How does human hearing estimates sleepiness from speech?

Vincent P Martin, Salin Nathan, Beaumard Colleen, Jean-Luc Rouas

Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

How does human hearing estimates sleepiness from speech?

Vincent P. Martin [1], *Nathan Salin* [2], *Colleen Beaumard* [2,3], *Jean-Luc Rouas* [2]

[1] Deep Digital Phenotyping unit, Department of Precision Health, Luxembourg Institute of Health, L-1445 Strassen, Luxembourg
[2]Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France
[3]Univ. Bordeaux, CNRS, SANPSY, UMR 6033, F-33000 Bordeaux, France
vincentp.martin@lih.lu, jean-luc.rouas@labri.fr

Abstract

Excessive sleepiness is a major public and personal health burden that would benefit from being measured in ecological and passive setups. Speech recording is implemented in all smartphones and is thus a relevant tool to do so. To evaluate the feasibility of detecting sleepiness from speech by the human perception, two previous perceptual studies on 90 samples from the SLEEP corpus have been conducted (Huckvale et al. 2020, Martin et al. 2023), which yielded contrasting results. A way to investigate the origin of this disagreement would have been to study on which speech characteristics the listeners have based their estimation. However, none of these studies have collected such information. In this study, we identify these characteristics by extracting speech features from the recordings, and training simple and explainable machine learning models to reproduce the annotation of each listener. Then, we measure the contribution of each feature to the decision of each model, and identify the most important ones. We then perform hierarchical clustering to draw listeners' profiles, depending on the features they rely on to identify sleepiness.

Index Terms: perceptual study, sleepiness, computational paralinguistics

1. Introduction

1.1. Context

Excessive sleepiness is both a major public health burden [1, 2] and a serious personal health indicator linked with metabolic, cardiovascular, neurological, and psychiatric disorders, increasing the risk of disability and mortality [3, 4]. Because of its high prevalence in the general population (up to one person over three [5]), clinicians need tools to measure the sleepiness level of their patients as regularly as possible, in ecological conditions (e.g. at home), in a passive way (i.e. without a dedicated task). In this regard, voice and speech recordings are a candidate of choice: their collection is implemented in all smartphones, they can be recorded in passive setups, and they have already been linked to multiple disorders [6], including sleepiness.

Indeed, sleepiness detection using voice recordings has already been the focus of two Interspeech challenges in 2011 and 2019, respectively relying on the Sleep Language Corpus (SLC) [7] and the SLEEP corpus [8]. Both corpus are labeled with a subjective measurement (self-evaluating questionnaire) [9], the Karolinska Sleepiness Scale (KSS) [10]. The best system on the Interspeech 2011 challenge reached an Unweighted Average Recall (UAR) of 71.7% [11] on the binary classification of sleepiness. On the SLEEP corpus, the task of the Interspeech 2019 challenge was estimating the degree of sleepiness. The winner of the challenge reached a Spearman correlation of $\rho = 0.387$ between estimation and ground truth [12]. This simple approach has never been outperformed despite the use of cutting-edge deep learning techniques (e.g. $\rho = 0.325$ in [13], $\rho = 0.367$ in [14], $\rho = 0.365$ in [15] or $\rho = 0.383$ in [16]). More recently, the Voiceome dataset, a new large corpus recorded in ecological conditions using smartphones has been introduced [17]. The team working on it reported an F1-score of 81.3% on the binary classification of sleepiness, as measured by the Stanford Sleepiness Scale [18]. In parallel with this work focusing on short-term sleepiness, it is noteworthy to mention the work from a team of the sleep clinic of Bordeaux (France), who has recorded the Multiple Sleep Latency Test corpus (MSLTc), containing voice recordings of hypersomniac patients labeled with both short- and long-term subjective (questionnaires) and physiological (sleep latency measured by electroencephalography) sleepiness. Using this corpus, they reached a UAR of 73.2% in differentiating patients affected by excessive sleep propensity [19].

Yet, most of the research using these corpora over the past decade has focused on the development of machine learning algorithms to estimate sleepiness from the speech recordings contained in these corpora. By contrast, very limited attention has been paid to elucidating the link between sleepiness and speech behavior. Since the seminal work of Krajewski et al. in 2009 [20], very few studies have sought to clarify the mechanisms underlying the expression of sleepiness in speech. In parallel with this machine learning work, two recent perceptual studies were carried out on the SLEEP corpus to determine whether the human ear can estimate sleepiness from voice samples. These two studies, based on 99 samples of the SLEEP corpus, produced contradictory results: the study by Huckvale et al., involving 26 annotators, concluded that it was feasible to recognize somnolence in the corpus recordings [21]. By contrast, the study of Martin et al., based on the annotations by 30 naive listeners, concluded that the task was not feasible [22]. Since the listeners of Huckvale were English native speakers and those of Martin et al. spoke French, this divergence between the studies could have been explained by differences in the speech characteristics used by the listeners to estimate sleepiness, but none of these studies collected such feedback.

1.2. Objectives

The goal of this paper is to investigate how sleepiness is expressed through voice by re-analyzing the data of the two previous perceptual studies on the SLEEP corpus [21, 22] to determine, a posteriori, the speech characteristics used by the listeners to estimate sleepiness. To do so, based on a minimal feature set extracted from the audio recordings, we trained several

10.21437/SpeechProsody.2024-33

machine learning pipelines to reproduce the annotations of the listeners (one "cloning" machine learning system per listener). Both the features and the machine learning pipeline have been chosen simple and perfectly explainable, allowing the extraction and the interpretation of the relative importance of each feature in the imitation of the listeners' annotations. Finally, based on these features, we draw listeners' profiles, based on the way they identify sleepiness in voice recordings.

2. Method

An overview of our Method is represented in Figure 1.

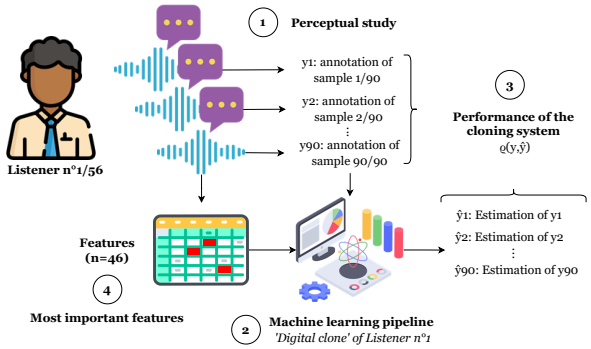

Figure 1: *Overview of our method to estimate the features used by the listeners of the perceptual studies to estimate sleepiness from speech samples*

checkpoint

2.1. Corpus and audio samples

We focus in this paper on the two perceptual studies involving the SLEEP corpus [21, 22]. The entire corpus contains more than 16,464 samples from 915 German-speaking subjects, recorded on different but unknown tasks [9]. All the samples are shorter than five seconds, with an average duration of 3.87 seconds. These samples have been annotated using the Karolinska Sleepiness Scale (KSS)[10], a questionnaire measuring subjective instantaneous sleepiness [23] using a 9-points Lickert-like scale. The two perceptual studies have used the same subset of 99 samples of the SLEEP corpus, 9 to train the listeners (one for each sleepiness level), and 90 (ten for each sleepiness level) for the experiment itself. Our analysis focuses on the 90 samples used for the experiment.

2.2. Perceptual study and listeners

During the two perceptual studies, the listeners were asked to estimate the sleepiness of the speaker from the audio recordings using a 9-point KSS. The samples were in the same order for the two studies, and listeners could not browse back. The two studies had different conclusions: while the annotations of the Huckvale et al. study [21], after applying a Wisdom of the Crowd algorithm, gave very convincing performances ($\rho = 0.72$ between estimation and the ground truth labels), the listeners of the replication study of Martin et al. [22] did not have the same success ($\rho = 0.41$).

Moreover, the study of Martin et al. was the only one to collect information about the characteristics of each listener. These included their genre (13F/17M), musical sensitivity (n=14 had music-related hobbies or profession; n=16 did not have), and their understanding of the German language ("at least a little", n= 11; "not at all", n=19). The other characteristics of each study are described in detail in another paper [22].

2.3. Voice features

Since the selected subcorpus from the SLEEP corpus has few samples per listener (90), and to allow the interpretation of the identified listeners' profiles, we limited ourselves to 46 features. They include the average and standard deviation of low-level features (n=40) and the temporal features (n=6) from the GEMAPS feature set, extracted using the Opensmile toolbox [24].

2.4. Machine learning pipeline

To obtain insights from the coefficient of the different parts of the pipeline, we chose simple algorithms that have previously shown efficient for regression estimation from small corpora:

(a) Lasso ($\alpha = 0.1$).

(b) PCA (80% of variance) + linear regression

(c) PCA (80% of variance) + Support Vector Regressor ($C = 1$)

A different classifier was trained for each annotator (n=26 for Huckvale et al. 2020, n=30 for Martin et al. 2023). Moreover, to compare the classifiers' performances with the state of the art in automatic sleepiness detection from voice, we also trained a classifier to reproduce the labels of the IS2019 challenge. Thus, a total of 171 classifiers (57 annotations set $\times$ 3 classifiers) have been trained.

2.5. Cross-validation and performance metric

In order to avoid overlearning, the performances were computed within a 5-fold cross-validation procedure, repeated 10 times. Since the sample size was small, we aggregated the estimation and corresponding ground truth and computed the performances on the aggregated labels. Since we did not finetune hyperparameters, we only performed simple cross-validation to evaluate the performances of the models. In the same way as in the IS2019 Challenge, the chosen performance metric was Spearman's ρ between estimated and ground-truth labels. The higher the value of ρ, the better the estimator. Both the labels and the input features were normalized (z-score).

2.6. Contribution of each feature

For each annotator, we measured the contribution of each feature in the pipeline trained to imitate him/her. For the pipeline using only Lasso for classification (a), we considered the L1-normalized weights of the classifiers. For the other pipelines [PCA and linear regression (b) or PCA and SVR (c)], we computed the L1-normalized cross-product of the PCA coefficient and classifier coefficients. In doing so, we measure the contribution of each feature to a given dimension of the PCA, which is weighted by the contribution of this PCA dimension to the classification. For each of these coefficients, we interpreted separately the absolute value – which is linked to the relative contribution of the feature to the classification – and the sign – which indicates the direction of the link between sleepiness and the voice feature.

Table 1: *Performance of the machine-learning pipeline trained to imitate listeners of the perceptual studies depending on the chosen model. Values are computed on the aggregation from a 5-fold cross-validation repeated ten times, and represented as Mean ± standard-deviation [min-max]*

Ref	Model	IS19 challenge	Huckvale et al. (n=26)	Martin et al. 2023 (n=30)
(a)	Lasso ($\alpha = 0.1$)	$\rho = 0.437$	$\rho = 0.049 \pm 0.166$ $[-0.412, 0.273]$	$\rho = 0.356 \pm 0.116$ $[0.107, 0.537]$
(b)	PCA (0.8) + Lin. regression	$\rho = 0.459$	$\rho = 0.066 \pm 0.164$ $[-282, 0.283]$	$\rho = 0.323 \pm 0.100$ $[0.115, 0.5]$
(c)	PCA (0.8) + SVR ($C = 1$)	$\rho = 0.447$	$\rho = 0.051 \pm 0.143$ $[-0.268, 0.267]$	$\rho = 0.289 \pm 0.095$ $[0.027, 0.424]$

2.7. Link between performances and annotators characteristics

Since the characteristics of the annotators were collected in the Martin et al. 2023 study, we computed Mann-Whitney tests in order to shed light on a possible link between the genre, the understanding of the language or the musical sensitivity of the listeners, and the performances of the pipeline trained to reproduce their annotations.

2.8. Profiles of annotators using hierarchical clustering

In order to draw profiles of listeners, we selected the most important features, i.e. those having a median value of absolute normalized contribution across the 57 annotation sets higher than 0.05. We then computed listeners profiles using hierarchical clustering using the `linkage` function from the `cluster.hierarchy` library of `scipy` [25]. The clustering has been performed using the ward method and an Euclidean metric. In order to get a more thoughtful insight into these profiles, we identified the profiles of annotators, i.e. the groups as returned by the `linkage` function. For each profile, we took into account the performances of the corresponding pipelines to be sure that no profile was dedicated to low-performance pipelines and that, on the contrary, every profile was represented by a diversity of performances.

3. Results

3.1. Pipeline performances

The mean, standard deviation, minimum, and maximum performances of the pipelines are reported in Table 1.

On replicating the IS19 challenge labels of the SLEEP subcorpus, our three pipelines obtain performances higher than the state-of-the-art systems on the whole corpus (cf. Introduction), confirming they are indeed suited for this task. On the annotations of the Huckvale et al. perceptual study, no classifier gives satisfying estimations of the labels: all the pipeline performances are below ρ=0.283 and most of them are negative, indicating that the pipeline did not generalize anything. As a consequence, we did not use them in the following. Contrastively, pipeline (a) achieves an average correlation coefficient of $\rho = 0.356$ when replicating the labels of the Martin et al. perceptual study, which is in the range of the usually obtained performances on the whole corpus (see Introduction).

3.2. Influence of speakers characteristics

We do not find any difference in the pipelines' performances depending on the sex (MW, $U = 147, p = 0.132$), musical sensitivity (MW, $U = 85, p = 0.271$), or the German understanding level (MW, $U = 145, p = 0.085$) of the listeners of the Martin et al. study. We thus infer that these variables do not bias our interpretation of the speech features implied in the estimation of sleepiness by the listeners.

3.3. Most prominent features

Over the 46 extracted features, six are identified as the most prominent, i.e. having a median absolute L1-normalized value across listeners higher than 0.05. They are reported in Table 2.

3.4. Hierarchical clustering

Hierarchical clustering was performed on these six features to draw profiles of listeners in the Martin et al. study. The clustering tree is represented in Figure 2. We identify three main profiles, which are represented along with the most prominent features and the performance of each pipeline in Figure 3.

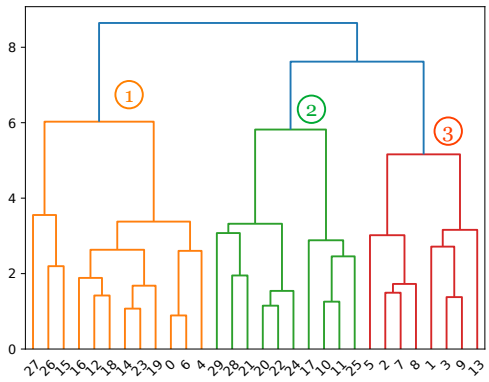

Figure 2: *Hierarchical clustering of the pipelines trained to reproduce listener's annotation depending on the six most prominent features*

The first group of listeners (Profile n°1, n=12) associate sleepiness with a voice having longer unvoiced segments, and a softer (loudnessPeksPerSec) and less expressive (slopeIV0-500 and shimer) voice, with a stronger focus on loudness. By contrast, listeners of Profile n°2 (n=10) estimate sleepiness using prosodic information (unvoiced segment length but also the number of voiced segments per second), voice expressiveness (slopeIV0-500, shimmer), but also voice purity (HNR variations). Finally, listeners of Profile n°3 (n=8) do not rely on the length of the unvoiced segment to identify sleepiness, but they focus on the number of voiced segments per second and pitch variability (slopeIV0-500).

4. Comparison with automatic approaches

To our knowledge, no previous system working on the SLEEP corpus has studied the contribution of descriptors to the estimation of sleepiness. However, a previous approach on the reading tasks of the SLC (same sleepiness label as the SLEEP corpus) reported the correlation between acoustic descriptors and

Table 2: *Most prominent features in the pipeline trained to imitate listeners of the perceptual study of Martin et al. 2023 [22]. Negative values mean that the value of the feature decreases when sleepiness increases.*

Name	Description	median value
`HNRdBACF_sma3nz_stddevNorm`	Standard deviation of the HNR	-0.102
`shimmerLocaldB_sma3nz_amean`	Average of the shimmer	-0.099
`slopeIV0-500_sma3nz_amean`	Frequency slope in the [0,500Hz] bandwidth	-0.095
`loudnessPeaksPerSec`	Average loudness peaks per second	-0.09
`VoicedSegmentsPerSec`	Number of voiced segments per second	-0.065
`MeanUnvoicedSegmentLength`	Average length of unvoiced segments	0.075

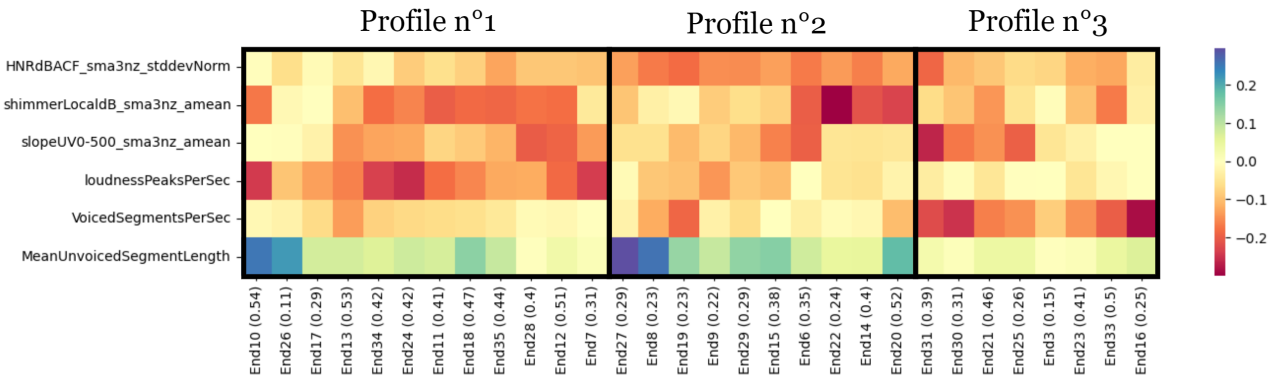

Figure 3: *Profiles of annotators as identified by the hierarchical clustering of the pipelines trained to imitate them.*

sleepiness [26]. In this work, the features correlating the most with sleepiness were mainly linked to F0 (mean, max, min), the frequency of the first formant (F1), and the energy range. On the reverse, the HNR and the duration of voiced or unvoiced segments were not among the features the most related to sleepiness. Moreover, applying the same methodology to our data, the features correlating the most with the ground truth given alongside the corpus are partly those identified as prominent in the imitation of listeners. Indeed, while the F0 slope ($\rho = -0.40$), the shimmer ($\rho = -0.34$) and the HNR ($\rho = -0.27$) are highly correlated with the sleepiness label, loudness peaks ($\rho = 0.10$), the duration of unvoiced segments ($\rho = 0.13$) and the number of voiced segments ($\rho = -0.10$) are not among the most prominent features in this view.

These results question the link between the ground truth given with the corpus and what listeners have detected. The high overall inter-annotator agreement reported by Martin et al. (ICC = 0.975) indicates that the listeners seem to have identified the same phenomena through voice, which is not completely what is represented by the measurement tool used to operationalize sleepiness in the SLEEP corpus. However, this label is criticized in the literature [9], since it is not a validated, used, and recognized measure of sleepiness in sleep medicine [23], and has never been used elsewhere to our knowledge than in the two IS2011 and IS2019 corpora. Furthermore, another perceptual study led by Martin et al. on the MSLTc [27], which contains validated measurements of sleepiness [9], has concluded the feasibility of human hearing to detect sleepiness using speech samples. We therefore interpret this difference between the features used by the annotators and the features correlated with the label supplied with the corpus as coming from the sleepiness measurement tool used in the SLEEP corpus.

5. Conclusion and perspectives

By training machine learning algorithms to reproduce the assessments of annotators in a perceptual study, we were able to identify the features they relied on to produce this assessment; and thus indirectly the cues they used to estimate sleepiness. We identified six features, related to energy stability (shimmer and energy peaks), the HNR, the variability of F0 (F0 slope), and the respective ratio and duration of the voiced and unvoiced segments.

Our next works will concentrate on including other dimensions such as reading pauses [28] or phonetic realization [29] into this perception-related study to better characterize sleepy speech.

This research is supported by the CNRS through the MITI PRIME 80 DSM-HEALTH and the French Research Agency ANR through the axis “Autonom-Health” of the PEPR “Santé Numérique”, Grant agreement n°ANR-22-PESN-000X and by the European Union’s Horizon Europe research and innovation program through the Marie Skłodowska-Curie grant agreement No. 101106577.

6. References

[1] D. Léger, V. Bayon, J. P. Laaban, and P. Philip, “Impact of sleep apnea on economics,” *Sleep Medicine Reviews*, vol. 16, no. 5, pp. 455–462, Oct. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1087079211000992

[2] C. M. Barnes and N. F. Watson, “Why healthy sleep is good for business,” *Sleep Medicine Reviews*, vol. 47, pp. 112–118, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1087079219300449

[3] M. Jike, O. Itani, N. Watanabe, D. J. Buysse, and Y. Kaneita, “Long sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression,” *Sleep Medicine Reviews*, vol. 39, pp. 25–36, 2018.

[4] A. J. Scott, T. L. Webb, M. Martyn-St James, G. Rowse, and S. Weich, "Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials," *Sleep Medicine Reviews*, vol. 60, p. 101556, 2021.

[5] B. P. Kolla, J.-P. He, M. P. Mansukhani, M. A. Frye, and K. Merikangas, "Excessive sleepiness and associated symptoms in the U.S. adult population: prevalence, correlates, and comorbidity," *Sleep Health*, vol. 6, no. 1, pp. 79–87, 2020.

[6] G. Fagherazzi, L. Zhang, A. Elbéji, E. Higa, V. Despotovic, M. Ollert, G. A. Aguayo, P. Nazarov, and A. Fischer, "A Voice-Based Biomarker for Monitoring Symptom Resolution in Adults with COVID-19: Findings from the Prospective Predi-COVID Cohort Study," *SSRN Electronic Journal*, 2021. [Online]. Available: https://www.ssrn.com/abstract=3949487

[7] B. Schuller, S. Steidl, A. Batliner, F. Schiel, and J. Krajewski, "The INTERSPEECH 2011 Speaker State Challenge," in *Interspeech 2011*, 2011, pp. 3201–3204.

[8] B. Schuller, A. Batliner, C. Bergler, F. B. Pokorny, J. Krajewski, M. Cychocz, R. Vollman, S.-D. Roelen, S. Schnieder, E. Bergelson, A. Cristia, A. Seidl, A. Warlaumont, L. Yankowitz, E. Nöth, S. Amiriparian, S. Hantke, and M. Schmitt, "The INTERSPEECH 2019 Computational Paralinguistics Challenge: Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity," in *Interspeech 2019*, 2019.

[9] V. P. Martin, J.-L. Rouas, J.-A. Micoulaud-Franchi, P. Philip, and J. Krajewski, "How to Design a Relevant Corpus for Sleepiness Detection Through Voice?" *Frontiers in Digital Health*, vol. 3, p. 686068, 2021.

[10] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual." *Int J Neurosci*, vol. 52, pp. 29–37, 1990.

[11] D.-Y. Huang, S. S. Ge, and Z. Zhang, "Speaker State Classification Based on Fusion of Asymmetric SIMPLS and Support Vector Machines," in *Interspeech 2011*, 2011, p. 4.

[12] G. Gosztolya, "Using Fisher Vector and Bag-of-Audio-Words Representations to Identify Styrian Dialects, Sleepiness, Baby & Orca Sounds," in *Interspeech 2019*, 2019, pp. 2413–2417.

[13] J. Fritsch, S. P. Dubagunta, and M. Magimai.-Doss, "Estimating the Degree of Sleepiness by Integrating Articulatory Feature Knowledge in Raw Waveform Based CNNS," in *ICASSP 2020*, Barcelona, Spain, May 2020, pp. 6534–6538. [Online]. Available: https://ieeexplore.ieee.org/document/9053351/

[14] S. Amiriparian, P. Winokurow, V. Karas, S. Ottl, M. Gerczuk, and B. W. Schuller, "A Novel Fusion of Attention and Sequence to Sequence Autoencoders to Predict Sleepiness From Speech," arXiv 2005.08722, 2020, _eprint: 2005.08722.

[15] J. V. Egas-López, R. Busa-Fekete, and G. Gosztolya, "On the Use of Ensemble X-Vector Embeddings for Improved Sleepiness Detection," in *Speech and Computer*, ser. Lecture Notes in Computer Science, S. R. M. Prasanna, A. Karpov, K. Samudravijaya, and S. S. Agrawal, Eds. Cham: Springer International Publishing, 2022, pp. 178–187.

[16] E. L. Campbell, L. Docio-Fernandez, C. Garcia-mateo, A. Wittenborn, J. Krajewski, and N. Cummins, "Automatic detection of short-term sleepiness state. Sequence-to-Sequence modelling with global attention mechanism." in *Workshop on Speech, Music and Mind*, 2022.

[17] B. Tran, Y. Zhu, X. Liang, J. W. Schwoebel, and L. A. Warrenburg, "Speech Tasks Relevant to Sleepiness Determined With Deep Transfer Learning," in *ICASSP 2022*, May 2022, pp. 6937–6941, iSSN: 2379-190X.

[18] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W. C. Dement, "Quantification of Sleepiness: A New Approach," *Psychophysiology*, vol. 10, no. 4, pp. 431–436, 1973. [Online]. Available: http://doi.wiley.com/10.1111/j.1469-8986.1973.tb00801.x

[19] V. P. Martin, J.-L. Rouas, F. Boyer, and P. Philip, "Automatic Speech Recognition systems errors for objective sleepiness detection through voice," in *Interspeech 2021*, Brno, 2021, pp. 2476–2480.

[20] J. Krajewski, A. Batliner, and M. Golz, "Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach," *Behavior Research Methods*, vol. 41, no. 3, pp. 795–804, 2009.

[21] M. Huckvale, A. Beke, and M. Ikushima, "Prediction of Sleepiness Ratings from Voice by Man and Machine," in *Interspeech 2020*, 2020.

[22] V. P. Martin, A. Ferron, J.-L. Rouas, and P. Philip, ""Prediction of Sleepiness Ratings from Voice by Man and Machine": a perceptual experiment replication study," in *ICASSP 2023*, 2023.

[23] V. P. Martin, R. Lopez, Y. Dauvilliers, J.-L. Rouas, P. Philip, and J.-A. Micoulaud-Franchi, "Sleepiness in adults: An umbrella review of a complex construct," *Sleep Medicine Reviews*, vol. 67, p. 101718, 2023.

[24] F. Eyben and B. Schuller, "Opensmile," *ACM SIGMultimedia Records*, vol. 6, pp. 4–13, 2015.

[25] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," 2011, publisher: arXiv Version Number: 1. [Online]. Available: https://arxiv.org/abs/1109.2378

[26] V. P. Martin, J.-L. Rouas, P. Thivel, and J. Krajewski, "Sleepiness detection on read speech using simple features," in *10th Conference on Speech Technology and Human-Computer Dialogue*, Timisoara, Romania, 2019.

[27] V. P. Martin, A. Ferron, J.-L. Rouas, T. Shochi, L. Dupuy, and P. Philip, "Physiological vs. Subjective sleepiness: what can human hearing estimate better?" in *ICPhS 2023*, 2023, pp. 196–200.

[28] V. P. Martin, B. Arnaud, J.-L. Rouas, and P. Philip, "Does sleepiness influence reading pauses in hypersomniac patients?" in *Speech Prosody 2022*. ISCA, 2022, pp. 62–66.

[29] C. Beaumard, V. P. Martin, Y. Wu, J.-L. Rouas, and P. Philip, "Automatic detection of schwa in French hypersomniac patients," in *Journée Santé et Intelligence Artificielle (Evènement affilié à PFIA 2023)*, 2023.