimated by each scorer (the mean number of KCs per minute during N2-sleep is $1-3$ [6], which is fulfilled by V1, see Table I).

B. KCWavelet (data from DREAMS© and DCSM)

Fig. 4 shows an example of detection from the KCWavelet algorithm applied on data from DREAMS©. The top graph shows the raw EEG together with green markings indicating KC annotations made by V1. The middle graph shows pseudo-KCs found in the Extraction step, while the bottom graph shows the final output. In this case the Rejection step sufficiently rejects the two *FP*s found in the first step.

Table II shows the performance values achieved by training and testing on the five subjects from DREAMS© with respect to *MCC* and F_2 separately. The training yielded a mean of 0.618 for *MCC*, and a mean of 0.716 for F_2. The performance values presented are the *TPR*s and *PPV*s for the subjects individually together with the overall mean.

The combination of feature thresholds used to achieve the results in Table II was: *Slope* $=250$, *Range* $=4$, *MeanTHR* $=25$, *PeakAmp* $=87.5$ and *RelDelta* $=0.95$.

It can be argued that a high *TPR* is more important than a high *PPV* for an unbalanced detection scenario, which is why F_2 also was used for optimization in the training. If an algorithm provides a high *TPR* but low *PPV*, sleep experts

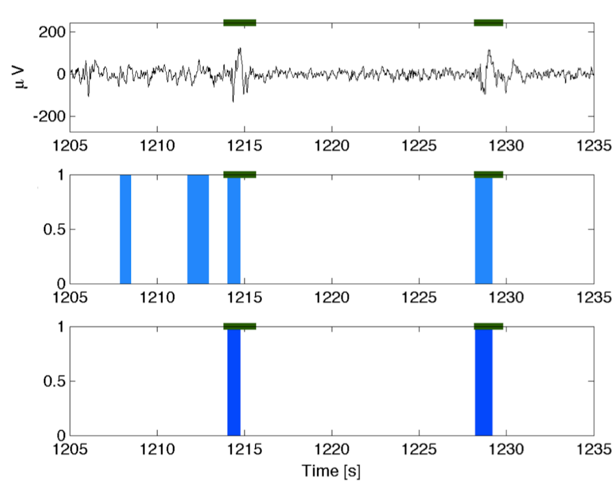

Fig. 4. KCWavelet detection example applied on data from DREAMS©.

TABLE II

DREAMS© PERFORMANCE EVALUATION.

	No. KCs (V1)	KCWavelet (*MCC*)		KCWavelet (F_2)	
		TPR	***PPV***	***TPR***	***PPV***
Subject 1	34	58.8 %	62.5 %	82.4 %	50.0 %
Subject 2	31	41.9 %	72.2 %	67.7 %	67.7 %
Subject 3	12	25.0 %	100 %	50.0 %	85.7 %
Subject 4	45	71.1 %	64.0 %	75.6 %	65.4 %
Subject 5	38	83.8 %	66.0 %	91.9 %	56.7 %
Mean	**32**	**56.1 %**	**72.9 %**	**73.5 %**	**65.0 %**

have to double-check all positives to eliminate the *FP*s. If it, on the other hand, provides a low *TPR*, the sleep experts would have to look through the entire dataset to find the *FN*s. How to balance the *TPR*/*PPV* is, however, completely dependent on what the detector should be used for, but high values of both *PPV* and *TPR* are always desired.

When evaluating the performance of this algorithm it is important to be aware of the on-going discussion of sleep experts' tendency to score too few KCs [3]. Thus, if the algorithm is trained to aim at a high *PPV*, one might risk that the algorithm discards KCs that are actually true but not marked by the scorer – and thus one prevents the possibility of the algorithm to outperform the human eye.

To evaluate the general applicability of the proposed procedure, the algorithm was also trained and tested on data from DCSM. Here a rough training procedure and visual inspection yielded an optimal combination of feature thresholds at *Slope* $= 200$, *Range* $= 4$, *MeanTHR* $= 32$ and *RelDel* $= 0.7$. In Table III is provided the performance measures obtained when using the KCWavelet algorithm with the new combination of parameters. As seen from the performance measures, the mean *TPR* reached an acceptable value, whereas the *PPV* is somewhat low.

A visual false-detection analysis showed that some of the detection errors were due to steep fluctuations coursing inconvenient baseline crossings for the algorithm, while not changing the overall KC morphology. But by far, most errors were due to great inter-subject differences seen in the EEG. This weak point of the algorithm is also seen clearly by comparing the performance level for Subject 1 and 3 in Table II. From further investigation of the EEG from Subject 3, it was seen that the morphology of the KCs differed significantly from those from the other subjects, which might be due to the subject's age, 47, 18 years older than the mean. Many studies have showed that EEG frequency characteristics and signal strength change significantly with age [6]. This is a clear indicator of the necessity of implementing better relative features for a better general applicability.

TABLE III

DCSM PERFORMANCE EVALUATION.

	No. KCs	KCWavelet	
		TPR	***PPV***
Subject 1	64	93.6 %	28.6 %
Subject 2	40	52.5 %	80.8 %
Subject 3	35	71.4 %	44.6 %
Mean	**46**	**72.6 %**	**51.3 %**

When comparing the feature thresholds' efficiency, the features *PeakAmp* and *MeanTHR* showed to be most effective in the rejection of *FP*s. The *RelDelta* did not have as large an influence as expected in data from DREAMS©, however, this might be because the data only contained N2-sleep. When the algorithm was tested on data from DSCM containing all sleep stages, this feature threshold was much more influential.

The data from DCSM and DREAMS© differ in terms of data collection, the EEG channel used, and the age of the subjects, which induced the necessity of manually adjusting the feature thresholds for the DCSM database. Still the performances achieved on these completely different datasets are indeed comparable. This indicates a good general applicability of the procedure.

V. CONCLUSION

The KCWavelet algorithm yielded a mean *TPR* of 74 % and a mean *PPV* of 65 % (see Table II), which is comparable to those presented in other studies [1][3-7], and significantly higher than the agreement rate between visual scorers.

This study needs first and foremost to be seen as a precursor for the further development of an automatic KC detection algorithm with high reliability and general applicability. In specific, future research should focus on automatic setting of subject-specific feature thresholds. Furthermore, future studies should aim for larger datasets, while persisting critical approach towards the Gold Standard.

REFERENCES

[1] A. Erdamar et al., A wavelet and teager energy operator based method for automatic detection of K-complex in sleep EEG, Elsevier Expert Systems with Applications, vol. 39, pp. 1284-1290, 2012.

[2] R. B. Berry et al., The AASM Manual for the Scoring of Sleep and Associated Events Version 2.0. American Academy of Sleep Medicine, 2012.

[3] Q. V. Huy et al., K-complex detection using a hybrid-synergic machine learning method, IEEE transactions on the systems, man, and cybernetics, vol. 42(6), pp. 1478-1490, 2012.

[4] A. Kam et al., Detection of K-complexes in sleep EEG using CD-HMM, 26th Annual Int. Conf. of the IEEE EMBS, pp. 33-36, 2004.

[5] M. Gala et al., Detection of K-complex in the EEG signal, IFMBE Proceedings, vol. 25(4), pp. 1170-1173, 2009.

[6] S. Devuyst et al., Automatic K-complexes detection in sleep EEG recordings using likelihood thresholds, IEEE Engineering in medicine and biology society conference proceedings, pp. 4658-4661, 2010.

[7] I. N. Bankman et al., Detection of the EEG K-complex wave with neural networks, IEEE transactions on biomedical engineering, vol. 39(12), pp. 1305-1310, 1992.

[8] S. Devuyst. The DREAMS© K-complexes Database. URL: http://www.tcts.fpms.ac.be/∼devuyst/Databases/DatabaseKcomplexes, July 2012.

[9] Glostrup Hospital. About the Danish Center for Sleep Medicine. URL: http://www.glostruphospital.dk/english/menu/Departments/Danish +Centre+for+Sleep+Medicine/, May 2013

[10] J. A. E. Christensen et al., Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease, Clin Neurophysiol., vol. 125(3), pp. 512-519, 2014.

[11] G. Strang et al. Wavelets and Filter Banks. Wellesley College; 2nd edition, 1996.

[12] P. Baldi et al., Assessing the accuracy of prediction algorithms for classification: an overview, Bioinformatics Review, vol. 16(5), pp. 412-424, 2000.