SOMVOICE: A First Dataset to Study the Effects of Sleep Deprivation on Voice Characteristics of Healthy French Speakers

Vincent P. Martin[1], **Jean-Luc Rouas**[2], **Colleen Beaumard**[2,3], **Pierre Philip**[3,4]
[1] Univ. Lorraine CNRS, Inria, LORIA, F-54000 Nancy, France
[2] LaBRI CNRS UMR 5800, Univ. Bordeaux, Bordeaux INP, F-33405 Talence, France,
[3] SANPSY CNRS UMR 6033, Univ. Bordeaux, F-33000 Bordeaux, France,
[4] CHU Bordeaux, University Department of Sleep Medicine, F-33000 Bordeaux, France
vpmartin@inria.fr; {jean-luc.rouas, colleen.beaumard}@labri.fr;
pierre.philip@u-bordeaux.fr

Abstract

Excessive sleepiness is a significant public health issue and a critical personal health indicator associated with various disorders. Given its high prevalence in the general population, clinicians need tools to regularly measure patients' sleepiness levels in natural settings, such as automatic speech analysis. In this article, we introduce the SOMVOICE corpus, the first French corpus containing read-speech recordings from the same participants either after a normal night or after a night of total sleep deprivation. Participants were included according to strict inclusion and exclusion criteria based on both medical characteristics and reading proficiency. The recordings were labelled with both objective and subjective measures of sleepiness, as well as fatigue and anxiety. After introducing the data-collection methodology, we use linear mixed models to conduct a preliminary investigation of the effect of total sleep deprivation on the collected sleepiness-related measures and on participants' reading behaviour. Doing so, we found that sleep deprivation strongly influences objective and subjective sleepiness measurements as well as fatigue self-reports, but has a lesser effect on anxiety. Regarding reading behaviour, sleep deprivation is associated with a lower speech rate (duration of the recordings and phoneme rate) and more pauses (number of pauses and pause ratio).

Keywords: total sleep deprivation, read speech, sleepiness

1. Introduction

1.1. Context

Excessive sleepiness is a significant public health issue and a critical personal health indicator associated with various disorders, including metabolic, cardiovascular, neurological, and psychiatric conditions. It increases the risk of disability and mortality (Jike et al., 2018; Scott et al., 2021). Given its high prevalence in the general population, with up to one in three people affected (Kolla et al., 2020), clinicians need tools to regularly measure patients' sleepiness levels in natural settings, such as at home, without requiring dedicated tasks. Voice and speech recordings are promising candidates for this purpose: they can be easily collected using smartphones, recorded passively, and have already been linked to multiple health conditions, including sleepiness (Ramanarayanan et al., 2022).

1.2. Previous corpora

Previous corpora for studying manifestations of sleepiness in the voice of healthy adults have been designed around three main approaches, depending on how sleepiness is measured or induced: i) collecting speech samples from general-population participants and labelling them with sleepiness measurements (with a perspective of *ecological* sleepiness estimation); ii) laboratory experiments submitting healthy subjects to (partial) sleep deprivation; or iii) recording patients with sleep disorders.

General population. Among the corpora belonging to this first category, two were developed by the same team at the University of Düsseldorf for Interspeech challenges in 2011 and 2019. The Sleepy Language Corpus, introduced in 2011, contains 9,089 recordings (approx. 21 hours) from 99 German-speaking speakers, collected across a wide range of tasks (Schuller et al., 2011). In 2019, the SLEEP corpus was released: it contains 10,892 samples (approx. 12 hours) recorded by 915 speakers (Schuller et al., 2019). For both corpora, sleepiness was annotated as the mean of three Karolinska Sleepiness Scale (KSS) scores, using an original methodology not recognised in sleep medicine (Martin et al., 2021, 2023b).

More recently, Tran et al. (2022) published the Voiceome dataset. Containing 186.2 hours of recordings from 1,828 participants from the general population, the level of sleepiness in this corpus was annotated using the Stanford Sleepiness Scale (SSS). A major limitation of this annotation is that it does not allow the distinction between fatigue and sleepiness, a difference of particular importance in sleep medicine (Shen et al., 2006).

Laboratory studies. In parallel with these works, studies have investigated the effects of (partial)

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 5597–5606
11–16 May 2026.

sleep deprivation on vocal characteristics, with the aim of identifying vocal surrogates for performance measurements.

This is the case for Golz et al. (2007), who correlated vocal descriptors from 24 German subjects with micro-sleep events during a driving task; Dhupati et al. (2010), who linked electroencephalographic descriptors (alpha-band energy) and vocal descriptors in 20 healthy English-speaking subjects during a 36-hour procedure designed to induce physical and mental fatigue; and Boyer et al. (2016), who recorded 22 French subjects after one night of total sleep deprivation, with vigilance estimated using a custom version of the Karolinska Drowsiness Test. All these studies share a focus on performance, which is a construct distinct from sleepiness (Martin et al., 2021).

More recently, Thoret et al. (2024) recorded 22 healthy French women under a partial sleep-deprivation protocol (3 hours of sleep, 03:00–06:00, for 2 nights). On day 1 and day 3, participants read different chapters from *Le Comte de Monte Cristo* (A. Dumas, 1844) at three different times (09:00, 15:00, 17:00). Unfortunately, these recordings were labelled only with the SSS (and thus suffer from its limitations), and they were not annotated with a standardised or recognised objective clinical sleepiness measurement.

Pathological sleepiness. Finally, we introduced in 2020 the Multiple Sleep Latency Test corpus (MSLTc), which contains recordings from 131 French patients (80 F) followed up for sleep disorders at the Bordeaux University Hospital sleep clinic (Martin et al., 2020, 2021). A key strength of this corpus is that it includes both *subjective* (questionnaire-based) and *physiological* (electroencephalography-based) measures of sleepiness, enabling direct comparison between them (Martin et al., 2023a). Moreover, these measurements are validated and widely used in sleep medicine (Martin et al., 2023b).

However, although including patients makes it possible to observe higher sleepiness levels than in the general population, it also reduces variability in sleepiness levels, which can weaken associations between voice and sleepiness fluctuations.

1.3. Originality

The aforementioned studies either focused on performance, did not include objective clinical tools for assessing sleepiness, or did not consider healthy participants (i.e. with inclusion and exclusion criteria verifying the absence of sleep disorders). Moreover, since we aim to work with French sleep specialists, we need our corpus to include French-speaking participants. To our knowledge, the SOMVOICE corpus introduced in this article is the only one to combine the following characteristics:

1. Each participant is recorded after both a normal night and a night of total sleep deprivation (randomised order);
2. Strict inclusion and exclusion criteria were applied, both regarding the medical characteristics of participants and, for the first time, regarding participants' reading proficiency;
3. Sleepiness is measured using an objective clinical tool recognised in sleep medicine (Martin et al., 2023b);
4. Each recording is associated with measures of sleepiness, fatigue, and performance;
5. The participants speak French.

1.4. Objectives

In this paper, we first present the methodology used for collecting the corpus (Section 2). Then, we conduct a preliminary investigation of the effect of total sleep deprivation on the sleepiness-related measures collected in the corpus (Section 3) and on the participants' reading behaviour (Section 4). Finally, we conclude and outline our future work in Section 5.

2. Data collection

2.1. Recruitment

The participants were recruited by word of mouth and poster campaigns. The first participant was included on 7 July 2022, and the last on 10 June 2024. Participants who completed the four visits (2 inclusion visits + 2 experimental visits) received financial compensation of 400€ (13.9% of the French gross salary in 2021).

2.2. Inclusion and exclusion criteria

To our knowledge, SOMVOICE is the first corpus to apply inclusion and exclusion criteria regarding both medical characteristics (e.g. absence of sleep or mood disorders) and reading proficiency (evaluation test). This ensures that sleep disorders or poor sleep hygiene do not interfere with the effect of total sleep deprivation, and that the speech descriptors we aim to link with sleepiness and sleep deprivation are independent of education level or language disorders. Details of all inclusion and exclusion criteria for the recruited population are available on the ClinicalTrials.gov platform[1].

---

[1] https://clinicaltrials.gov/study/NCT04942574

Evaluation	Task	Inclusion	M (n=12)	F (n=16)
Lexical decoding level	Reading *L'Alouette*	MCLM ≥ 144	171.0 (33.9)	171.4 (25.3)
		Errors < 6	2.1 (1.16)	1.2 (1.2)
Lexical procedure	Reading 20 regular words	Corr. ≥ 19	19.75 (0.45)	19.7 (0.6)
		Dur. $\leq$ 13s	9.8 (2.4)	10.8 (2.0)
Lexical procedure	Reading 20 irregular words	Corr. ≥ 18	19.5 (0.8)	19.7 (0.48)
		Dur. $\leq$ 14s	10.1 (2.4)	9.8 (2.1)
Lexical procedure	Reading 20 pseudo-words	Corr. ≥ 19	19.5 (0.8)	19.5 (0.6)
		Dur. $\leq$ 21s	16.25 (4.5)	16.75 (3.4)
Phonological loop	Repeating 20 pseudo-words	n ≥ 19	19.7 (0.5)	19.75 (0.45)
Sequential visual attention	Symbols crossing in 90s	n > 25	34.5 (5.7)	38.2 (5.0)

Table 1: Descriptive statistics and inclusion criteria of the participants. A small yet significant difference is observed between sexes regarding the *symbol crossing* task (Mann-Whitney test, $W = 51.5, p = 0.04$). *MCML*: Number of correctly read words per minute; *Corr.*: number of correctly read words; *Dur.*: Duration.

Medical characteristics. Participants underwent a medical interview to rule out the presence of hypersomnolence, insomnia, circadian disorders, psychiatric disorders, addictive behaviours, and any pharmaceutical treatment associated with these disorders. This initial screening was completed with seven-night actimetry and ambulatory polygraphy, ensuring sufficient regularity and sleep efficiency ($\geq$ 85%), and ruling out sleep apnea and restless legs syndrome.

Reading proficiency. All potential participants in our study were screened for reading proficiency. To do this, they completed four tasks from the ECLA16+ assessment battery[2]. First, they read aloud *L'Alouette*, a non-semantic text usually used for screening reading difficulties such as dyslexia. Then, they read as fast as possible 10 regular words, 10 irregular words, and 10 pseudo-words, and they performed a pseudo-word repetition task. These four tasks allow the assessment of lexical and phonological pathways. Finally, they performed a symbol-crossing task: they had to cross out as many occurrences as possible of two specified symbols on an A3 paper sheet filled with different symbols, measuring their sequential visual attention. Normative values for the tasks in this battery were previously established by another team for 311 French individuals aged 16 (158 women), drawn from the general population and assessed during their day of introduction to national defense and the French armed forces[3].

Potential participants in our study had to perform above the 25th percentile of these normative values (i.e. better than at least 75% of the reference participants). The inclusion criteria for the SOMVOICE study based on these tests, and the descriptive statistics of the included participants' performances, are reported in Table 1.

In addition to these inclusion criteria, potential participants were excluded if they had an ear, nose, or throat disorder that could cause voice changes; treatment that interfered with normal muscle function (e.g. muscle relaxants); dental care planned during the study follow-up period; or if French was not their native language.

Results. We ultimately included 28 participants. Their age, sex, and Body Mass Index (BMI) distributions are reported in Table 2.

Factor	Inclusion	M (n=12)	F (n=16)
Age	[18, 50]	28.3 (7.2)	35.2 (10.9)
BMI	[18.5, 25]	22.4 (2.2)	21.7 (2.2)

Table 2: Descriptive statistics of the participants

2.3. Recordings and environment

The corpus was recorded in a sleep research laboratory, specifically equipped to guarantee optimal sleep conditions. The rooms are acoustically isolated from the outside and are equipped with a dimmer, allowing total obscurity. The rooms are furnished with a bed and a desk, and participants have access to a lounge, equipped with a sofa, a kitchen and a television. Although inside a research laboratory, the experimental zone can be totally isolated from the other parts of the laboratory.

The corpus was recorded by Clinical Research Assistants employed by the University Hospital of Bordeaux, who were specifically trained in speech-recording techniques. During speech recording and questionnaire completion, participants were systematically seated at the desk in their room, with the microphone pointed toward them at approximately 30 cm from their mouth. The microphone was an omnidirectional Audio-Technica AT4022 connected to a Tascam DR-100 MKIII audio recorder. At the beginning of each recording,

---

[2] [fr] https://www1.ac-grenoble.fr/article/cognisciences-121593

[3] These normative values are available online: https://www1.ac-grenoble.fr/media/14947/download

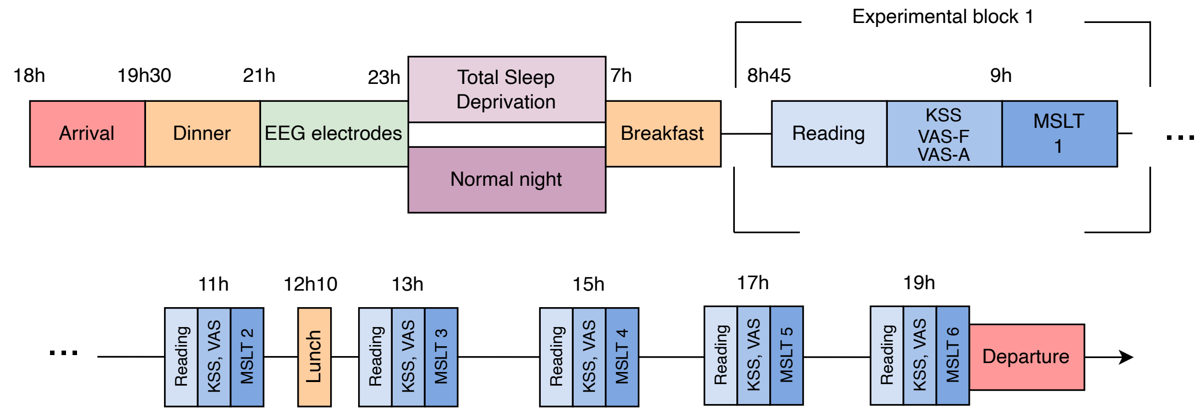

Figure 1: Experimental procedure for the two experimental visits of the participants.

the clinical assistant said aloud the anonymized participant ID and the reference of the text, to ensure robust identification of the recording and safe merging with clinical data. The recordings were stored in 48 kHz, 16-bit WAV PCM format.

2.4. Experimental visits

A schematic view of the experimental visits is shown in Figure 1. The procedure was the same for the two experimental visits, except for the night before the experimental blocks: participants either underwent total sleep deprivation or spent a normal night at the lab.

2.4.1. Night before the evaluation

In both experimental conditions, participants were welcomed the evening before the experimental blocks. After being installed in their room and having a meal, they were equipped with electroencephalography (EEG) electrodes linked to an ambulatory polysomnograph. The remainder of the evening differed across experimental conditions.

When participants were assigned to spending a normal night at the lab, lights were switched off at 23:00, and the required minimum sleep time of 7 hours was checked using the EEG recordings from that night during the following day.

Otherwise, participants were assigned to total sleep deprivation. Under this condition, participants were kept under surveillance by a dedicated research assistant during the evening and the whole night, to ensure that they did not fall asleep. Participants were allowed to perform calm activities (e.g. reading, watching TV) and have snacks. They were allowed to drink any beverage as long as it did not contain caffeine, theine, or alcohol. Work or learning activities were not allowed.

In both conditions, participants were given the opportunity to have breakfast between 07:00 and 07:45 the following day, before the experimental blocks.

2.4.2. Experimental blocks

The experimental day was paced by six similar experimental blocks, corresponding to the six iterations of a Multiple Sleep Latency Test (cf. Section 3.2). First, participants read a text aloud. Then, they completed three questionnaires: the Karolinska Sleepiness Scale (KSS) about their instantaneous sleepiness (Åkerstedt and Gillberg, 1990), and two Visual Analog Scales for fatigue (VAS-F) (Lee et al., 1991) and anxiety (VAS-A) (Williams et al., 2010). Finally, they lay on the bed, with lights switched off, for the sleep latency test (cf. Section 3.2).

3. Effect of sleep deprivation on medical labels

In this section, we describe the sleepiness-related measurements collected from participants during the evaluation day, and we measure the impact of total sleep deprivation on them.

3.1. Statistical model

Each sleepiness-related measurement (measure) was modelled using a linear mixed-effects model including, as fixed effects, total sleep deprivation status (tsd: 0 or 1), recording time to account for temporal evolution (time: 09:00–19:00), and visit (visit: v1 or v2), along with a random intercept per participant (ID) to account for inter-individual variability and the repeated nature of the measurements:

$$measure \sim tsd(0|1) + time[9-19] + visit(v1|v2)$$

The statistical models' p-values were corrected using the Benjamini–Hochberg correction.

3.2. Multiple Sleep Latency Test

Method. During their experimental visits, participants' day was structured around a Multiple Sleep Latency Test, performed according to international recommendations (Krahn et al., 2021). This test consists of six 20-minute nap opportunities every two hours (from 09:00 to 19:00): participants are put to bed in darkness and asked to close their eyes and try to sleep. Sleep onset[4] – if it occurred – was identified from the EEG signal by a trained specialist according to international standards (Rechtschaffen and Kales, 1968). The delay between the beginning of the nap opportunity and sleep onset is called *sleep latency*. Mean Sleep Latency (MSL) is computed as the average of the six sleep latencies and is the gold-standard measurement in sleep medicine (Martin et al., 2023b). If a participant fell asleep, they were immediately (and gently) woken up by the research assistant, in order to maintain sufficient sleep pressure and avoid recovery, thus favouring more severe sleepiness states.

Results. Sleep latencies as a function of visit and deprivation status are reported in Figure 2 (top). We did not find any significant difference between visits, showing no habituation effect. Sleep latency measured during the last nap opportunity (19:00) is significantly higher than baseline ($\beta = 3.6, p < 0.001$). This observation has already been described in the literature as the 'last nap effect': since participants are impatient to complete the testing day – and thus to go back home – they struggle to relax sufficiently to fall asleep quickly (Goddard et al., 2021).

After controlling for variations related to time, order of visits and intra-subject variations, sleep latencies are in average 9.3min shorter during the total sleep deprivation condition ($\beta = -9.3, p < 0.001$), indicating a higher level of physiological sleepiness.

3.3. Karolinska Sleepiness Scale

Method. The Karolinska Sleepiness Scale (Åkerstedt and Gillberg, 1990) is a one-item questionnaire asking participants to rate their "degree of sleepiness during the last 10 minutes" on a scale from 1 ('Extremely alert', 'Parfaitement éveillé' in French) to 10 ('Extremely sleepy, can't keep awake', 'Extrêmement somnolent, ne peut rester éveillé' in French). In our study, participants completed the French version translated, validated, and used at Bordeaux University Hospital.

Results. The KSS scores are reported in Figure 2 (bottom). The KSS score is slightly yet significantly lower during the 13:00 experimental block ($\beta = -0.64, p < 0.05$), after lunch. We did not find any significant difference between visits, showing no habituation effect. After controlling for all other effects, participants under total sleep deprivation had KSS scores that were on average 3.48 points higher ($\beta = 3.48, p < 0.001$).

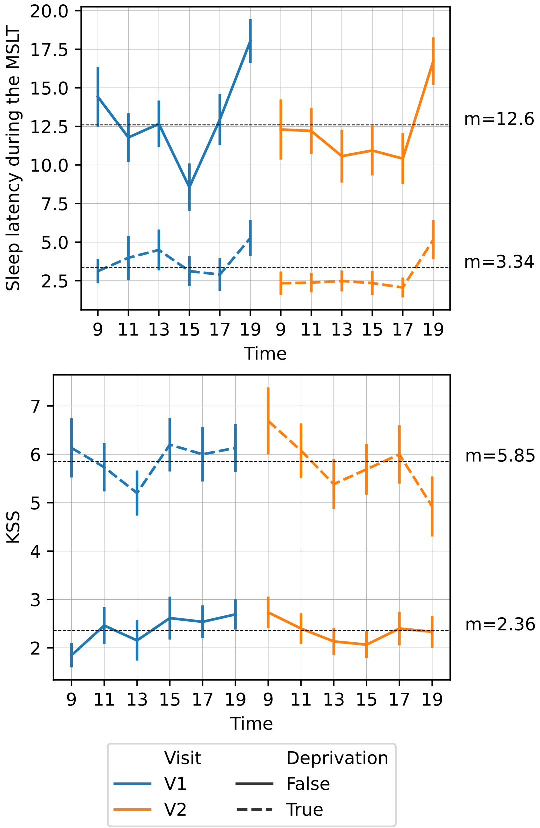

Figure 2: Physiological (top) and subjective (bottom) sleepiness measurements

3.4. Fatigue and anxiety

Method. In addition to the KSS, participants were also asked to complete, during each experimental block, two other questionnaires: the Visual Analogue Scales for fatigue (VAS-F) and anxiety (VAS-A). These are 10 cm lines on which participants place a mark to indicate symptom intensity between two anchors: 'Very fit' ('En excellente forme physique') and 'Very tired' ('Très fatigué') for fatigue; 'Very relaxed' ('Très détendu') and 'Very anxious' ('Très anxieux') for anxiety.

Results. The VAS-F and VAS-A scores are reported in Figure 3. For both measurements, we did not observe any significant effect of either time or visit. After controlling for all other variables, fatigue scores were significantly higher when participants were under total sleep deprivation by an average

---

[4] 3 epochs of 30s in sleep stage 1 or 1 epoch of any other stage

of 44.4 points ($\beta = 44.4, p < 0.001$).

Similarly, but to a lesser extent, anxiety scores were slightly higher under total sleep deprivation ($\beta = 3.3, p < 0.05$). Based on the trajectories displayed in Figure 3 (bottom), this difference in axnsity scores seems to apply mostly to participants who underwent sleep deprivation during their first visit (dashed blue line). A complementary linear mixed model that accounted for the interaction between *privation* and *visit* confirmed this hypothesis (privation:V1, $\beta = 9.33, p < 0.03$).

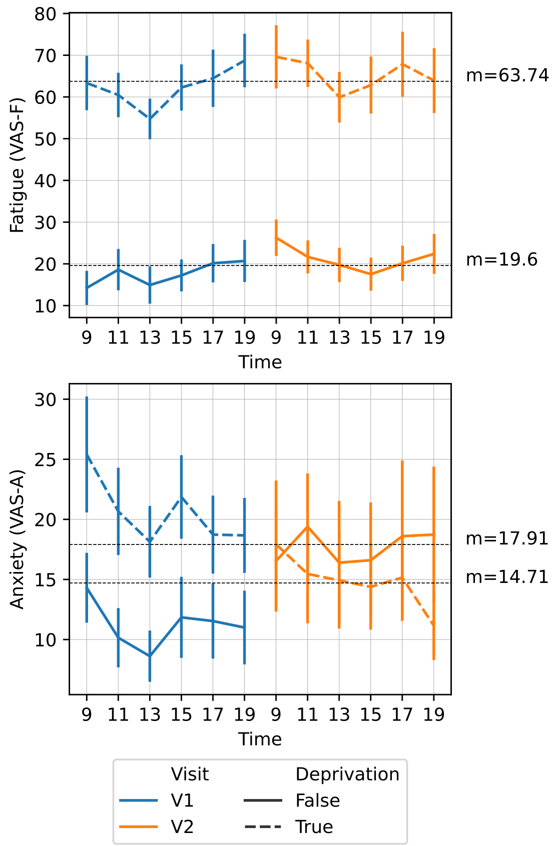

Figure 3: Fatigue (top) and anxiety (bottom) Visual Analog Scales scores

4. Effect of sleep deprivation on reading behaviour

In this section, we describe the texts read by participants during their experimental visits. Then, we compute four metrics (duration, phoneme rate, number of pauses, and pause-time ratio) to describe participants' reading behaviour and evaluate the effect of sleep deprivation on these metrics.

4.1. Read texts

Each participant read a total of 12 texts, always in the same order, independently of sleep deprivation status. These texts were carefully selected by master's students in speech therapy to be accessible, to limit the induction of strong emotions (e.g. compared with newspaper reading), to be balanced in length (approximately 150 words), and to avoid dialogues. Moreover, to facilitate sharing, we limited our search to public domain texts. Finally, we wanted the 12 texts to be extracted from the same book in order to homogenise writing style – and therefore the corresponding speaking style (Martin et al., 2021) – while also motivating participants, who benefited from the continuity of a narration they followed across experimental days. These constraints led us to choose the novel *Le Petit Chose* by Alphonse Daudet (1868). To facilitate reading, we changed some archaisms and the names of some characters. The texts are available online[5].

The texts, their number of words and phones, and the duration of the corresponding recordings are reported in Table 3.

n°	**#W**	**#P**	**Dur. (s)**	**#P$_{obs}$**
01	175	531	59.8 (8.2)	531.1 (10.4)
02	181	585	66.1 (8.6)	593.0 (12.6)
03	181	563	57.9 (7.5)	568.0 (12.6)
04	185	560	56.7 (7.1)	563.9 (10.5)
05	192	540	57.4 (6.9)	547.1 (13.0)
06	179	568	59.1 (7.3)	575.2 (8.6)
07	181	553	60.1 (7.3)	556.1 (7.5)
08	171	568	60.7 (7.8)	574.3 (13.3)
09	181	583	62.2 (7.3)	592.4 (11.3)
10	182	573	62.2 (7.8)	581.5 (12.6)
11	190	592	60.7 (7.6)	593.7 (7.5)
12	184	565	58.0 (6.9)	565.4 (8.0)
TOTAL			**5 h 32 min 33 s**	**189,275**

Table 3: Characteristics of the texts read by the participants. #W: total number of words; #P: total number of expected phonemes; Dur.: Mean duration (std); #P$_{obs}$: mean observed number of phonemes (std)

4.2. Statistical model

Similarly to Section 3, speech characteristics (*speech charac.*) were modelled using a linear mixed-effects model including, as fixed effects, total sleep deprivation status (0 or 1) and the read text (text: 1–12), along with a random intercept per participant (ID) to account for inter-individual variability and the repeated nature of the measurements:

$$speech\ charac. \sim text[1-12] + tsd(0|1)$$

Similarly to the previous models, we corrected p-values using the Benjamini–Hochberg correction.

4.3. Duration and phoneme rate

In a first approach, we hypothesise that participants under sleep deprivation will show an overall slowing

---

[5] https://osf.io/z9h5m

of speech (Schuller et al., 2013). We quantify this effect by recording duration (i.e. overall reading speed) and phoneme rate (number of phonemes per second during inter-pause units).

Method. Recording duration was computed using SoX. Phoneme transcription and segmentation were obtained by running the Montreal Forced Aligner (McAuliffe et al., 2017) with the audio and a custom automatic speech recognition (ASR) transcription (see Appendix A). Phoneme rate was then computed by dividing the total number of phonemes found in each recording by its duration.

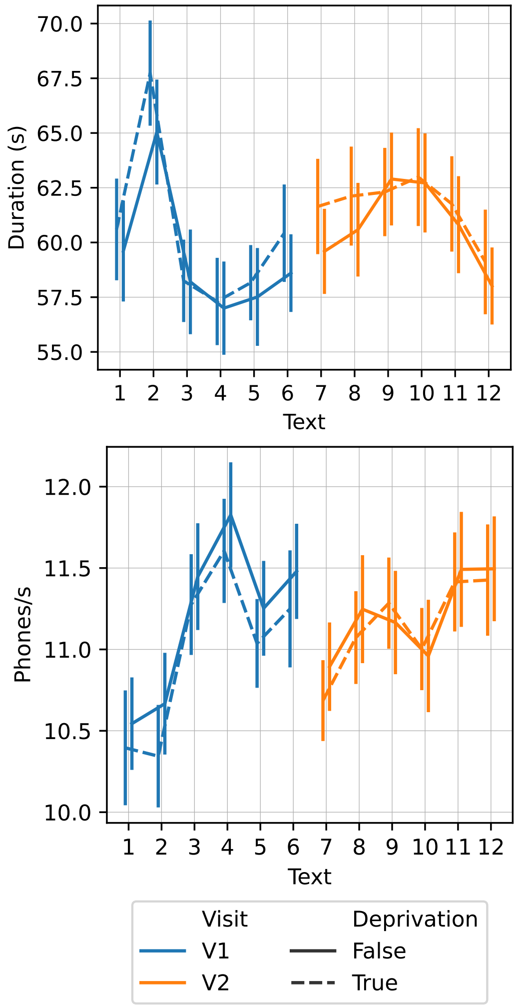

Figure 4: Duration of the recordings (top) and phoneme rate (bottom)

Results. Reading duration and phoneme rate are plotted in Figure 4. Unsurprisingly, although we carefully selected texts with approximately the same length (in terms of number of words), recording duration is highly sensitive to text: recordings corresponding to texts T2, T9, and T10 have significantly longer duration than the baseline (T1), whereas texts T3, T4, T5, and T12 have significantly shorter duration than the baseline. Text also has a slight effect on phoneme rate, with text T1 resulting in a significantly slower articulatory flow ($p < 0.001$) than all other texts except T2. We did not find any effect of the visit.

After controlling for all other variables, participants under total sleep deprivation spoke slightly yet significantly more slowly than in the normal night condition, with longer reading duration ($\beta = 0.99, p < 0.01$) and lower phoneme rate ($\beta = -0.14, p < 0.001$). This may reflect a compensatory strategy: after total sleep deprivation, reduced cognitive capacity may require readers to slow down to plan their oral reading more effectively. These results should, however, be interpreted with caution, as the main sources of variability are inter-individual differences and text effects.

4.4. Number of pauses and pause ratio

In a second approach, we hypothesise that participants under sleep deprivation will make more and longer pauses (Martin et al., 2022, 2024b), to gain the time required to plan the continuation of reading, which is made more difficult by lower cognitive resources.

Method. Pauses were extracted using `respiro-en` (Yang et al., 2024). We then counted them and extracted the ratio of pause duration to total recording duration.

Results. The number of pauses and pause-duration ratio are reported in Figure 5.

Similarly to speech rate, text has a significant effect on the number of pauses: texts T3–T9 and T12 are associated with significantly lower pause counts than T1. Regarding pause-duration ratio, we observed only T3 to have a significantly lower value than T1.

After controlling for all other variables, total sleep deprivation was significantly associated with a higher number of pauses ($\beta = 1.12, p < 0.001$) and a higher pause ratio ($\beta = 0.66, p < 0.05$), specifically in the second visit (Figure 5). These results are consistent with the previous findings on speech rate: after total sleep deprivation, participants need to slow down to plan reading and add pauses to manage the cognitive load of the reading-aloud task.

5. Conclusion and future works

To conclude, we introduced in this article the SOMVOICE corpus, a first dataset for studying the impact of total sleep deprivation on healthy French speech. This corpus contains recordings

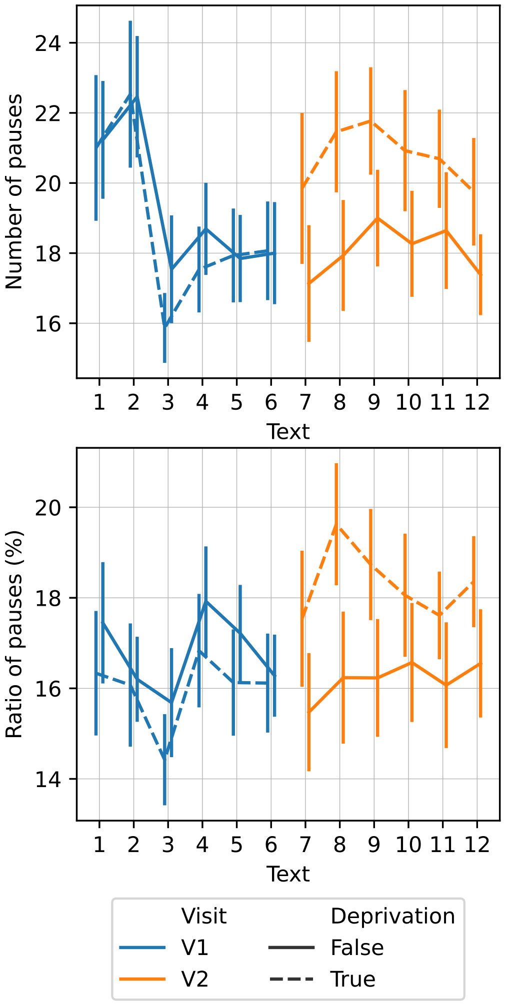

Figure 5: Number of pauses (top) and pause-duration ratio (bottom)

from 28 healthy speakers (16 F), who successfully passed strict inclusion and exclusion criteria related to both sleep medicine and reading abilities. They were recorded reading aloud texts during a Multiple Sleep Latency Test (6 nap opportunities), after either a normal night or a total sleep-deprivation night (randomised order). Our first analyses show, on the one hand, that the sleep-deprivation procedure was effective (sleepiness-related measurements were significantly higher after total sleep deprivation) and, on the other hand, that it significantly impacts speech flow and pauses.

Our future work includes evaluating the effect of total sleep deprivation on other speech characteristics (e.g. prosody or voice quality). Moreover, we plan to quantify the link between degree of sleepiness (i.e. sleepiness-related measurements) and speech characteristics, for example through the development of machine-learning algorithms that automatically estimate sleepiness severity from audio recordings. Finally, we plan to compare the SOMVOICE corpus, which contains recordings from healthy participants, with previous corpora focusing on pathological sleepiness (e.g. the MSLTc), in order to investigate whether sleepiness has the same impact on speech in healthy subjects as in patients affected by sleep disorders.

6. Acknowledgements

This research was supported by the SOMVOICE project sponsored by the Labex BRAIN (ANR-10-LABX-43); by the CNRS through the MITI PRIME 80 DSM-HEALTH; and by the French National Research Agency (ANR) through the “Autonom-Health” axis of the PEPR “Santé Numérique”, Grant Agreement No. ANR-22-PESN-000X.

7. Ethical considerations

The protocol was approved by the *Comité de Protection des Personnes Sud Ouest et Outre Mer IV* (No. CPP2021-01-009a) and registered on ClinicalTrials.gov (ID: NCT04942574).

8. Bibliographical References

Torbjorn Åkerstedt and Mats Gillberg. 1990. Subjective and objective sleepiness in the active individual. *Int J Neurosci*, 52:29–37.

S. Boyer, R. El-Yagoubi, M. Tiberge, R. Ruiz, and A. Daurat. 2016. Paramètres Acoustiques de la Voix et Privation de Sommeil. In *CFA/VISHNO*.

Lakshmi Swathi Dhupati, Sibsambhu Kar, Aparna Rajaguru, and Aurobinda Routray. 2010. A novel drowiness detection scheme based on speech analysis with validation using simultaneous EEG recordings. In *IEEE - Int. CASE*, pages 917–921.

Sylvain Galliano, Guillaume Gravier, and Laura Chaubard. 2009. The ESTER 2 Evaluation Campaign for the Rich Transcription of French Radio Broadcasts. In *Interspeech 2009*, pages 2583–2586.

John Goddard, George Tay, Jennifer Fry, Mark Davis, Deanne Curtin, and Irene Szollosi. 2021. Multiple Sleep Latency Test: When are 4 naps enough? *Journal of Clinical Sleep Medicine*, 17(3):491–497.

Martin Golz, David Sommer, Mo Chen, Dalino Mandic, and Udo Trutschel. 2007. Feature Fusion for the Detection of Microsleep Events. *Journal of VLSI Signal Processing*, 49:329–342.

Maki Jike, Osamu Itani, Norio Watanabe, Daniel J. Buysse, and Yoshitaka Kaneita. 2018. Long sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression. *Sleep Medicine Reviews*, 39:25–36.

Bhanu Prakash Kolla, Jian-Ping He, Meghna P. Mansukhani, Mark A. Frye, and Kathleen Merikangas. 2020. Excessive sleepiness and associated symptoms in the U.S. adult population: Prevalence, correlates, and comorbidity. *Sleep Health*, 6(1):79–87.

Lois E. Krahn, Donna L. Arand, Alon Y. Avidan, David G. Davila, William A. DeBassio, Chad M. Ruoff, and Christopher G. Harrod. 2021. Recommended protocols for the Multiple Sleep Latency Test and Maintenance of Wakefulness Test in adults: Guidance from the American Academy of Sleep Medicine. *Journal of Clinical Sleep Medicine*, 17(12):2489–2498.

Anne Lacheret-Dujour, Sylvain Kahane, and Paola Pietrandrea, editors. 2019. *Rhapsodie: A Prosodic and Syntactic Treebank for Spoken French*, volume 89 of *Studies in Corpus Linguistics*. John Benjamins Publishing Company, Amsterdam.

Kathryn A Lee, Gregory Hicks, and German Nino-Murcia. 1991. Validity and reliability of a scale to assess fatigue. *Psychiatry research*, 36(3):291–298.

Vincent P. Martin, Brice Arnaud, Jean-Luc Rouas, and Pierre Philip. 2022. Does sleepiness influence reading pauses in hypersomniac patients? In *Speech Prosody 2022*, pages 62–66. ISCA.

Vincent P. Martin, Colleen Beaumard, Jean-Luc Rouas, and Yaru Wu. 2024a. Is automatic phoneme recognition suitable for speech analysis? Temporal and performance evaluation of an Automatic Speech Recognition model in spontaneous French. In *Speech Prosody 2024*, pages 1120–1124.

Vincent P. Martin, Aymeric Ferron, Jean-Luc Rouas, Takaaki Shochi, Lucile Dupuy, and Pierre Philip. 2023a. Physiological vs. Subjective sleepiness: What can human hearing estimate better? In *International Conference on Phonetic Science (ICPhS) 2023*, pages 196–200.

Vincent P. Martin, Régis Lopez, Yves Dauvilliers, Jean-Luc Rouas, Pierre Philip, and Jean-Arthur Micoulaud-Franchi. 2023b. Sleepiness in adults: An umbrella review of a complex construct. *Sleep Medicine Reviews*, 67:101718.

Vincent P. Martin, Jean-Luc Rouas, Jean-Arthur Micoulaud-Franchi, and Pierre Philip. 2020. The Objective and Subjective Sleepiness Voice Corpora. In *Language Resources and Evaluation Conference 2020*, pages 6525–6533, Marseille, France.

Vincent P. Martin, Jean-Luc Rouas, Jean-Arthur Micoulaud-Franchi, Pierre Philip, and Jarek Krajewski. 2021. How to Design a Relevant Corpus for Sleepiness Detection Through Voice? *Frontiers in Digital Health*, 3:686068.

Vincent P. Martin, Jean-Luc Rouas, and Pierre Philip. 2024b. Automatic detection of sleepiness-related symptoms and syndromes using voice and speech biomarkers. *Biomedical Signal Processing and Control*, 91:105989.

Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Interspeech 2017*, pages 498–502. ISCA.

Vikram Ramanarayanan, Adam C. Lammert, Hannah P. Rowe, Thomas F. Quatieri, and Jordan R. Green. 2022. Speech as a Biomarker: Opportunities, Interpretability, and Challenges. *Perspectives of the ASHA Special Interest Groups*, 7(1):276–283.

Allan Rechtschaffen and A Kales. 1968. *A Manual for Standardized Terminology, Techniques and Scoring System for Sleep Stages in Human Subjects*, ucla edition. Los Angeles: BIS/BRI.

Björn Schuller, Anton Batliner, Christian Bergler, Florian B. Pokorny, Jarek Krajewski, Margaret Cychocz, Ralf Vollman, Sonja-Dana Roelen, Sebastian Schnieder, Elika Bergelson, Alejandrina Cristia, Amanda Seidl, Anne Warlaumont, Lisa Yankowitz, Elmar Nöth, Shahin Amiriparian, Simone Hantke, and Maximilian Schmitt. 2019. The INTERSPEECH 2019 Computational Paralinguistics Challenge: Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity. In *Interspeech 2019*.

Björn Schuller, Stefan Steidl, Anton Batliner, Florian Schiel, and Jarek Krajewski. 2011. The INTERSPEECH 2011 Speaker State Challenge. In *Interspeech 2011*, pages 3201–3204.

Björn Schuller, Stefan Steidl, Anton Batliner, Florian Schiel, Jarek Krajewski, Felix Weninger, and Florian Eyben. 2013. Medium-term speaker states-A review on intoxication, sleepiness and the first challenge. *Comput. Speech Lang.*, 28(2):346–374.

Alexander J. Scott, Thomas L. Webb, Marrissa Martyn-St James, Georgina Rowse, and Scott Weich. 2021. Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials. *Sleep Medicine Reviews*, 60:101556.

Jianhua Shen, Joseph Barbera, and Colin M. Shapiro. 2006. Distinguishing sleepiness and fatigue: Focus on definition and measurement. *Sleep Medicine Reviews*, 10(1):63–76.

Etienne Thoret, Thomas Andrillon, Caroline Gauriau, Damien Léger, and Daniel Pressnitzer. 2024. Sleep deprivation detected by voice analysis. *PLOS Computational Biology*, 20(2):e1011849.

Bang Tran, Youxiang Zhu, Xiaohui Liang, James W. Schwoebel, and Lindsay A. Warrenburg. 2022. Speech Tasks Relevant to Sleepiness Determined With Deep Transfer Learning. In *ICASSP 2022*, pages 6937–6941.

Valerie S. L. Williams, Robert J. Morlock, and Douglas Feltner. 2010. Psychometric evaluation of a visual analog scale for the assessment of anxiety. *Health and Quality of Life Outcomes*, 8:57.

Dong Yang, Tomoki Koriyama, and Yuki Saito. 2024. Frame-Wise Breath Detection with Self-Training: An Exploration of Enhancing Breath Naturalness in Text-to-Speech. In *Interspeech 2024*, pages 4928–4932. ISCA.

A. Automatic phoneme recognition and segmentation

This transcription was obtained using a character-based conformer end-to-end ASR system trained on the French ESTER corpus (Galliano et al., 2009). This model achieves a Word Error Rate (WER) of 14.6% and a Character Error Rate (CER) of 5.7% on the ESTER test set. The whole pipeline (ASR+MFA) achieves a Phoneme Error Rate of 10.1% and an F-score of 0.72 on the planned subcorpus of Rhapsodie (Lacheret-Dujour et al., 2019). We refer the reader to (Martin et al., 2024a) for more details about the method.