

Sleep and Arousal Scoring in HSAT: DeepResp with Nox BodySleep (v.2)



This document provides practical guidance on how to interpret and review sleep staging and arousal detection generated by Nox Medical's DeepResp™, powered by Nox BodySleep™ (v.2). DeepResp is an automated scoring solution that includes sleep staging, arousal scoring, and respiratory event scoring within home sleep apnea testing (HSAT).

It focuses on:

- » How automated sleep staging and arousal detection are presented in Noxturnal Web
- » How clinicians can review and validate these outputs
- » How these outputs can be incorporated into routine scoring workflows

Background: HSAT Scoring

The American Academy of Sleep Medicine (AASM) Scoring Manual defines cortical arousals in polysomnography (PSG) using EEG.

In home sleep apnea testing (HSAT), where EEG is not available, scoring has traditionally relied on respiratory signals and oxygen desaturation. As a result, HSAT has not included sleep staging or arousal scoring, which can contribute to underestimation of AHI.

Advances in signal analysis and machine learning now enable detection of autonomic arousal patterns and sleep states from respiratory inductance plethysmography (RIP) signals in HSAT systems, tuned to align with EEG-defined arousals – enabling more complete interpretation of HSAT recordings.

Interpreting Sleep and Arousal Signals in HSAT

Nox BodySleep (v.2), part of DeepResp, uses Nox RIP signals to identify:

- » Ventilatory patterns across sleep states (wake, NREM, REM)
- » Arousal-related respiratory responses from breathing patterns

By identifying sleep states and arousals using RIP signals, DeepResp enables scoring of hypopneas that terminate in arousal without oxygen desaturation — events that are often missed in traditional HSAT and in many cases represent a significant component of sleep-disordered breathing. This approach was validated against expert-scored PSG recordings with EEG¹.

The clinical relevance of capturing these events is reflected in current sleep scoring guidance. The AASM has emphasized the importance of identifying hypopneas associated with either oxygen desaturation or arousal, recognizing that reliance on desaturation-only definitions may miss symptomatic patients who could benefit from diagnosis and treatment. While this work has been developed in the context of PSG, the AASM Hypopnea Scoring Rule Task Force further highlighted that inclusion of arousal-based hypopneas affects OSA prevalence, treatment eligibility, and identification of patients who may otherwise be undiagnosed, especially in groups such as women and younger adults².

According to AASM HSAT Hypopnea Rule 2A, when sleep is recorded (with EEG), a respiratory event may be scored as a hypopnea when:

- a. Peak signal excursion drops $\geq 30\%$ from pre-event baseline,
- b. Duration of reduction is ≥ 10 seconds, and
- c. Event is associated with $\geq 3\%$ oxygen desaturation or an arousal

In standard PSG, the arousal criterion in (c) is assessed using EEG. As there is no sleep staging and arousal scoring in traditional HSAT, AASM does not recommend scoring arousal-associated hypopneas in HSAT studies. For most systems, this holds true and they can only apply the desaturation arm of this rule, meaning hypopneas that terminate in arousal without significant desaturation go undetected. These arousal-associated hypopneas are clinically relevant and, in many patients, represent a meaningful portion of the total AHI.

DeepResp addresses this gap by using Nox RIP signals to infer sleep states and detect arousals from characteristic breathing pattern changes, without EEG. This enables identification of arousal-associated hypopneas in the HSAT setting, applying the full Recommended hypopnea definition rather than only the desaturation option. DeepResp's HSAT arousal detection was validated against expert-scored PSG recordings with EEG¹.

Clinicians retain full control of scoring and may review confidence levels and adjust events prior to diagnosis. The following section outlines how these outputs are presented and reviewed in practice.

DeepResp's Autoscoring in Noxturnal Web

DeepResp's automatic analysis is fully integrated into Noxturnal Web through Nox Connect. For HSAT, Nox BodySleep (v.2), part of DeepResp, provides automated:

- » Sleep stage (REM, NREM, wake) scoring
- » Arousal scoring
- » Respiratory event scoring

These outputs are presented alongside standard respiratory signals and can be reviewed within the same workflow clinicians use today.

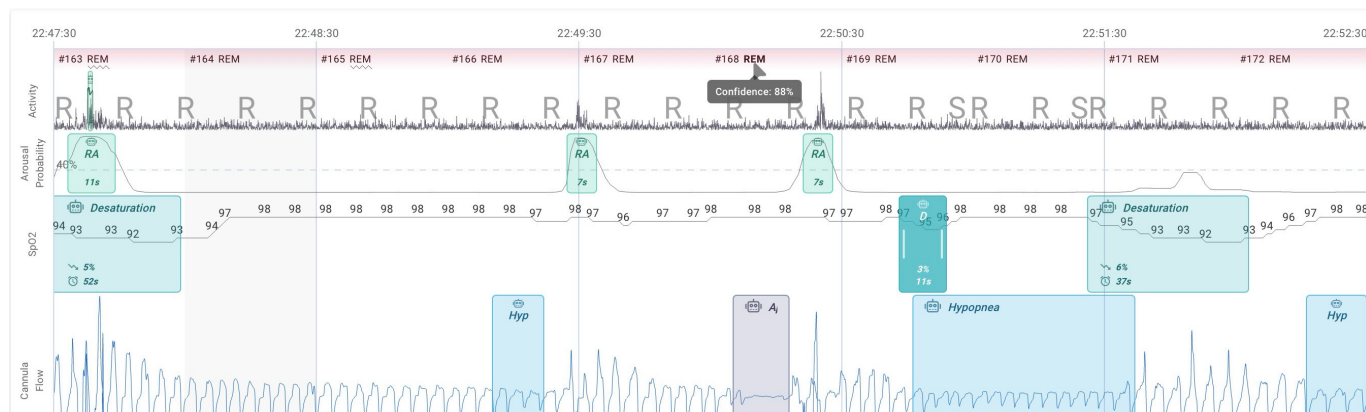
The system is fully transparent. Clinicians can:

- » Review underlying signals logic
- » Manually adjust any scored events as needed
- » Review the probability or likelihood of each scored sleep stage, arousal, or respiratory event

Sleep Staging (Epoch-by-Epoch)

1. Probability Assignment: The neural network analyzes the input signals and assigns probabilities to each potential sleep state: Wake, NREM, and REM.
2. Classification: The final sleep state for the 30-second epoch is initially determined by selecting the state with the highest probability calculated by the neural network.

Example: If Nox BodySleep calculates REM = 88% and other states are lower, the epoch is scored as REM.



Arousal Detection (Second-by-Second)

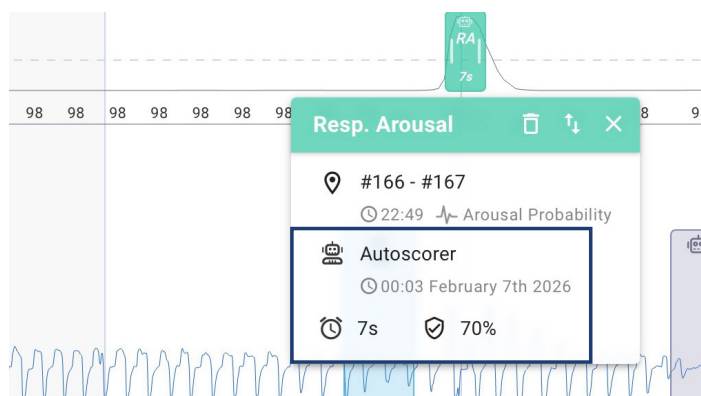
Arousals are tracked on a dedicated “Arousal Lane” with 1-second resolution. Clinicians have access to an Arousal Probability signal, which shows the strength of the detected arousal pattern over time.

An arousal event is scored when:

- » The probability crosses a defined threshold
- » For at least 3 consecutive seconds (AASM's minimum required duration of an arousal)

Example: Nox BodySleep scores arousal because the signal crossed the threshold for 7 seconds with a confidence level of 70%.

Even without EEG, clinicians can validate this by observing the recovery breaths in the RIP belts immediately following the hypopnea, a physiological hallmark of an arousal overcoming airway obstruction and restoring airway patency.



Respiratory Event Scoring

DeepResp scores AASM-defined respiratory events. Hypopneas are identified using both qualification pathways under the AASM recommended rule: the $\geq 3\%$ oxygen desaturation arm (when SpO2 is available) and the arousal arm.

Apneas are classified as obstructive or central. Central apnea detection reaches 99% overall agreement, supporting accurate derivation of both AHI and CAHI.

Combined with DeepResp's validated sleep staging (89–95% overall agreement for Wake, NREM, and REM)³, the result is an AHI built from reliable inputs on both sides of the equation: accurately scored respiratory events in the numerator and a total sleep time (TST) that closely matches the manually scored PSG reference in the denominator.

Example Interpretation

Point 1 (Increased Sensitivity):

This is an arousal-based hypopnea. There is significant flow reduction, but no oxygen desaturation. A standard HSAT would miss this event entirely. However, Nox BodySleep detects the arousal, allowing the hypopnea event to be correctly scored and contribute to a more accurate AHI.

Point 2 (High Specificity):

Here, we see a hypopnea followed by an oxygen desaturation (~4%). Note that the Arousal Probability signal remains low in this example. This demonstrates the analysis specificity; it does not over-score arousals where none exist, even when a desaturation event is present.

Point 3 (Dual Validation):

This event represents the intersection of criteria: a hypopnea accompanied by both a 3% SpO2 desaturation and a detectable arousal. DeepResp captures both, confirming the hypopnea event with high confidence.



Recommended Review Workflow

Review:

- » Arousal probability signal
- » RIP belt signals

Focus on:

- » Borderline events
- » Lower-confidence segments
- » Adjust scoring where clinically appropriate

All results must be reviewed by a qualified technologist and physician prior to diagnosis.

Validation and Regulatory Status

DeepResp SaMD is:

- » FDA-cleared (K241960, K252330), Health Canada licensed (number 113803), CE-marked (MDR)

Validation includes:

- » ~3,500 PSG studies (v1)
- » ~5,800 PSG studies, thereof ~4000 Type I PSG recordings, and ~1700 Type III recordings (with EEG and EOG channels added) (v2)

Arousal detection performance falls within reported human inter-scorer variability. DeepResp v.2's HSAT arousal scoring achieves 64% positive percentage agreement, 90.5% negative percentage agreement, and 83.1% overall percentage agreement — consistent with DeepResp v.1 and comparable to expert manual scoring. While the positive agreement may appear modest, this reflects a well-established limitation of arousal scoring itself, where even trained sleep technologists show moderate agreement.

Published multi-scorer studies demonstrate this variability, with reported agreement ranges of ICC ~0.41–0.68^{3,4,5} and κ ~0.40–0.59^{4,6,5}. These findings highlight that there is no single “ground truth” for arousal scoring, but rather a range of acceptable expert interpretations.

DeepResp's performance falls within these ranges and is also consistent with FDA-cleared, EEG-based systems, supporting clinically reliable results within the inherent limits of human scoring.

References

1. Finnsson E et al. Detecting arousals and sleep from respiratory inductance plethysmography. *Sleep Breath*. 2025 Apr 11;29(2):155.
2. Berry RB, Abreu AR, Krishnan V, Quan SF, Strollo PJ Jr, Malhotra RK. A transition to the American Academy of Sleep Medicine–recommended hypopnea definition in adults: initiatives of the Hypopnea Scoring Rule Task Force. *J Clin Sleep Med*. 2022;18(5):1419–1425.
3. DeepResp v.2 FDA 510k https://www.accessdata.fda.gov/cdrh_docs/pdf25/K252330.pdf
4. Pitkänen H, Nikkonen S, Rissanen M, Islind AS, Gretarsdottir H, Arnardottir ES, Leppänen T, Korkalainen H. Multi-centre arousal scoring agreement in the Sleep Revolution. *J Sleep Res*. 2024 Aug;33(4):e14127. doi: 10.1111/jsr.14127. Epub 2023 Dec 26. PMID: 38148632.
5. Ulysses J. Magalang, Ning-Hung Chen, Peter A. Cistulli, Annette C. Fedson, Thorarinn Gíslason, David Hillman, Thomas Penzel, Renaud Tamisier, Sergio Tufik, Gary Phillips, Allan I. Pack, for the SAGIC Investigators, Agreement in the Scoring of Respiratory Events and Sleep Among International Sleep Centers, *Sleep*, Volume 36, Issue 4, 1 April 2013, Pages 591–596, <https://doi.org/10.5665/sleep.2552>
6. Alvarez-Estevez D, Fernández-Varela I. Large-scale validation of an automatic EEG arousal detection algorithm using different heterogeneous databases. *Sleep Med*. 2019 May;57:6–14. doi: 10.1016/j.sleep.2019.01.025. Epub 2019 Jan 30. PMID: 30878899.