Actes de JEP-TALN-RECITAL 2024
Toulouse, France, 8 au 12 juillet 2024

Comment l'oreille humaine perçoit-elle la somnolence dans la parole ? Une analyse rétrospective d'études perceptuelles.

Vincent P. Martin[1] Colleen Beaumard[2,3] Jean-Luc Rouas[2]
(1) Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health, Strassen, L-1445, Luxembourg
(2) Univ. Bordeaux, LaBRI, CNRS UMR 5800, Bordeaux INP, Talence, F-33405, France
(3)Univ. Bordeaux, CNRS, SANPSY, UMR 6033, Bordeaux, F-33000, France
vincentp.martin@lih.lu, {colleen.beaumard, jean-luc.rouas}@labri.fr

RÉSUMÉ

La somnolence bénéficierait d'être mesurée dans des configurations écologiques, par exemple grâce à des enregistrements de parole. Pour évaluer la faisabilité de sa détection à partir de la parole par l'audition humaine, deux études perceptuelles précédentes ont produit des résultats contradictoires. Une façon de comprendre ce désaccord aurait pu être d'étudier sur quelles caractéristiques de la parole les annotateurs ont basé leur estimation, mais aucune étude n'a collecté cette information. Nous avons donc choisi d'extraire des descripteurs acoustiques des enregistrements annotés, et d'entraîner des modèles d'apprentissage automatique simples et explicables à reproduire l'annotation de chaque annotateur. Ensuite, nous mesurons la contribution de chaque caractéristique à la décision de chaque modèle, et identifions les plus importantes. Nous effectuons ensuite un regroupement hiérarchique pour dessiner les profils des annotateurs, en fonction des caractéristiques sur lesquelles ils s'appuient pour identifier la somnolence.

ABSTRACT

How does human hearing perceive sleepiness from speech ? A retrospective analysis of perceptual experiments

Excessive sleepiness would benefit from being measured in ecological settings, for example through speech recordings. To assess the feasibility of detecting sleepiness from speech by human hearing, two previous perceptual studies have yielded contradictory results. One way to investigate this disagreement would have been to look into which speech characteristics listeners based their estimation, but no study has collected this information. In this study, we extract acoustic descriptors from annotated recordings, and train simple and explainable machine learning models to reproduce the annotation of each annotator. Then, we measure the contribution of each feature to each model's decision, and identify the most important ones. We then perform hierarchical clustering to draw profiles of listeners, based on the features they rely on to identify sleepiness.

MOTS-CLÉS : Études perceptuelles, Somnolence, Modèle interprétable.

KEYWORDS: Perceptual studies, Sleepiness, Interpretable model.

Cet article est une traduction en français d'un article en cours d'évaluation par les pairs pour la conférence internationale *Speech Prosody 2024*. Il traite des éléments de la parole sur laquelle se sont basés les annotateurs d'une étude perceptuelle, que nous avions recrutés en 2021. Tous étaient francophones. Il nous semble important de promouvoir ce travail en français pour permettre un retour vers les personnes qui ont participé à l'étude (qui ne lisent en général pas l'anglais) et avoir des retours

Actes des 35ème Journées d'Études sur la Parole, pages 23–32.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

scientifiques de la communauté de recherche en parole sur la perception de la parole en français.

1 Introduction

Contexte. L'hypersomnie est un fardeau majeur à la fois pour la santé publique (Léger *et al.*, 2012; Barnes & Watson, 2019) et la santé personnelle, en lien avec des troubles métaboliques, cardiovasculaires, neurologiques et psychiatriques, augmentant le risque d'invalidité et de mortalité (Jike *et al.*, 2018; Scott *et al.*, 2021). En raison de sa forte prévalence dans la population générale (jusqu'à une personne sur trois (Kolla *et al.*, 2020)), les cliniciens ont besoin d'outils pour mesurer le niveau de somnolence de leurs patients aussi régulièrement que possible, dans des conditions écologiques (par exemple, à domicile), de manière passive (c'est-à-dire sans tâche dédiée). À cet égard, les enregistrements vocaux et de parole sont un candidat de choix : leur collecte est implémentée dans tous les smartphones, ils peuvent être enregistrés dans des configurations passives, et ils ont déjà été liés à de multiples troubles (Fagherazzi *et al.*, 2021), y compris la somnolence.

Précédents travaux. En effet, la détection de la somnolence à l'aide d'enregistrements vocaux a déjà été au centre de deux challenges Interspeech en 2011 et 2019, reposant respectivement sur le *Sleep Language Corpus* (SLC) (Schuller *et al.*, 2011) et le corpus SLEEP (Schuller *et al.*, 2019). Les deux corpus sont étiquetés avec une mesure subjective (questionnaire d'auto-évaluation) (Martin *et al.*, 2021), l'échelle de somnolence de Karolinska (KSS) (Åkerstedt & Gillberg, 1990).

Le meilleur système du challenge Interspeech 2011 a atteint un score de Rappel Moyen Non Pondéré (*Unweighed Average Recall*, UAR) de 71.7% (Huang *et al.*, 2011) sur la classification binaire de la somnolence. Sur le corpus SLEEP, la tâche du défi Interspeech 2019 était d'estimer le degré de somnolence. Les gagnants du challenge ont atteint une corrélation de Spearman $\rho = 0.387$ entre l'estimation produite par leur système et la vérité terrain (Gosztolya, 2019). Cette approche simple n'a jamais été surpassée dans des approches plus récentes utilisant les dernières techniques d'apprentissage profond (par exemple, $\rho = 0.325$ dans (Fritsch *et al.*, 2020), $\rho = 0.367$ dans (Amiriparian *et al.*, 2020), $\rho = 0.365$ dans (Egas-López *et al.*, 2022) ou $\rho = 0.383$ dans (Campbell *et al.*, 2022)).

Plus récemment, un nouveau grand corpus enregistré dans des conditions écologiques à l'aide de smartphones a été introduit : le corpus Voiceome (Tran *et al.*, 2022). L'équipe ayant développé ce corpus a rapporté un score F1 de 81.3% sur la classification binaire de la somnolence, mesurée par l'échelle de somnolence de Stanford (Hoddes *et al.*, 1973).

Parallèlement à ce travail se concentrant sur la somnolence à court terme, il est notable de mentionner nos précédents travaux sur le *Multiple Sleep Latency Test corpus* (MSLTc), contenant des enregistrements vocaux de patients hypersomniaques de la clinique du sommeil du CHU de Bordeaux étiquetés avec à la fois la somnolence subjective à court et à long terme (questionnaires) et physiologique (latence de sommeil mesurée par électroencéphalographie). Avec ces données, nous avons atteint des scores d'UAR supérieurs à 75% sur la détection de trois symptômes liés à la somnolence dans cette population (Martin *et al.*, 2024).

Limites. Au cours de la dernière décennie, la plupart des recherches utilisant ces corpus se sont concentrées sur le développement d'algorithmes d'apprentissage automatique pour estimer la

somnolence à partir des enregistrements de parole contenus dans ces corpus. En revanche, très peu d'attention a été accordée à l'élucidation du lien entre la somnolence et le comportement vocal. Depuis le travail fondateur de Krajewski *et al.* (2009), très peu d'études ont cherché à clarifier les mécanismes sous-jacents à l'expression de la somnolence dans la parole.

Parallèlement à ce travail d'apprentissage automatique, deux études perceptuelles ont récemment été menées sur le corpus SLEEP pour déterminer si l'oreille humaine peut estimer la somnolence à partir d'échantillons de parole. Ces deux études, basées sur 99 échantillons du corpus SLEEP, ont produit des résultats contradictoires : l'étude de Huckvale et al., impliquant 26 annotateurs, a conclu qu'il était possible de reconnaître la somnolence dans les enregistrements du corpus (Huckvale *et al.*, 2020). En revanche, notre étude de réplication, basée sur les annotations de 30 annotateurs naïfs, a obtenu des résultats moins enthousiastes (Martin *et al.*, 2023c). Puisque les participants de l'étude menée par Huckvale et al. étaient anglophones natifs et ceux de notre étude de réplication parlaient français, une explication à cette divergence entre les études aurait pu être expliquée par des différences dans les caractéristiques de la parole utilisées par les annotateurs pour estimer la somnolence, mais aucune de ces études n'a collecté de tels retours.

Objectif. L'objectif de cet article est d'identifier, a posteriori, les caractéristiques de la parole sur lesquelles les annotateurs se sont appuyés pour identifier la somnolence en réanalysant les données des deux études perceptuelles précédentes sur le corpus SLEEP (Huckvale *et al.*, 2020; Martin *et al.*, 2023c). Pour ce faire, sur la base d'un ensemble minimal de descripteurs extraits des enregistrements audio, nous avons entraîné plusieurs systèmes d'apprentissage automatique pour reproduire les annotations (un système d'apprentissage automatique de "clonage" par annotateur). Les caractéristiques extraites et le système d'apprentissage automatique ont été choisis simples et parfaitement explicables, permettant l'extraction et l'interprétation de l'importance relative de chaque caractéristique de parole dans l'imitation des annotateurs. Cette technique nous permet de dresser des profils d'annotateurs, déterminés en fonction de la manière dont ils identifient la somnolence à partir des enregistrements vocaux.

2 Méthode

Un aperçu de notre méthode est représenté dans la Figure 1.

Corpus et échantillons audio. Nous nous concentrons dans cet article sur les deux études perceptuelles impliquant le corpus SLEEP (Huckvale *et al.*, 2020; Martin *et al.*, 2023c). Le corpus entier contient plus de 16.464 échantillons de 915 sujets germanophones, enregistrés sur différentes tâches inconnues (Martin *et al.*, 2021). Tous les échantillons sont inférieurs à cinq secondes, avec une durée moyenne de 3.87 secondes. Ces échantillons ont été annotés en utilisant l'échelle de somnolence de Karolinska (KSS) (Åkerstedt & Gillberg, 1990), un questionnaire mesurant la somnolence subjective instantanée (Martin *et al.*, 2023b) utilisant une échelle de Lickert à 9 points. Les deux études perceptuelles ont utilisé le même sous-ensemble de 99 échantillons du corpus SLEEP, 9 pour familiariser les annotateurs avec la tâche (un pour chaque niveau de somnolence), et 90 (dix pour chaque niveau de somnolence) pour l'expérience elle-même. Notre analyse se concentre sur les 90 échantillons utilisés pour l'expérience.

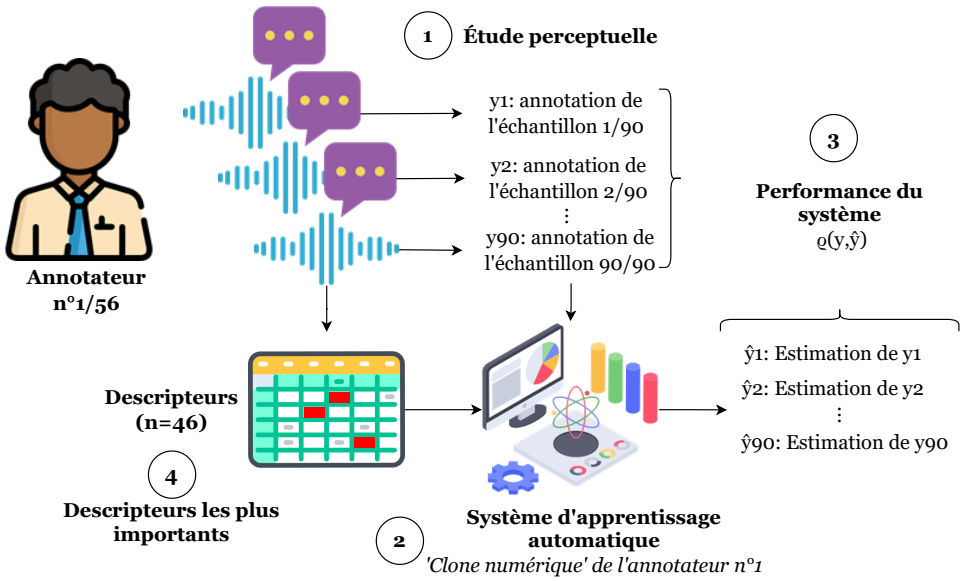

FIGURE 1 – Représentation schématique de notre méthode pour estimer les caractéristiques utilisées par les annotateurs des études perceptuelles pour estimer la somnolence à partir d'échantillons de parole

Études perceptuelles et annotateurs. Lors des deux études perceptuelles, les annotateurs étaient invités à estimer la somnolence de l'orateur à partir des enregistrements audio en utilisant une KSS à 9 points. Les échantillons étaient dans le même ordre pour les deux études, et les annotateurs ne pouvaient pas revenir en arrière. Les deux études sont arrivées à des conclusions différentes : alors que les annotations de l'étude de Huckvale et al. (Huckvale *et al.*, 2020), après application d'un algorithme de 'Sagesse des foules' (*Wisdom of the Crowd*), ont donné des performances très convaincantes ($\rho = 0.72$ entre l'estimation et la vérité terrain), les annotateurs de notre étude de réplication (Martin *et al.*, 2023c) n'ont pas atteint les mêmes performances ($\rho = 0.41$).

De plus, notre étude est la seule à avoir collecté les caractéristiques de chaque annotateur. Celles-ci incluaient leur genre (13F/17M), la sensibilité musicale (n=14 avaient des hobbies ou une profession liés à la musique ; n=16 n'en avaient pas), et leur compréhension de la langue allemande (« au moins un peu », n= 11 ; « pas du tout », n=19). Les autres caractéristiques de chaque étude sont décrites en détail dans un autre article (Martin *et al.*, 2023c).

Caractéristiques vocales. Puisque le sous-corpus sélectionné du corpus SLEEP contient peu d'échantillons par annotateur (90), et pour permettre l'interprétation des profils des annotateurs identifiés, nous nous sommes limités à 46 descripteurs extraits des enregistrements vocaux. Ils incluent la moyenne et l'écart type des caractéristiques de bas niveau (n=40) et les caractéristiques temporelles (n=6) de l'ensemble de descripteurs GEMAPS, extraits à l'aide de la boîte à outils `Opensmile` (Eyben & Schuller, 2015).

Systèmes d'apprentissage automatique. Pour pouvoir interpréter les coefficients des différentes parties du système d'apprentissage, nous avons choisi des algorithmes simples, qui ont précédemment montré leur efficacité sur des petits corpus :

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

(b) Analyse en Composantes Principales (ACP, 80% de variance) + régression linéaire
(c) ACP (80% de variance) + régression à vastes marges (SVR, $C = 1$)

Un système différent a été entraîné pour chaque annotateur (n=26 pour Huckvale et al. 2020, n=30 pour Martin et al. 2023). De plus, pour comparer les performances des systèmes d'apprentissage automatique avec ceux de l'état de l'art pour la détection automatique de la somnolence à partir de la voix (cf. Introduction), nous avons également entraîné un système à reproduire les labels fournis avec le challenge IS2019. Ainsi, un total de 171 systèmes d'apprentissages (57 ensembles d'annotations $\times$ 3 systèmes) ont été entraînés.

Validation croisée et métrique de performance. Afin d'éviter un surapprentissage, les performances ont été calculées dans une procédure de validation croisée 5-fold, répétée 10 fois. En raison de la faible taille de l'échantillon, nous avons agrégé les estimations et les vérités terrain correspondantes, et calculé les performances sur les labels agrégées. De la même manière que pour le challenge IS2019, la métrique de performance choisie était la corrélation de Spearman ρ entre les labels estimés et la vérité terrain. Plus la valeur de ρ est élevée, meilleur est l'estimateur. Les étiquettes et les caractéristiques d'entrée ont été normalisées (z-score).

Contribution de chaque caractéristique. Pour chaque annotateur, nous avons mesuré la contribution de chaque caractéristique dans la chaîne de traitement entraîné pour l'imiter. Pour la chaîne de traitement utilisant uniquement un Lasso pour la classification (a), nous avons considéré les poids normalisés (norme L1) des classificateurs. Pour les autres régresseurs [ACP et régression linéaire (b) ou ACP et SVR (c)], nous avons calculé le produit croisé des coefficients de l'ACP et des coefficients du classificateur, que nous avons normalisé (norme L1). Ce faisant, nous mesurons la contribution de chaque caractéristique à une dimension donnée de l'ACP, qui est pondérée par la contribution de cette dimension de l'ACP à la classification. Pour chacun de ces coefficients, nous avons interprété séparément la valeur absolue – qui est liée à la contribution relative de la caractéristique à la classification – et le signe – qui indique la direction du lien entre la somnolence et la caractéristique vocale.

Lien entre les performances et les caractéristiques des annotateurs. Puisque les caractéristiques des annotateurs ont été collectées dans notre étude perceptuelle, nous avons calculé des tests de Mann-Whitney (MW) afin de mettre en lumière un lien possible entre le genre, la compréhension de la langue ou la sensibilité musicale des annotateurs, et les performances des systèmes entraînés pour reproduire leurs annotations.

Profils d'annotateurs. Afin de dresser les profils des annotateurs, nous avons sélectionné les caractéristiques les plus importantes, c'est-à-dire celles ayant une valeur médiane de contribution normalisée absolue supérieure à 0.05. Nous avons ensuite calculé les profils des annotateurs en utilisant le regroupement hiérarchique avec la fonction `linkage` de la bibliothèque `cluster.hierarchy` de `scipy` (Müllner, 2011). Le regroupement a été effectué en utilisant la méthode de Ward et une métrique euclidienne. Nous avons ensuite identifié les profils des annotateurs, c'est-à-dire les groupes tels que renvoyés par la fonction `linkage`. Pour chaque profil, nous avons pris en compte les performances des systèmes de régression correspondants pour être sûr qu'aucun profil n'était exclusivement constitué des descripteurs des systèmes ayant des performances faibles et que, au contraire, chaque profil était représenté par une diversité de performances.

3 Résultats

3.1 Performances des chaînes de traitement

Les moyennes et écarts-types des performances des sont rapportés dans le Tableau 1.

Ref	Modèle	Défi IS19	Huckvale et al. (n=26)	Martin et al. 2023 (n=30)
(a)	Lasso ($\alpha = 0.1$)	$\rho = 0.437$	$\rho = 0.049 \pm 0.166$	$\rho = 0.356 \pm 0.116$
(b)	ACP (0.8) + Régr. linéaire	$\rho = 0.459$	$\rho = 0.066 \pm 0.164$	$\rho = 0.323 \pm 0.100$
(c)	ACP (0.8) + SVR ($C = 1$)	$\rho = 0.447$	$\rho = 0.051 \pm 0.143$	$\rho = 0.289 \pm 0.095$

TABLE 1 – Performance des systèmes d'apprentissage automatique entraînés à imiter les annotateurs des études perceptuelles. Les valeurs sont calculées sur l'agrégation d'une validation croisée à 5-fold répétée dix fois, représentées sous la forme *Moyenne* $\pm$ *écart-type*

Sur le sous-corpus de 90 échantillons du corpus SLEEP, nos trois chaînes de traitement obtiennent des performances supérieures aux systèmes état-de-l'art sur l'ensemble du corpus (cf. Introduction), confirmant leur pertinence pour la tâche.

Sur les annotations de l'étude perceptuelle de Huckvale et al., aucun classificateur ne donne d'estimation satisfaisante des labels : tous les systèmes obtiennent des performances inférieures à ρ=0.283 et la plupart d'entre elles sont négatives, indiquant que le système n'a rien généralisé. En conséquence, nous ne les avons pas utilisés dans la suite. En revanche, le système (a) atteint un coefficient de corrélation moyen de $\rho = 0.356$ lors de la réplication des labels de notre étude perceptuelle, ce qui est dans l'ordre de grandeur des performances habituellement obtenues sur l'ensemble du corpus (cf. Introduction).

3.2 Influence des caractéristiques des auditeurs

Nous ne trouvons aucune différence dans les performances des systèmes de régression en fonction du sexe (MW, $U = 147, p = 0.132$), de la sensibilité musicale (MW, $U = 85, p = 0.271$), ou du niveau de compréhension de l'allemand (MW, $U = 145, p = 0.085$) de notre étude perceptuelle (Martin *et al.*, 2023c). Nous en déduisons donc que ces variables ne biaisent pas notre interprétation des caractéristiques de la parole impliquées dans l'estimation de la somnolence par les annotateurs.

3.3 Caractéristiques les plus saillantes

Parmi les 46 caractéristiques extraites, six sont identifiées comme les plus saillantes, c'est-à-dire ayant une valeur médiane de contribution normalisée absolue à travers les annotateurs supérieure à 0.05. Elles sont rapportées dans le Tableau 2.

3.4 Regroupement hiérarchique

Le regroupement hiérarchique a été effectué sur ces six caractéristiques pour dresser les profils des annotateurs dans notre étude perceptuelle. Nous avons identifié trois profils principaux, qui sont

représentés avec les caractéristiques les plus saillantes et la performance de chaque système de régression dans la Figure 2.

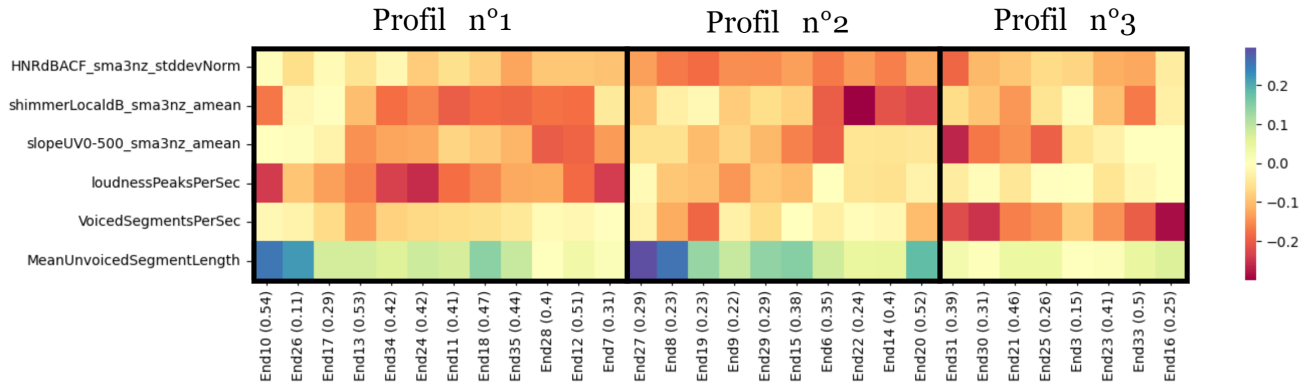

FIGURE 2 – Profils des annotateurs identifiés par le regroupement hiérarchique des systèmes entraînés à les imiter. Chaque ligne correspond à un descripteur, chaque colonne à un annotateur. La performance du système entraîné à reproduire chaque annotateur est indiquée entre parenthèse (ρ de Spearman).

Le premier groupe d'annotateurs (Profil n°1, n=12) associe la somnolence à une voix ayant des segments non voisés plus longs, et une voix plus douce (loudnessPeksPerSec) et moins expressive (slopeUV0-500 et shimmer), avec un accent particulier sur le volume. En revanche, les annotateurs du Profil n°2 (n=10) estiment la somnolence en utilisant des informations prosodiques (longueur des segments non voisés mais aussi le nombre de segments voisés par seconde), l'expressivité de la voix (slopeUV0-500, shimmer), mais aussi la pureté de la voix (variations de HNR). Enfin, les annotateurs du Profil n°3 (n=8) ne se basent pas sur la longueur du segment non voisé pour identifier la somnolence, mais se concentrent sur le nombre de segments voisés par seconde et la variabilité de la hauteur (slopeUV0-500).

4 Comparaison avec les approches automatiques

À notre connaissance, aucun système précédent travaillant sur le corpus SLEEP n'a étudié la contribution des descripteurs à l'estimation de la somnolence. Cependant, dans un de nos précédents travaux sur les tâches de lecture du *Sleepy Language Corpus* (même étiquette de somnolence que le corpus SLEEP) nous avions rapporté la corrélation entre les descripteurs acoustiques et la somno-

Nom	Description	médiane
`HNRdBACF_sma3nz_stddevNorm`	Écart-type du HNR	-0.102
`shimmerLocaldB_sma3nz_amean`	Moyenne du shimmer	-0.099
`slopeUV0-500_sma3nz_amean`	Pente de fréquence dans la bande passante [0,500Hz]	-0.095
`loudnessPeaksPerSec`	Pics d'énergie par seconde (moy.)	-0.09
`VoicedSegmentsPerSec`	Nombre de segments voisés par seconde	-0.065
`MeanUnvoicedSegmentLength`	Longueur moyenne des segments non voisés	0.075

TABLE 2 – Caractéristiques les plus saillantes dans la chaîne de traitement formé pour imiter les annotateurs dans notre étude perceptuelle (Martin *et al.*, 2023c). Les valeurs négatives signifient que la valeur de la caractéristique diminue lorsque la somnolence augmente.

lence (Martin *et al.*, 2019). Dans ce travail, les caractéristiques les plus corrélées à la somnolence étaient principalement liées à la fréquence fondamentale F0 (moyenne, max, min), la fréquence du premier formant (F1) et la plage d'énergie. À l'inverse, le HNR et la durée des segments voisés ou non voisés n'étaient pas parmi les caractéristiques les plus liées à la somnolence. De plus, en appliquant la même méthodologie à nos données, les caractéristiques les plus corrélées avec la vérité terrain donnée avec le corpus sont en partie celles identifiées comme saillantes dans l'imitation des annotateurs. En effet, alors que la pente de la fréquence fondamentale F0 ($\rho = -0.40$), le shimmer ($\rho = -0.34$) et le HNR ($\rho = -0.27$) sont fortement corrélés avec l'étiquette de somnolence, les pics de volume ($\rho = 0.10$), la durée des segments non voisés ($\rho = 0.13$) et le nombre de segments voisés ($\rho = -0.10$) ne sont pas parmi les caractéristiques les plus saillantes avec cette méthode.

Ces résultats questionnent le lien entre la vérité terrain donnée avec le corpus et ce que les annotateurs ont détecté. Le haut degré global d'accord inter-annotateurs rapporté dans notre étude perceptuelle (Martin *et al.*, 2023c) (ICC = 0.975) indique que les annotateurs semblent avoir identifié le même phénomène à travers la voix, qui n'est lui-même pas complètement représenté par l'outil de mesure utilisé pour opérationnaliser la somnolence dans le corpus SLEEP. Cependant, ce label est critiqué dans la littérature (Martin *et al.*, 2021), puisqu'il n'est pas une mesure de la somnolence validée, utilisée et reconnue en médecine du sommeil (Martin *et al.*, 2023b), et n'a jamais été utilisée ailleurs à notre connaissance que dans les deux corpus IS2011 et IS2019. De plus, une autre étude perceptuelle que nous avons menée sur le MSLTc (Martin *et al.*, 2023a), qui contient des mesures validées de la somnolence (Martin *et al.*, 2021), a conclu à la faisabilité de la détection de la somnolence par l'audition humaine à l'aide d'échantillons de parole. Nous interprétons donc cette différence entre les caractéristiques utilisées par les annotateurs et les caractéristiques corrélées avec l'étiquette fournie avec le corpus comme provenant de l'outil de mesure de la somnolence utilisé dans le corpus SLEEP.

5 Conclusion et perspectives

En entraînant des algorithmes d'apprentissage automatique à reproduire les annotations d'une étude perceptuelle sur la somnolence, nous avons pu identifier les caractéristiques sur lesquelles les annotateurs se sont appuyés pour produire cette évaluation ; et ainsi indirectement les indices qu'ils ont utilisés pour estimer la somnolence. Nous avons identifié six caractéristiques, liées à la stabilité de l'énergie (shimmer et pics d'énergie), le HNR, la variabilité de la fréquence fondamentale (pente de F0), et le ratio et la durée des segments voisés et non-voisés.

Nos prochains travaux se concentreront sur l'inclusion d'autres dimensions telles que les pauses de lecture (Martin *et al.*, 2022) ou la réalisation phonétique (Beaumard *et al.*, 2023) dans l'imitation du comportement d'annotation dans ces études perceptuelles.

Remerciements

Cette recherche est financée par l'Agence Nationale de la Recherche (ANR) dans le cadre de l'axe Autonom-Health du PEPR Santé Numérique, convention de subvention n°ANR-22-PESN-0009. VPM a reçu le soutien financier du programme de recherche et d'innovation européen Horizon Europe à travers le projet Marie Skłodowska-Curie MATER (No. 101106577). CB a reçu le soutien financier de la MITI du CNRS (projet PRIME 80 DSM-HEALTH).

Références

AMIRIPARIAN S., WINOKUROW P., KARAS V., OTTL S., GERCZUK M. & SCHULLER B. W. (2020). *A Novel Fusion of Attention and Sequence to Sequence Autoencoders to Predict Sleepiness From Speech*. arXiv 2005.08722. _eprint : 2005.08722.

BARNES C. M. & WATSON N. F. (2019). Why healthy sleep is good for business. *Sleep Med. Rev.*, **47**, 112–118. DOI : 10.1016/j.smrv.2019.07.005.

BEAUMARD C., MARTIN V. P., WU Y., ROUAS J.-L. & PHILIP P. (2023). Automatic detection of schwa in French hypersomniac patients. In *Journée Santé et Intelligence Artificielle (Evènement affilié à PFIA 2023)*.

CAMPBELL E. L., DOCIO-FERNANDEZ L., GARCIA-MATEO C., WITTENBORN A., KRAJEWSKI J. & CUMMINS N. (2022). Automatic detection of short-term sleepiness state. Sequence-to-Sequence modelling with global attention mechanism. In *Workshop on Speech, Music and Mind*. DOI : 10.21437/SMM.2022-2.

EGAS-LÓPEZ J. V., BUSA-FEKETE R. & GOSZTOLYA G. (2022). On the Use of Ensemble X-Vector Embeddings for Improved Sleepiness Detection. In S. R. M. PRASANNA, A. KARPOV, K. SAMUDRAVIJAYA & S. S. AGRAWAL, Éds., *Speech and Computer*, Lecture Notes in Computer Science, p. 178–187, Cham : Springer International Publishing. DOI : 10.1007/978-3-031-20980-2_16.

EYBEN F. & SCHULLER B. (2015). Opensmile. *ACM SIGMultimedia Records*, **6**, 4–13.

FAGHERAZZI G., ZHANG L., ELBÉJI A., HIGA E., DESPOTOVIC V., OLLERT M., AGUAYO G. A., NAZAROV P. & FISCHER A. (2021). A Voice-Based Biomarker for Monitoring Symptom Resolution in Adults with COVID-19 : Findings from the Prospective Predi-COVID Cohort Study. *SSRN Journal*. DOI : 10.2139/ssrn.3949487.

FRITSCH J., DUBAGUNTA S. P. & MAGIMAI.-DOSS M. (2020). Estimating the Degree of Sleepiness by Integrating Articulatory Feature Knowledge in Raw Waveform Based CNNS. In *ICASSP 2020*, p. 6534–6538, Barcelona, Spain. DOI : 10.1109/ICASSP40776.2020.9053351.

GOSZTOLYA G. (2019). Using Fisher Vector and Bag-of-Audio-Words Representations to Identify Styrian Dialects, Sleepiness, Baby & Orca Sounds. In *Interspeech 2019*, p. 2413–2417. DOI : 10.21437/Interspeech.2019-1726.

HODDES E., ZARCONE V., SMYTHE H., PHILLIPS R. & DEMENT W. C. (1973). Quantification of Sleepiness : A New Approach. *Psychophysiology*, **10**(4), 431–436. DOI : 10.1111/j.1469-8986.1973.tb00801.x.

HUANG D.-Y., GE S. S. & ZHANG Z. (2011). Speaker State Classification Based on Fusion of Asymmetric SIMPLS and Support Vector Machines. In *Interspeech 2011*, p.4̃.

HUCKVALE M., BEKE A. & IKUSHIMA M. (2020). Prediction of Sleepiness Ratings from Voice by Man and Machine. In *Interspeech 2020*. DOI : 10.21437/Interspeech.2020-1601.

JIKE M., ITANI O., WATANABE N., BUYSSE D. J. & KANEITA Y. (2018). Long sleep duration and health outcomes : A systematic review, meta-analysis and meta-regression. *Sleep Med. Rev.*, **39**, 25–36. DOI : 10.1016/j.smrv.2017.06.011.

KOLLA B. P., HE J.-P., MANSUKHANI M. P., FRYE M. A. & MERIKANGAS K. (2020). Excessive sleepiness and associated symptoms in the U.S. adult population : prevalence, correlates, and comorbidity. *Sleep Health*, **6**(1), 79–87. DOI : 10.1016/j.sleh.2019.09.004.

KRAJEWSKI J., BATLINER A. & GOLZ M. (2009). Acoustic sleepiness detection : Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods*, **41**(3), 795–804.

LÉGER D., BAYON V., LAABAN J. P. & PHILIP P. (2012). Impact of sleep apnea on economics. *Sleep Med. Rev.*, **16**(5), 455–462. DOI : 10.1016/j.smrv.2011.10.001.

MARTIN V. P., ARNAUD B., ROUAS J.-L. & PHILIP P. (2022). Does sleepiness influence reading pauses in hypersomniac patients ? In *Speech Prosody 2022*, p. 62–66 : ISCA. DOI : 10.21437/SpeechProsody.2022-13.

MARTIN V. P., FERRON A., ROUAS J.-L., SHOCHI T., DUPUY L. & PHILIP P. (2023a). Physiological vs. Subjective sleepiness : what can human hearing estimate better ? In *International Conference on Phonetic Science (ICPhS) 2023*, p. 196–200.

MARTIN V. P., LOPEZ R., DAUVILLIERS Y., ROUAS J.-L., PHILIP P. & MICOULAUD-FRANCHI J.-A. (2023b). Sleepiness in adults : An umbrella review of a complex construct. *Sleep Medicine Reviews*, **67**, 101718. DOI : 10.1016/j.smrv.2022.101718.

MARTIN V. P., ROUAS J.-L., FERRON A. & PHILIP P. (2023c). "Prediction of sleepiness ratings from voice by man and machine" : the Endymion replication perceptual study. In *International Conference on Phonetic Science (ICPhS) 2023*, p. 201–205.

MARTIN V. P., ROUAS J.-L., MICOULAUD-FRANCHI J.-A., PHILIP P. & KRAJEWSKI J. (2021). How to Design a Relevant Corpus for Sleepiness Detection Through Voice ? *Front. Digit. Health*, **3**, 686068. DOI : 10.3389/fdgth.2021.686068.

MARTIN V. P., ROUAS J.-L. & PHILIP P. (2024). Automatic detection of sleepiness-related symptoms and syndromes using voice and speech biomarkers. *Biomedical Signal Processing and Control*, **91**, 105989. DOI : https ://doi.org/10.1016/j.bspc.2024.105989.

MARTIN V. P., ROUAS J.-L., THIVEL P. & KRAJEWSKI J. (2019). Sleepiness detection on read speech using simple features. In *10th Conference on Speech Technology and Human-Computer Dialogue*, Timisoara, Romania. DOI : 10.1109/SPED.2019.8906577.

MÜLLNER D. (2011). Modern hierarchical, agglomerative clustering algorithms. Publisher : arXiv Version Number : 1, DOI : 10.48550/ARXIV.1109.2378.

SCHULLER B., BATLINER A., BERGLER C., POKORNY F. B., KRAJEWSKI J., CYCHOCZ M., VOLLMAN R., ROELEN S.-D., SCHNIEDER S., BERGELSON E., CRISTIA A., SEIDL A., WARLAUMONT A., YANKOWITZ L., NÖTH E., AMIRIPARIAN S., HANTKE S. & SCHMITT M. (2019). The INTERSPEECH 2019 Computational Paralinguistics Challenge : Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity. In *Interspeech 2019*. DOI : 10.21437/Interspeech.2019-1122.

SCHULLER B., STEIDL S., BATLINER A., SCHIEL F. & KRAJEWSKI J. (2011). The INTERSPEECH 2011 Speaker State Challenge. In *Interspeech 2011*, p. 3201–3204. DOI : 10.21437/Interspeech.2011-801.

SCOTT A. J., WEBB T. L., MARTYN-ST JAMES M., ROWSE G. & WEICH S. (2021). Improving sleep quality leads to better mental health : A meta-analysis of randomised controlled trials. *Sleep Med. Rev.*, **60**, 101556. DOI : 10.1016/j.smrv.2021.101556.

TRAN B., ZHU Y., LIANG X., SCHWOEBEL J. W. & WARRENBURG L. A. (2022). Speech Tasks Relevant to Sleepiness Determined With Deep Transfer Learning. In *ICASSP 2022*, p. 6937–6941. ISSN : 2379-190X, DOI : 10.1109/ICASSP43922.2022.9747000.

ÅKERSTEDT T. & GILLBERG M. (1990). Subjective and objective sleepiness in the active individual. *Int J Neurosci*, **52**, 29–37. DOI : 10.3109/00207459008994241.