Medispeech: a French Reading and Spontaneous Speech Corpus for Sleepiness Estimation

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

Abstract

Excessive Daytime Sleepiness (EDS) is associated with several diseases and therefore negatively affects the daily life of impacted people. Its diagnosis and follow-up are difficult because they require testing at the hospital for one full day. Monitoring patients regularly in ecological conditions may be done through speech analysis. Although several corpora containing speech from sleepy subjects exist, they do not suit ecological requirements regarding either the device used for recording or the speech elicitation tasks. In this paper, we introduce the Medispeech corpus containing reading, daily-life semi-spontaneous, and medically-oriented spontaneous tasks. Fifty-nine French subjects were recorded with both a professional-quality microphone and a smartphone using a dedicated application, resulting in 1,729 recordings for a total duration of 21 hours. Their EDS diagnosis was assessed by both a physiological objective measurement (mean sleep latency measured during a clinical test) and a subjective questionnaire (Karolinska Sleepiness Scale). Phenotyping of subjects is assured by collecting socio-demographic and medical data related to diverse dimensions of sleepiness, comorbidities, and addictions. Finally, we analyse the validity of our data collection protocol by measuring the effective duration of speech (after discarding pauses) and assessing its links with the collected subjects' characteristics.

Keywords: excessive sleepiness, ecological monitoring, spontaneous speech

1. Introduction

Excessive Daytime Sleepiness (EDS) is associated with several diseases (e.g. neurological or cardiovascular disorders) and affects the daily life of affected people. It also increases the risk of mortality for vehicle accidents (Bioulac et al., 2017) and of disability (Jike et al., 2018), which are public health concerns. The diagnosis and follow-up of EDS symptoms is difficult given that tests are undergone during one full day at the hospital (Sateia, 2014), which is both time-consuming for the subject and costly for the hospital. Thus, clinicians need a tool to collect those symptoms regularly in ecological conditions. Such a tool could also broaden the diagnostic reach to the general population, thereby providing more people with access to appropriate medical care.

Collecting the necessary data can be partially achieved using Ecological Momentary Assessment through questionnaires. However, gathering additional information confirming the answers to questionnaires is critical. To this end, researchers have been studying speech recordings to detect sleep-related conditions (e.g. (Thoret et al., 2024; Zhang et al., 2024)). Speech is indeed a good candidate to monitor EDS for two main reasons: its low-cost measurement characteristics, and the ease of data collection via smartphones in the subject's usual living environment.

Hence, we aim at studying if spontaneous speech recorded on smartphones can be used to estimate the sleepiness level in the same way as it can be measured during clinical interviews. We focus on spontaneous speech since it is the most commonly used style of speech in everyday life; but also because previous works have shown it to be more useful than read speech for detecting depression, due to its higher inherent variability (Alghowinem et al., 2013).

To do so, we need a corpus satisfying four criteria. First, the corpus must contain a spontaneous speech task since we want to study this particular type of speech. Second, the spontaneous task have to be recorded at least with a smartphone to respect the ecological condition we aim for. Thirdly, the used sleepiness measurement must be medically validated and specific to this construct. To this end, collaboration with somnologists is required. Finally, the participants must be French native speakers, since we work we work with French somnologists and we want to discriminate speech alterations due to sleepiness and these due to the linguistic background.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 5737–5748
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

In the following section, we review existing corpora, and discuss their limitations with respect to our objectives (Section 2). Section 3 explains the corpus methodological choices, with inclusion and exclusion criteria and its integration with the clinical EDS assessment. Phenotyping of subjects, including socio-demographic and medical data are presented in Section 4. The different speech tasks, protocol and duration analysis are presented in Section 5. Section 6 concludes about the Medispeech corpus.

2. Existing corpora

The Sleepy Language Corpus (SLC) and the SLEEP corpus provided during Interspeech challenges (Schuller et al., 2011, 2019) comprise read speech and (semi-)spontaneous tasks. The SLC was recorded with a microphone, while the SLEEP corpus was recorded on a computer with a microphone on a headset. They are both in German, and their sleepiness measure is not medically validated (Martin et al., 2021, 2023).

The Voiceome dataset (Tran et al., 2022) contains several read and (semi-)spontaneous speech tasks from 6,650 English participants, recorded with smartphones. However, sleepiness was assessed with the Stanford Sleepiness Scale (SSS) (MacLean et al., 1992) which cannot distinguish sleepiness from fatigue: it is thus not specific enough.

The EmoV_DB (Moon et al., 2022) is a database about emotion detection, "sleepy" being among them. 4 actors were recorded for a total of 1,744 English sentences. Similarly, the Thorsten-voice dataset (Müller and Kreutz, 2021) contains 300 sentences recorded by an actor who acted sleepy. There are several issues with both corpora: first, sleepiness is acted out and not spontaneous; second, both were recorded with a microphone; third, they don't have sleepiness measurement since sleepiness is acted out; finally, they are not in our target language.

Thoret et al. (2024) recorded 22 healthy women with a microphone reading different chapters from a French novel ("Le Comte de Monte-Cristo") during 10 minutes before and after sleep deprivation. Their sleepiness was evaluated with the SSS, which is not specific enough for proper sleepiness estimation. Thus, we cannot use this corpus despite it being in our target language.

Finally, the Multiple Sleep Latency Test corpus (MSLTc) (Martin et al., 2020) contains 660 recordings of 132 French hypersomniac subjects labelled with the Mean Sleep Latency (MSL), a reference measurement of sleepiness (Arand et al., 2005; Martin et al., 2023). Subjects were recorded with a professional-quality microphone reading aloud texts from *Le Petit Prince* (A. de Saint-Exupéry). While this corpus has allowed for obtaining promising results on the effect of sleepiness on read speech, it does not contain spontaneous speech recordings nor smartphone recordings.

Since no existing corpus fulfills all of our constraints, we created Medispeech: it is the first French corpus containing spontaneous and read speech, labelled with medically validated sleepiness measurements, and recorded with a professional-quality microphone and a smartphone. Our objective in this article is to present its methodology and to provide preliminary analyses.

3. Corpus design

Subjects were included to the Medispeech corpus under several criteria, and recorded with both a professional-quality microphone and a smartphone. We developed a smartphone application to automate speech data collection and to record self-reported sleepiness levels through a built-in questionnaire. The objective is to replicate as closely as possible how a participant would use the application in a home environment while undergoing the clinical assessment.

3.1. Inclusion & exclusion criteria

Subjects were included in this research based on 5 criteria. First, they have to be hospitalized for a Multiple Sleep Latency Test (MSLT) for the diagnosis or follow-up of Idiopathic Hypersomnia (IH). Second, they must be above 18 years old to be able to consent to participate in our study. Third, they must be French native speakers. Fourth, they need to have a sufficient reading proficiency to perform the reading-aloud task and understand written instructions. Finally, subjects must complete our study (no missing data). Any speech disorder is an exclusion criterion.

3.2. Integration into existing medical protocol

The Medispeech recordings took place before each session of the MSLT (Arand et al., 2005), the gold-standard test for diagnosing Idiopathic Hypersomnia. The MSLT consists of 5 sessions of 20 minutes, from 9 AM to 5 PM every 2 hours, after a normal night. While alone in a dark and quiet room of the sleep clinic, the subjects are asked to lie down on their bed and not fight sleep. The sleep latency, i.e. the duration between the beginning of each session and sleep onset, is measured objectively using electroencephalography (EEG). The 5 sleep latencies, one for each session, are averaged to measure the mean sleep latency (MSL). A MSL equal to or lower than 8 is a required criterion for the diagnosis of IH (Sateia, 2014).

Few minutes before each session of the MSLT, the subjects answered the French version of the Karolinska Sleepiness Scale questionnaire (KSS) (Åkerstedt and Gillberg, 1990), which measures their subjective instantaneous sleepiness on a Likert scale (from 1 “Extremely alert” to 9 “Very sleepy, great effort to keep awake, fighting sleep”). Then, they were recorded as they completed the tasks specified in the Medispeech protocol.

3.3. Recording protocol

The recordings were simultaneously captured using a smartphone and a professional-quality microphone plugged into a portable recorder. The smartphone is an entry-level Redmi Note 11, and was set to airplane mode to prevent any unintended data transfer. The microphone is a Beyerdynamic MPR 210, and the portable recorder is a TASCAM DR-40X.

The protocol is schematised in Figure 1. The subject was seated on their hospital bed and took the smartphone (the blue box “2” in the Figure) in their hand. It was placed around 15cm away from their face, depending on their grip. The subjects were autonomous in the handling of the smartphone. The microphone (the red box “1” in the Figure) was placed on a table in front of the subject, 20cm away from their mouth. The investigator was present in front of the subject across the table and manually activated the recorder. All the recordings were recorded in 48kHz 16-bit PCM WAV format, before being post-processed to mono 16kHz and normalised to -3 dB with Audacity. They were cut at the beginning to keep only the subjects’ voices if needed.

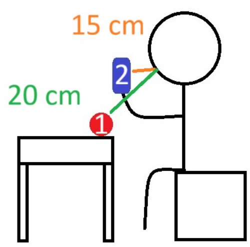

Figure 1: Recording protocol schema. The red circle “1” is the recorder with the microphone placed on a table and the blue box “2” is the smartphone.

3.4. Smartphone application

In order to automate data collection, we designed a smartphone application to carry out the protocol and to encourage the subjects to record themselves in complete autonomy. The application was created on Android Studio with Java. All tasks (KSS, reading, semi-, and spontaneous questions) were adapted to this device: the subjects were able to fill the KSS questionnaire, read the texts displayed on the smartphone screen, and use buttons to start and stop the recordings. Furthermore, during the semi-spontaneous speech tasks, a timer was displayed on the smartphone screen to ensure that participants spoke for more than 30 seconds. The audio files were stored locally during the session, transferred afterward to a secured computer, and deleted from the local device.

4. Socio-demographic and medical data

We completed the profile of the participants with socio-demographic data. Moreover, their medical phenotype was completed with questionnaires about other sleepiness dimensions, possible comorbidities, and addictions.

Table 1 reports the median values and their Interquartile Range (IQR) of socio-demographic and medical data, along with the statistical differences between IH patients (MSL$\leq$ 8 minutes) and non-IH ones (MSL$>$ 8 minutes), performed with χ^2 tests for categorical data, and Mann-Whitney tests (MW) for continuous ones.

4.1. Socio-demographic data

We collected subjects’ age, sex, height, weight, Body Mass Index (BMI), level of education and neck circumference – which is associated with obstructive sleep apnea (Davies and Stradling, 1990). The level of education corresponds to the number of years in school from the *Brevet des collèges*, a French diploma obtained at 15 years at the end of middle school.

We observed a difference in level of education (MW, $p < 0.05$) between the two diagnostic groups, patients fulfilling an IH diagnostic having a lower educational level, which is in line with recent sleep health research (Galušková et al.)

4.2. Medical data

The collected medical data are divided into 3 categories: sleepiness-related measurements, comorbidities questionnaires, and addiction-related questionnaires.

Sleepiness-related measurements The first category contains the clinical test and questionnaires measuring sleepiness. The MSLT and the KSS belong to this category.

	Range	Median value (IQR)		Diff.
		MSL≤ 8 minutes (n=13)	MSL> 8 minutes (n=46)	
Socio-demographic data				
Age (years)		26.0 (33.0)	32.5 (17.0)	
Sex		female=9 (69.2%) male=4 (30.8%)	female=34 (73.9%) male=12 (26.1%)	
Height (m)		1.7 (0.1)	1.7 (0.1)	
Weight (kg)		70.0 (39.0)	64.5 (17.0)	
BMI (kg/m^2)		25.6 (6.5)	23.3 (6.4)	
Level of education (years)		3.0 (2.0)	5.0 (3.5)	*
Neck circumference (cm)		39.8 (4.9)	37.0 (4.8)	
Medical data: Sleepiness				
Per session (n=290)				
SL (minutes)	[0-20]	6.0 (4.5)	14.8 (11.1)	***
KSS (Åkerstedt and Gillberg, 1990)	[1-9]	4.0 (2.0)	4.0 (3.0)	
PSS (Maldonado et al., 2004)	[0-4]	1.0 (1.0)	1.0 (1.0)	
Per patient (n=59)				
ESS (Johns, 1991)	[0-24]	16.0 (6.0)	13.0 (6.0)	
BSI (Guaita et al., 2015)	[0-6]	3.0 (1.0)	2.0 (2.0)	
Hobson Scale (Hobson et al., 2002)	[0-12]	6.0 (5.0)	3.0 (4.0)	
HSI (Fernandez-Mendoza et al., 2021)	[0-36]	21.0 (5.0)	21.0 (7.4)	
ISI (Bastien, 2001)	[0-36]	13.0 (3.0)	14.0 (8.0)	
FOSQ-10 (Chasens et al., 2009)	[10-40]	22.0 (7.0)	21.0 (8.8)	
BOSS (Philip et al., 2023)	[0-8]	3.0 (2.0)	2.0 (3.0)	
Medical data: Comorbidities				
FSS (Krupp, 1989)	[9-63]	49.5 (7.2)	51.0 (19.2)	
THAT (Shapiro et al., 2006)	[0-50]	19.0 (22.0)	21.0 (12.8)	
HAD-Anxiety	[0-21]	5.5 (7.0)	7.0 (5.8)	
HAD-Depression (Zigmond and Snaith, 1983)	[0-21]	5.0 (11.0)	5.5 (5.0)	
Medical data: Addictions				
CAGE (Dhalla and Kopec, 2007)	[0-4]	0.0 (0.0)	0.0 (0.0)	
Number of alcoholics drinks per day		0.0 (0.0)	0.0 (0.1)	
CDS5 (Fu et al., 2022)	[5-25]	5.0 (0.0)	5.0 (0.0)	
Number of cigarettes per day		0.0 (0.0)	0.0 (0.0)	

Table 1: Median values for socio-demographic and medical data with their range for Medispeech corpus for IH (MSL≤ 8 minutes) and non-IH (MSL> 8 minutes) subjects. χ_2 and Mann-Whitney tests were performed on socio-demographic and medical data respectfully. Any statistical differences are reported in the "Diff." column. *: $p < 0.05$; ***: $p < 0.001$.
IQR: Interquartile Range. MSL: Mean Sleep Latency. SL: Sleep Latency. KSS: Karolinska Sleepiness Scale. PSS: Pictorial Sleepiness Scale Based on Cartoon Faces. ESS: Epworth Sleepiness Scale. BSI: Barcelona Sleepiness Index. HSI: Hypersomnia Severity Index. ISI: Insomnia Severity Index. FOSQ10: Functionnal Outcomes of Sleep Questionnaire. BOSS: Bordeaux Sleepiness Scale. FSS: Fatigue Severity Scale. THAT: Toronto Hospital Alertness Test. HAD: Hospital Anxiety and Depression scale. CAGE: Alcohol dependence to alcohol questionnaire. CDS5: Cigarette Dependence Scale-5.

We also collected, at each iteration of the MSLT, the Pictorial Sleepiness Scale Based on Cartoon Faces (PSS), which measures the subjective sleepiness as for the KSS, using faces pictures instead of a textual question, potentially allowing a different expression of subjective sleepiness.

Regarding long-term sleepiness, the Epworth Sleepiness Scale (ESS) is a very standard questionnaire in sleep clinic practice, as it measures a similar construct as the MSLT: the propensity to sleep during the day. Similarly, the Barcelona Sleepiness Index (BSI) is a very short questionnaire (2 items) which has been designed to have a strong external validity with the MSLT and the ESS. A variant exist for the ESS measuring the propensity to sleep in *active* situations: the Hobson scale.

With regards to clinical assessment, we collected the Hypersomnia Severity Index (HSI) and the Insomnia Severity Index (ISI), which measure respectively the severity of hypersomnia or insomnia, through their impact on several aspects on daily life.

Moreover, the Functional Outcomes of Sleep

Questionnaire (FOSQ10) evaluates the functional impact of sleepiness. The Bordeaux Sleepiness Scale (BOSS) evaluates more specifically the accidental risk while driving when sleepy.

Finally, we considered circadian rhythm by gathering specific questions from 3 different questionnaires (the Munich Chronotype Questionnaire, the Morningness-Eveningness Questionnaire and the Pittsburgh Sleep Quality Index) into one named "Sleeping Habits Questionnaires" (SHQ). Those questions being open-ended oriented, we thus cannot consider them for the statistical analysis.

We found significant differences between the two diagnosis groups for sleep latency (MW, $p < 0.001$, unsurprisingly). Surprisingly, we did not find any differences in the instantaneous sleepiness scores (KSS, PSS) nor regarding the functional impact of sleepiness on the daily life of participants (FOSQ-10, BOSS) or the measurements of sleep propensity (ESS, BDI, Hobson scale).

Comorbidities questionnaires We also collected measurements of subjects' comorbidities, such as the impairment caused by their fatigue (Fatigue Severity Scale - FSS), their ability to maintain alertness (Toronto Hospital Alertness Test), and their anxiety or depression status (HAD).

Previous studies have shown links between sleepiness behaviors and mental health (Baglioni et al., 2016; Barateau et al., 2025) but no significant differences between the two diagnosis groups were found in Medispeech corpus. This result is expected since the population we study report low depressive symptomatic levels: in the sleep clinics from where they come, patients are pre-screened and treated for their psychiatric comorbidities before undergoing the MSLT for a further investigation of potential idiopathic hypersomnia.

Addiction-related questionnaires The last category evaluates consumption (number per day) and addiction to alcohol (CAGE questionnaire) and cigarettes (CDS-5 questionnaire). Indeed, several studies showed a link between their consumption and either sleep troubles inducing EDS, either with EDS itself (Stein and Friedmann, 2005; Wetter and Young, 1994).

However we did not find any significant differences in the addiction questionnaires between IH and non-IH subjects. These results can be explained by the fact that while these addictions have a significant impact on sleep architecture (Gardiner et al., 2024), the recruited participants had consumptions of alcohol and cigarettes that could be considered as negligible for both groups.

5. Speech data

With the objective of comparing which speaking style would be more useful for ecological assessment of IH diagnosis, the Medispeech corpus contain the recordings of the participant on three tasks: the first consists in reading a text aloud, the second task is a semi-spontaneous speech task about daily questions while the third contains sleepiness-oriented questions designed to elicitate spontaneous speech. Each are described and analysed in the following subsections.

Table 2 shows the number of microphone and smartphone recordings with the total duration for all and each task. The reading aloud task is the one with the highest speech duration (6 hours) while the (semi-)spontaneous tasks each have a duration of only 2 hours.

Task	# recordings	Total duration
	MICROPHONE	
Reading	n=274	5 hrs 56 min 32 sec
Semi-Spont.	n=274	2 hrs 06 min 25 sec
Spont.	n=272	1 hrs 56 min 04 sec
All	n=820	9 hrs 52 min 02 sec
	SMARTPHONE	
Reading	n=271	5 hrs 51 min 22 sec
Semi-Spont.	n=272	2 hrs 04 min 41 sec
Spont.	n=272	1 hrs 54 min 49 sec
All	n=815	9 hrs 50 min 52 sec

Table 2: Number of microphone (up) and smartphone (below) recordings and their total duration for all and each task.

While we recorded the speech of the participants using both a professional-quality microphone and a smartphone, we focus in this paper on the recordings from the smartphone, which are closer to the final use of the application (ecological monitoring using a smartphone). The comparison between the two recording devices will be the focus of future works.

5.1. Validation of collected data

We validated the collected data in terms of audio duration and effective speech duration, to ensure that each task results into recordings of sufficient duration for voice analysis. For instance, in a previous article about pathological sleepiness, Martin et al. (2021) had found a minimal duration of 20 seconds to be able to extract stable speech features – which is required for them to be potential biomarkers.We thus have to check that our samples have a speech duration longer than this threshold.

We first extracted speech segments from the audio files using the Python library rVADfast (Tan et al., 2020)[1]. From this extraction, we computed the total duration and the effective speech duration (the sum of the duration of all the detected speech segments). The total duration and effective speech duration for each session according to the type of speech are reported in Table 3.

Then, for each type of speech, we further investigated if these durations are influenced by socio-demographic characteristics of the speakers using a *linear mixed-effect model* (LMM), with the session (1–6), the age, the sex (M or F) and the education level (edu., 0–9) as fixed effects, and a random intercept per participant (ID) to account for inter-individual variability and the repeated nature of the experiment:

$$duration \sim session + age + sex + edu.$$

5.2. Reading Task

The first task for each session is reading a text aloud. We included a reading task in order to collect planned and controlled speech. Collecting read speech will allow to compare our results with other research, for example using the MLSTc (Martin et al., 2020).

Method For each session, the participants read a text extracted from *Le Petit Prince* of Antoine de Saint-Exupéry. Each text has a slightly different number of words: 232 for the 1st; 238 for the 2nd; 219 for the 3rd; 230 for the 4th; 248 for the 5th.

Results The recordings of the reading tasks have a duration between 52.8s and 173.6s, with an effective speech duration between 48.0s and 139.8s. The median duration per text is between 70.5s and 77.4s, and the median effective speech duration by text is between 59.1s and 66.6s: these two metrics are largely above the 20s threshold (Table 3).

Regarding the effect of text and socio-demographic characteristics, we did not find any effect of the session and the age on the duration of the recordings ($p > 0.05$), but we found a positive association of the duration with sex ($\beta = 9,570, p < 0.05$): recordings made by men tend to have a shorter duration than women. Moreover, the level of education has a negative impact on the duration ($\beta = -1.769, p < 0.05$): the greater level of education of a participant, the shorter the duration of the recording.

We found similar effects of education level ($\beta = -1.531, p < 0.05$) on the effective speech duration, but we no longer observe any effect of sex ($p > 0.05$).

[1] github.com/zhenghuatan/rVADfast

The effect of the level of education on the median effective speech duration is clearly visible on on Figure 2.

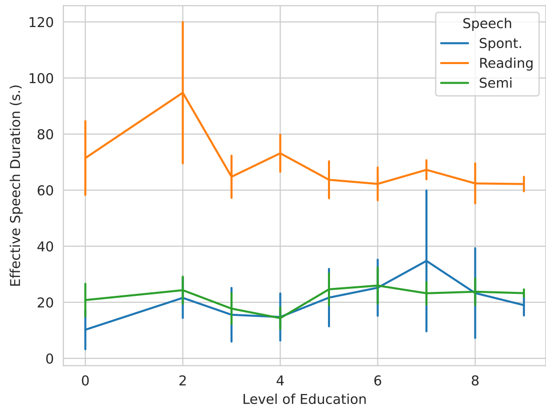

Figure 2: Median effective speech duration in seconds for each task according to level of education. Error bars represent standard error. *Spont.*: spontaneous speech. *Semi*: semi-spontaneous speech.

5.3. Semi-spontaneous task

After the reading task, the participants chose one question among the five proposed. While the reading task was first introduced to ease the speech recordings of subjects that may not be compliant to speak for more than 20 seconds, semi-spontaneous tasks offer a more spontaneous manner to collect speech.

Design According to our target population and our recording protocol, we set four constraints for the semi-spontaneous task: being implementable on a smartphone, a low-memory load to avoid bias between subjects, a speech duration of 20-30s (Martin et al.; Rutowski et al., 2019) and not inducing emotions to avoid changing participant's state.

We reviewed the speech tasks used in the literature and compared them with our constraints. Among them, there is the description of a picture (Albuquerque et al., 2021; Clarke et al., 2021) or a video (Tovar et al., 2020), interviews (Yamamoto et al., 2020; Aharonson et al., 2020; Ceccarelli and Mahmoud, 2021), a narrative (Figueroa Saavedra et al., 2021; Demiroglu et al., 2020; Ceccarelli and Mahmoud, 2021), free speech (Agurto et al., 2020) and questions about health (Tan et al., 2021; Lu et al.) or relatives (Schultebraucks et al., 2020). None of those fulfilled all our constraints.

We thus created 15 questions about daily aspects of life, that we evaluated with 38 native French speakers from the general population. We aimed

Task	Duration					Effective speech duration				
	Session					Session				
	1	2	3	4	5	1	2	3	4	5
Reading	76.6 (12.7)	77.8 (15.5)	70.9 (11.1)	74.7 (12.3)	77.4 (13.4)	65.2 (9.8)	66.7 (12.4)	59.3 (7.0)	62.3 (9.7)	65.6 (11.0)
Semi	28.8 (5.6)	29.2 (4.3)	28.0 (4.9)	28.5 (7.7)	28.1 (7.2)	24.3 (5.9)	24.0 (6.3)	22.6 (4.9)	22.4 (7.5)	22.9 (7.8)
Spont.	25.5 (22.5)	21.6 (12.7)	24.5 (16.8)	21.4 (13.2)	22.8 (16.0)	21.5 (17.5)	17.2 (10.9)	19.1 (13.6)	17.7 (12.3)	18.9 (13.6)

Table 3: Duration (left) and effective speech duration (right) median values (and IQR) for each task and each session.

at maximizing speech duration (automatically extracted using rVADfast (Tan et al., 2020)) and minimizing emotional arousal (automatically extracted using SpeechDimEmo (Evain et al., 2021)). Two tasks were excluded based on these criteria, the 13 other questions are reported in Table 4.

Method During each of the five recording sessions (corresponding to the five iterations of the MSLT), five questions were randomly selected from the remaining unanswered questions. Then, subjects picked the question they preferred, and were asked to answer step by step with as much detail as possible. Subjects were recorded with the instruction to talk for at least 30 seconds, and a timer was displayed on the smartphone screen to indicate the remaining time.

Results The median duration (respectively effective speech duration) of the semi-spontaneous task are reported in the second row, left side (resp. right side), of Table 3. The median duration and median effective speech duration of all the tasks are above 20 seconds, as for the reading aloud task. Regarding socio-demographic effects, the level of education is the only factor with a positive impact on speech duration ($\beta = 0.782, p < 0.05$) and effective speech duration ($\beta = 0.768, p < 0.01$): the greater level of education of a subject, the longer the duration of the answer and the duration of analysable speech. Moreover, the session have a negative impact for both speech duration and effective speech duration ($\beta = -0.339, p < 0.05$ and $\beta = -0.407, p < 0.01$). We did not find any significant effect of the sex or the age of the participants.

Interestingly, the effect of the level of education is opposite to the one found for the reading task: subjects with higher education level are faster for the reading task and longer for the semi-spontaneous task. The impact of the level of education on the median effective speech duration is displayed in Figure 2.

Choice of semi-spontaneous questions Since participants chose the questions they answered among a selection, we can further study which questions are most often chosen and if there is a link between questions and the measured speech durations.

Figure 3 shows the median duration (in blue) and the median effective speech duration (in orange) for each question. Questions are sorted by the number of answers, indicated below the question number. One question was answered by more than 50% of the subjects: Q7 with 30 answers. On the contrary, Q5 and Q1 were the ones with the least number of answers (6 and 8 respectfully).

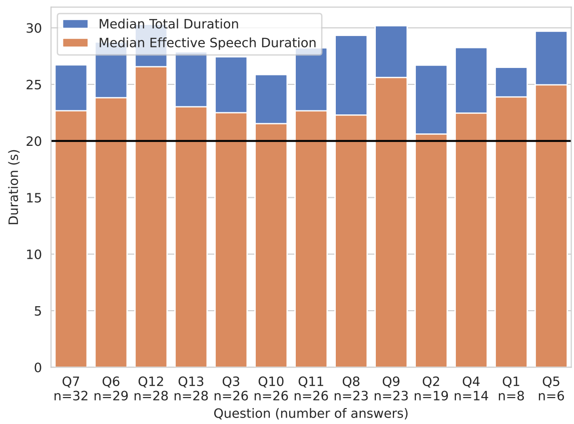

Figure 3: Median duration and median effective speech duration (s.) for each question of the semi-spontaneous task, sorted by number of answers. The black line is the 20 seconds threshold.

Regarding the duration of the recordings associated to each task, two Kruskal-Wallis tests indicated that all the 13 questions did not have the same duration ($\chi^2 = 37.19, p < 0.001, df = 12$) or effective speech duration ($\chi^2 = 39.44, p < 0.001, df = 12$). However, even if they have a different duration, all the questions result into median effective speech duration greater than 20 seconds: the 30 seconds timer displayed on the smartphone screen seems to be efficient for ensuring that the subject keep speaking long enough. Q2 is the question with the lowest median effective speech median (20.6s), and Q12 is the one with the highest value (26.8s).

It could be hypothesized that the most popular tasks are the one leading to greater duration or

Id.	Question
Q1	*Quelles sont les règles d'un jeu de votre choix (échecs, Uno…) ?* What are the rules of a game of your choice (chess, Uno, etc.)?
Q2	*Comment prendre un repas au restaurant ?* How to eat at a restaurant?
Q3	*Qu'est-ce qui se trouve dans votre champ de vision ?* What is in your field of vision?
Q4	*Comment vous préparez-vous à recevoir des invités chez vous ?* How do you prepare to welcome guests into your home?
Q5	*Quelles sont les règles d'un sport de votre choix (football, rugby…) ?* What are the rules of a sport of your choice (soccer, rugby, etc.)?
Q6	*Comment faites-vous vos courses ?* How do you do your shopping?
Q7	*Comment vous rendez-vous depuis votre domicile jusqu'au lieu dans lequel vous vous trouvez ?* How do you get from your home to the place where you are?
Q8	*De quoi a besoin une maison selon vous pour être habitable, et pourquoi ?* What do you think a house needs to be livable, and why?
Q9	*Quelle est la recette d'un plat de votre choix (crêpes…) ?* What is the recipe for a dish of your choice (crepes, etc.)?
Q10	*Que devez-vous faire avec un vêtement sale avant de le ranger ?* What should you do with dirty clothes before putting them away?
Q11	*Comment démarrer et conduire une voiture ?* How to start and drive a car?
Q12	*Quelle est votre journée type (travail, trajets, tâches ménagères…) ?* What is your typical day like (work, commuting, household chores, etc.)?
Q13	*Comment préparez-vous un voyage à l'étranger ?* How do you prepare for a trip abroad?

Table 4: Semi-spontaneous questions. Participants had access to the French version, the English version is only provided for this article.

effective speech duration, e.g. because the participants have a special affect with it. However, we did not find any correlation between the number of time each question has been chosen and the median duration (Spearman, $\rho = 0.05, p = 0.86$) or the median effective speech duration ($\rho = 0.04, p = 0.89$).

Since they all lead to effective speech duration higher than the 20s threshold and their popularity is independent from the resulting speech duration, we did not select a subset of questions among the 13 and would advice to keep a broad diversity of questions to pick among for the participants.

5.4. Spontaneous task

After the semi-spontaneous task of each session, the participants answered a spontaneous (free-speech) task in order to capture their behaviour during a task mimicking clinical interviews.

Method We designed five questions with the goal to capture spontaneous speech, one for each session (see Table 5). They were created in collaboration with a sleep medicine specialist, and they evaluate different aspects of EDS: the first question emphasizes the onset and the evolution of sleepiness; the second researches its determinants; the third its functional impact; the fourth explores countermeasures taken by subjects; the last one investigates the distress related to sleepiness. Subjects could take time to think before answering. Contrary to the previous task, we did not display any timer for a minimal duration.

Session	Question
1	*Depuis quand est apparue votre somnolence et quelle est son évolution ?* When did your drowsiness start, and how has it progressed?
2	*Quels sont les facteurs susceptibles d'influencer votre niveau de somnolence ?* Which factors are likely to influence your level of sleepiness?
3	*Dans quelle activité personnelle ou professionnelle êtes-vous particulièrement limité(e) par votre somnolence ?* In which personal or professional activity are you particularly limited by your sleepiness?
4	*Quelles stratégies mettez-vous en place pour limiter l'impact de la somnolence dans des situations qui requièrent une vigilance soutenue ?* What strategies do you implement to limit the impact of drowsiness in situations that require sustained alertness?
5	*Quel est l'impact de votre somnolence sur votre bien-être et plus largement sur votre santé ?* How does your sleepiness affect your well-being and, more broadly, your health?

Table 5: Medical questions related to EDS for the collection of spontaneous speech. Questions were asked in French, the English translation is only reported for this article.

Results The median duration (respectively median effective speech duration) for each spontaneous task are showed in the third row, left side (resp. right side), of Table 3. Contrary to the other tasks, the median duration are closer to 20 seconds (with wider IQR), and, except session 1, all median effective speech durations are below 20 seconds with wide IQR.

As for the semi-spontaneous task, the level of education is the only factor impacting positively the total duration ($\beta = 2.050, p < 0.01$) and the effective speech duration ($\beta = 1.754, p < 0.01$) while the session have a negative impact ($\beta = -0.573, p < 0.05$ and $\beta = -0.522, p < 0.05$): the higher the level of education, the longer the recordings and the more sessions, the shorter the recordings. The impact of the level of education on the median effective speech duration per session is displayed in Figure 2.

Without previous studies on the minimum effective speaking time for spontaneous speech tasks, it is difficult to conclude whether the tasks themselves led to shorter recordings, or whether it is the absence of a timer displayed on the screen that explains the difference between semi-spontaneous and spontaneous speech recording duration.

6. Conclusion

This paper introduced the Medispeech corpus, which contains reading, semi-spontaneous, and spontaneous recordings of patients at a sleep clinic. In order to collect the different speech styles, we chose relevant texts; we designed semi-spontaneous questions about the daily life of the participants; and we designed spontaneous questions about the impact of sleepiness on their daily life. The 59 participants were recorded with both a professional-quality microphone and a smartphone, with a dedicated application to simulate an ecological data collection at home (total duration: 19 hours and 49 minutes). We only focused on the recordings made by the smartphone in this paper, other works are ongoing to compare the acoustic quality of speech from both recording devices. Their sleepiness was assessed with both a subjective questionnaire (KSS) and the objective MSL measured during the MSLT, the gold-standard measurement in sleep medicine and a required test for the diagnosis of Idiopathic Hypersomnia. Additional information were also collected in order to phenotype the subjects as much as possible, including socio-demographics and medical questionnaires on sleepiness, comorbidities and addictions.

To validate our speech recording collection, we measured the duration and the effective speech duration (after applying a voice activity detector) of each task for each session, ensuring that the recordings contain enough speech to allow robust features extraction. Moreover, we performed statistical analysis to investigate if any of the socio-demographic data influence the recording durations. We found out that a higher level of education implies shorter durations for the reading task and longer ones for the (semi-)spontaneous tasks. Furthermore, we observed that the answers to spontaneous questions are much more variable in duration than the answers to semi-spontaneous ones. However, we could not conclude if this phenomenon is due to the nature of the questions or the fact that no minimal duration guidelines were given to the participants for the spontaneous task.

7. Perspectives

Other statistical analyses are foreseen, such as the relation between the sleepiness measurements and the vocal characteristics extracted for each task. In addition to the MSLT, subjects may take another test, the Maintenance of Wakefulness Test (MWT). It measures the subject's ability to stay awake in conditions conducive to falling asleep. Recording subjects undergoing this test would allow comparison between the two tests (MSLT-MWT) to find vocal characteristics specific to each one. Moreover, subjects can undergo two MWT, the first to assess IH, and the second to evaluate the treatment they were given. Both recordings of the same subject could be compared to assess if the treatment changed their vocal characteristics.

8. Ethical statement

The protocol has been evaluated as 'not involving the human person' by the ethics committee of the French National Centre for Scientific Research (CNRS) (MR004 protocol). It is being carried out in accordance with the GDPR regulation, and has been registered and authorized by the DPO of the University of Bordeaux ("Medispeech" study). Due to the sensitive nature of the data contained in the Medispeech corpus, we are not allowed to share it with the community.

9. Acknowledgments

This project has received financial support from the CNRS through the MITI interdisciplinary programs through its exploratory research program (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.

10. Bibliographical References

Carla Agurto, Guillermo A. Cecchi, Raquel Norel, Rachel Ostrand, Matthew Kirkpatrick, Matthew J. Baggott, Margaret C. Wardle, Harriet de Wit, and Gillinder Bedi. 2020. Detection of acute 3,4-methylenedioxymethamphetamine (MDMA) effects across protocols using automated natural language processing. *Neuropsychopharmacol.*, 45(5):823–832.

Vered Aharonson, Alexandra de Nooy, Seth Bulkin, and Gareth Sessel. 2020. Automated Classification of Depression Severity Using Speech - A Comparison of Two Machine Learning Architectures. In *ICHI*, pages 1–4.

Torbjörn Åkerstedt and Mats Gillberg. 1990. Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, 52(1-2):29–37.

Luciana Albuquerque, Ana Rita S. Valente, António Teixeira, Daniela Figueiredo, Pedro Sa-Couto, and Catarina Oliveira. 2021. Association between acoustic speech features and non-severe levels of anxiety and depression symptoms across lifespan. *PLOS ONE*, 16(4):e0248842.

Sharifa Alghowinem, Roland Goecke, Michael Wagner, Julien Epps, Michael Breakspear, and Gordon Parker. 2013. Detecting depression: A comparison between spontaneous and read speech. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7547–7551, Vancouver, BC, Canada. IEEE.

Donna Arand, Michael Bonnet, Thomas Hurwitz, Merrill Mitler, Roger Rosa, and R. Bart Sangal. 2005. The Clinical Use of the MSLT and MWT. *Sleep*, 28(1):123–144.

Chiara Baglioni, Svetoslava Nanovska, Wolfram Regen, Kai Spiegelhalder, Bernd Feige, Christoph Nissen, Charles F. Reynolds, and Dieter Riemann. 2016. Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychological bulletin*, 142(9):969–990.

Lucie Barateau, Marie-Gwenaëlle Paryl, Sofiene Chenini, Carole Pesenti, Lily Guiraud, Séverine Béziat, Isabelle Jaussent, and Yves Dauvilliers. 2025. Depressive symptoms and suicidal thoughts in idiopathic hypersomnia: a cross-sectional and longitudinal study. *Sleep*, page zsaf215.

C Bastien. 2001. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, 2(4):297–307.

Stéphanie Bioulac, Jean-Arthur Micoulaud-Franchi, Mickael Arnaud, Patricia Sagaspe, Nicholas Moore, Francesco Salvo, and Pierre Philip. 2017. Risk of Motor Vehicle Accidents Related to Sleepiness at the Wheel: A Systematic Review and Meta-Analysis. *Sleep*, 40(10).

Francesco Ceccarelli and Marwa Mahmoud. 2021. Multimodal temporal machine learning for Bipolar Disorder and Depression Recognition. *Pattern. Anal. Appl.*

Eileen R. Chasens, Sarah J. Ratcliffe, and Terri E. Weaver. 2009. Development of the FOSQ-10: A Short Version of the Functional Outcomes of Sleep Questionnaire. *Sleep*, 32(7):915–919.

N. Clarke, T.R. Barrick, and P. Garrard. 2021. A Comparison of Connected Speech Tasks for Detecting Early Alzheimer's Disease and Mild Cognitive Impairment Using Natural Language Processing and Machine Learning. *Front. Comput. Sci.*, 3.

R. J. Davies and J. R. Stradling. 1990. The relationship between neck circumference, radiographic pharyngeal anatomy, and the obstructive sleep apnoea syndrome. *The European Respiratory Journal*, 3(5):509–514.

Cenk Demiroglu, Asli Besirli, Yasin Ozkanca, and Selime Çelik. 2020. Depression level assessment from multi-lingual conversational speech data using acoustic and text features. *JASM*, 2020(1):17.

Shayesta Dhalla and Jacek A. Kopec. 2007. The CAGE Questionnaire for Alcohol Misuse: A Review of Reliability and Validity Studies. *Clinical and Investigative Medicine*, 30(1):33–41.

Solène Evain, Manh Ha Nguyen, Hang Le, Marcely Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, and Laurent Besacier. 2021. Task Agnostic and Task Specific Self-Supervised Learning from Speech with LeBenchmark.

Julio Fernandez-Mendoza, Kristina Puzino, Gregory Amatrudo, Elizaveta Bourchtein, Susan L. Calhoun, David T. Plante, and Kate Kaplan. 2021. The Hypersomnia Severity Index: Reliability, construct, and criterion validity in a clinical sample of patients with sleep disorders. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 17(11):2249–2256.

Carla Figueroa Saavedra, Tamara Otzen Hernández, Camila Alarcón Godoy, Arlette Ríos Pérez, Damaris Frugone Salinas, and Roberto Lagos Hernández. 2021. Association between suicidal ideation and acoustic parameters of university students' voice and speech: a pilot study. *Logop. Phoniatr.*, 46(2):55–62.

Marcela Fu, Dolors Carnicer-Pont, Yolanda Castellano, Montse Ballbè, Xisca Sureda, Antònia Raich, Jose Martínez-Sánchez, Cristina Martínez, Antoni Baena, Anna Riccobene, Marta Enríquez, and Esteve Fernández. 2022. Measuring cigarette dependence: A comparison of two scales in two different groups of smokers. *Tobacco Prevention & Cessation*, 8(April):1–8.

K. Galušková, S. Nevšímalová, I. Příhodová, S. Dostálová, A. Mazouchová, and K. Šonka. Quality of life and its associates in narcolepsy 1 and 2 types and idiopathic hypersomnia. 129:31–39.

C. Gardiner, J. Weakley, Louise M. Burke, G. Roach, C. Sargent, N. Maniar, Minh Huynh, Dean J. Miller, A. Townshend, and S. Halson. 2024. The effect of alcohol on subsequent sleep in healthy adults: A systematic review and meta-analysis. *Sleep medicine reviews*, 80:102030.

Marc Guaita, Manel Salamero, Isabel Vilaseca, Alex Iranzo, Josep M. Montserrat, Carles Gaig, Cristina Embid, Montserrat Romero, Mònica Serradell, Carme León, Joan De Pablo, and Joan Santamaria. 2015. The Barcelona Sleepiness Index: A New Instrument to Assess Excessive Daytime Sleepiness in Sleep Disordered Breathing. *Journal of Clinical Sleep Medicine*, 11(11):1289–1298.

Douglas E. Hobson, Anthony E. Lang, W. R. Wayne Martin, Ajmal Razmy, Jean Rivest, and Jonathan Fleming. 2002. Excessive Daytime Sleepiness and Sudden-Onset Sleep in Parkinson Disease: A Survey by the Canadian Movement Disorders Group. *JAMA*, 287(4):455.

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.

Murray W. Johns. 1991. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, 14(6):540–545.

Lauren B. Krupp. 1989. The Fatigue Severity Scale: Application to Patients With Multiple Sclerosis and Systemic Lupus Erythematosus. *Archives of Neurology*, 46(10):1121.

Y. Lu, A. Harati, T. Rutowski, R. Oliveira, P. Chlebek, and E. Shriberg. Robust Speech and Natural Language Processing Models for Depression Screening. In *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–5. IEEE.

A. W. MacLean, G. C. Fekken, P. Saskin, and J. B. Knowles. 1992. Psychometric evaluation of the Stanford Sleepiness Scale. *Journal of Sleep Research*, 1(1):35–39.

Claudia C. Maldonado, Alison J. Bentley, and Duncan Mitchell. 2004. A Pictorial Sleepiness Scale Based on Cartoon Faces. *Sleep*, 27(3):541–548.

Vincent P. Martin, Régis Lopez, Yves Dauvilliers, Jean-Luc Rouas, Pierre Philip, and Jean-Arthur Micoulaud-Franchi. 2023. 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 *12th Edition of Its Language Resources and Evaluation Conference.*, 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. Détection de la somnolence dans la voix : nouveaux marqueurs et nouvelles stratégies.

Jihye Moon, Youngsun Kong, and Ki H. Chon. 2022. Language-Independent Sleepy Speech Detection. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1981–1984, Glasgow, Scotland, United Kingdom. IEEE.

Thorsten Müller and Dominik Kreutz. 2021. Thorsten-Voice Dataset 2021.06 emotional.

Pierre Philip, Jean-Arthur Micoulaud-Franchi, Jacques Taillard, Julien Coelho, Camille Tisserand, Yves Dauvilliers, and Patricia Sagaspe. 2023. The Bordeaux Sleepiness Scale (BOSS): A new questionnaire to measure sleep-related driving risk. *Journal of Clinical Sleep Medicine*, 19(5):957–965.

Tomasz Rutowski, Amir Harati, Yang Lu, and Elizabeth Shriberg. 2019. Optimizing Speech-Input Length for Speaker-Independent Depression Classification. In *Interspeech 2019*, pages 3023–3027. ISCA.

Michael J Sateia. 2014. International Classification of Sleep Disorders-Third Edition. *Chest*, 146(5):1387–1394.

Björn Schuller, Stefan Steidl, Anton Batliner, Florian Schiel, and Jarek Krajewski. 2011. The INTERSPEECH 2011 speaker state challenge. In *Interspeech 2011*, pages 3201–3204. ISCA.

Björn W. Schuller, Anton Batliner, Christian Bergler, Florian B. Pokorny, Jarek Krajewski, Margaret Cychosz, Ralf Vollmann, Sonja-Dana Roelen, Sebastian Schnieder, Elika Bergelson, Alejandrina Cristia, Amanda Seidl, Anne S. 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*, pages 2378–2382. ISCA.

Katharina Schultebraucks, Vijay Yadav, Arieh Y. Shalev, George A. Bonanno, and Isaac R. Galatzer-Levy. 2020. Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. *Psychol. Med.*, pages 1–11.

Colin M. Shapiro, Christine Auch, Marlene Reimer, Leonid Kayumov, Ronald Heslegrave, Nada Huterer, Helen Driver, and Gerald M. Devins. 2006. A new approach to the construct of alertness. *Journal of Psychosomatic Research*, 60(6):595–603.

Michael D. Stein and Peter D. Friedmann. 2005. Disturbed sleep and its relationship to alcohol use. *Substance Abuse*, 26(1):1–13.

Eric J. Tan, Denny Meyer, Erica Neill, and Susan L. Rossell. 2021. Investigating the diagnostic utility of speech patterns in schizophrenia and their symptom associations. *Schizophr*, 238:91–98.

Zheng-Hua Tan, Achintya Kr. Sarkar, and Najim Dehak. 2020. rVAD: An unsupervised segment-based robust voice activity detection method. *Computer Speech & Language*, 59:1–21.

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

Antonia Tovar, Aina Garí Soler, Jesús Ruiz-Idiago, Celia Mareca Viladrich, Edith Pomarol-Clotet, Joana Rosselló, and Wolfram Hinzen. 2020. Language disintegration in spontaneous speech in Huntington's disease: a more fine-grained analysis. *J. Commun. Disord.Disorders*, 83:105970.

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 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6937–6941, Singapore, Singapore. IEEE.

D.W. Wetter and T.B. Young. 1994. The Relation Between Cigarette Smoking and Sleep Disturbance. *Preventive Medicine*, 23(3):328–334.

Mao Yamamoto, Akihiro Takamiya, Kyosuke Sawada, Michitaka Yoshimura, Momoko Kitazawa, Kuo-Ching Liang, Takanori Fujita, Masaru Mimura, and Taishiro Kishimoto. 2020. Using speech recognition technology to investigate the association between timing-related speech features and depression severity. *PloS One*, 15(9):e0238726.

Boyu Zhang, Joseph Ronda, Robin Yuan, Jeanne Duffy, and Charles Czeisler. 2024. 0292 Detecting Sleep Deficiency with Voice Biomarkers and Machine Learning. *Sleep*, 47(Supplement_1):A126.

A. S. Zigmond and R. P. Snaith. 1983. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 67(6):361–370.