

Détection de la somnolence par estimation d'erreurs de lecture

Vincent P. Martin¹ Gabrielle Chapouthier² Mathilde Rieant² Jean-Luc Rouas¹
Pierre Philip³

(1) LaBRI - Univ. Bordeaux - Bordeaux INP - CNRS - UMR5800 - F-33400 Talence, France

(2) CFUOB - Univ. Bordeaux Sengalen - F-33076 Bordeaux, France

(3) SANPSY - CNRS - USR 3413 - Univ. Bordeaux - CHU Pellegrin - F-33000 Bordeaux, France

{vincent.martin, rouas}@labri.fr, {gabrielle.chapouthier,
mathilde.rieant}@etu.u-bordeaux.fr, pierre.philip@u-bordeaux.fr

RÉSUMÉ

La détection automatique de la somnolence peut aider le suivi de patients souffrant de maladies neuro-psychiatriques chroniques. Des recherches précédentes ont déjà montré que cela est possible en utilisant des enregistrements vocaux. Dans cet article, nous proposons d'étudier les erreurs de lecture effectuées par des patients souffrant de Somnolence Diurne Excessive (SDE) sur le corpus TILE, enregistré à l'hôpital de Bordeaux. Avec des orthophonistes, nous avons défini et compté les erreurs de lecture des patients et les avons confrontées aux différentes mesures de somnolence du corpus. Nous montrons ici que relever ces erreurs peut être utile pour élaborer des marqueurs robustes de la somnolence objective mais aussi pour définir des critères d'exclusion des locuteurs n'ayant pas un niveau de lecture suffisant.

ABSTRACT

Sleepiness detection through reading errors estimation

Automatic detection of sleepiness can help to improve the follow-up of patients suffering from chronic diseases. Previous research on sleepiness detection has shown that this task is feasible using voice recordings. In this paper, we propose to study the reading errors made by patients suffering from Excessive Daytime Sleepiness (EDS) on the MSLT Corpus, collected at the Bordeaux hospital. With the help of speech therapists, we defined and counted reading errors and confront these numbers with sleepiness measurements. We show that evaluating these reading errors can be useful to elaborate robust markers of objective sleepiness but also to elaborate exclusion criteria of the speakers not having a sufficient reading level.

MOTS-CLÉS : Erreurs de lecture, Détection de la somnolence, Prosodie.

KEYWORDS: Reading Mistakes, Sleepiness detection, Prosody.

1 Introduction

L'un des défis majeurs actuels du diagnostic et du traitement des maladies chroniques en neuro-psychiatrie est la quantification des symptômes et le suivi des patients, afin d'adapter leur traitement et de détecter précocement les rechutes. Ce suivi est aujourd'hui possible grâce à des dispositifs médicaux connectés (mesurant par exemple le poids, la pression sanguine ou l'activité physique) mais des informations cruciales telles que le niveau de fatigue ou de somnolence sont difficiles à

mesurer en milieu écologique. Par ailleurs, ces pathologies nécessitent des entretiens réguliers entre les médecins et les patients, mais ceux-ci ne permettent pas de mesurer les variations des symptômes en réponse au traitement lorsque les patients sont à leur domicile. De plus, le nombre grandissant de patients augmente le temps d'attente entre deux entretiens, ce qui implique des entretiens irréguliers avec les médecins et donc un suivi épisodique et incomplet des patients.

Les avancées récentes dans le domaine du traitement automatique de la voix rendent possible la détection des états du locuteur via l'analyse d'indices vocaux (Cummins *et al.*, 2018). Ces méthodes présentent de nombreux avantages : elles ne sont pas invasives, ne nécessitent ni capteurs spécifiques ni processus de calibration, et peuvent être mises en place dans des environnements variés. Elles permettent ainsi un suivi régulier et non restrictif des patients.

Notre objectif est d'estimer la somnolence objective de patients souffrant de Somnolence Diurne Excessive (SDE). Aucun corpus existant ne nous permettait d'élaborer un système répondant à cet objectif : la majorité des études portant sur la détection de la somnolence dans la voix sont basées soit sur le Sleepy Language Corpus (Martin *et al.*, 2019; Schuller *et al.*, 2011) soit plus récemment sur le SLEEP corpus (Schuller *et al.*, 2019). Or, ces corpus sont enregistrés sur des volontaires sains dont la seule mesure associée à la voix est la somnolence subjective. Afin d'avoir une annotation de la somnolence objective d'échantillons vocaux recueillis sur des patients souffrant de SDE, nous avons élaboré notre propre corpus, le corpus TILE (*MSLT corpus* en anglais) (Martin *et al.*, 2020), enregistré à la Clinique du Sommeil du Centre Hospitalier Universitaire de Bordeaux.

Par ailleurs, à notre connaissance, la majorité des travaux précédents ayant pour objectif de détecter la somnolence dans la voix sont basés sur des marqueurs vocaux concernant la qualité de la voix (énergie, fréquence, ...), généralement extraits avec la toolbox openSMILE (Eyben & Schuller, 2015). Cette étude tire bénéfice du fait que contrairement aux autres corpus, dans le corpus TILE tous les échantillons vocaux sont collectés lors d'une tâche de lecture, permettant d'annoter les erreurs de lecture faites par rapport au texte original. La nouveauté de cette approche réside dans le fait de se servir des erreurs de lecture comme nouveaux marqueurs de la somnolence. En effet, si les marqueurs de qualité vocale permettent d'étudier l'influence de la somnolence sur l'aspect neuro-musculaire de la voix (Krajewski *et al.*, 2009), nous pensons que les erreurs de lecture sont des marqueurs pertinents pour l'étude de l'influence de la somnolence sur les performances cognitives. Nous proposons ainsi une nouvelle méthode pour évaluer la somnolence d'un locuteur à partir de ses erreurs de lecture.

Cet article est organisé comme suit. Dans la Section 2, nous décrivons les cinq types d'erreurs de lecture que nous avons utilisées pour annoter notre corpus. Dans la Section 3, nous présentons brièvement notre corpus et les critères d'exclusion que nous avons mis au point. Les résultats sont présentés et discutés dans la Section 4. La Section 5 introduit une première tentative de classification. Enfin, la conclusion et nos futurs travaux sur le sujet sont présentés dans la Section 6.

2 Définition des erreurs de lecture

En travaillant avec des orthophonistes, nous avons sélectionné cinq types d'erreurs de lecture, qui sont à la fois suffisamment générales pour être représentées dans notre corpus et assez restrictives pour être spécifiques à certains comportements de lecture.

- Les achoppements (Ach) : "hésitation, coupure, dans le rythme de la parole" (Brin *et al.*, 2018).

Ces erreurs mesurent principalement la capacité d'*assemblage* du lecteur. L'*assemblage* est le fait de mettre bout à bout des syllabes pour former un mot : quand un lecteur commence à lire un mot, s'arrête, et reprend sa lecture, le processus d'*assemblage* a été interrompu, ce qui cause un achoppement. Nous avons choisi de ne pas prendre en compte les hésitations entre les mots (interruption dans le débit de parole), mais seulement les arrêts qui sont observés au milieu d'un mot ou les allongements artificiels de certaines voyelles, qui témoignent d'une hésitation. En effet, les différentes accents rencontrés lors des enregistrements induisent des hésitations plus ou moins appuyées entre les mots, ce qui n'est pas le cas des hésitations dans les mots. Les hésitations et reprises de bouts de phrases ou de phrases sont également prises en compte dans cette catégorie. Dans le cas d'une reprise, un seul achoppement est compté, quel que soit le nombre de mots repris.

- Les paralexies (Plx) : "erreur d'identification de mots écrits consistant à oraliser un mot écrit à la place d'un autre" (Brin *et al.*, 2018).

Contrairement aux achoppement, les paralexies reflètent les capacités d'*adressage* du lecteur. En effet, contrairement au processus d'*assemblage*, l'*adressage* est défini comme le fait de lire un mot dans sa globalité, sans le déchiffrer ou le découper en syllabes. Les paralexies sont des erreurs symptomatiques de ce type de lecture. Nous avons choisi de généraliser cette catégorie à toute prononciation d'un mot, existant ou non, qui est lu à la place du mot correct.

- Les oublis de mots (O) : cette erreur est observée lorsque le locuteur oublie de prononcer un mot et passe directement au suivant ou au début du mot suivant avant de se reprendre. Même s'il se corrige ensuite, l'oubli est compté.
- Les additions (Add) : cette erreur est observée lorsque le locuteur ajoute un mot qui n'est pas présent dans le texte. Même s'il se corrige ensuite, l'addition est prise en compte.
- Les inversions de mots (I) : cette erreur est observée lorsque le locuteur inverse plusieurs mots dans la phrase.

Si un locuteur se reprend après une paralexie, un oubli, une addition ou une inversion de mots, aucun achoppement n'est compté sauf s'il se trompe avec l'erreur correspondante lors de sa reprise.

3 Description de la base de données

3.1 Présentation du corpus

Le corpus utilisé dans cette étude est une version étendue du corpus TILE (Martin *et al.*, 2020). Il comprend les enregistrements de 115 patients enregistrés à la Clinique du Sommeil du Centre Hospitalier Universitaire de Bordeaux (France). Tous les patients ont émis des plaintes concernant des problèmes de sommeil et passent un Test Itératif de Latence d'Endormissement (Littner *et al.*, 2005) - TILE. Ce test consiste en 5 siestes espacées de 2h à partir de 9h du matin. Les patients sont largement phénotypés et leurs caractéristiques physiques ainsi que leurs résultats à des questionnaires subjectifs de dépression, de fatigue et de somnolence sur le long-terme sont collectés. Le principal avantage de ce corpus réside dans le fait qu'il associe à chaque échantillon audio deux mesures de somnolence : une mesure subjective (Karolinska Sleepiness Scale (Åkerstedt & Gillberg, 1990) - KSS) et une mesure objective (temps d'endormissement des patients à chaque sieste, appelé "valeur de TILE"). Cette double annotation est importante car si les sujets sains estiment correctement leur niveau de somnolence (Horne & Burley, 2010), ce n'est pas le cas des malades (Sangal, 1999). Les échantillons vocaux sont collectés durant la lecture d'un texte, qui est différent à chaque session mais le même pour tous les locuteurs à session constante. Afin d'éviter une trop grande valence émotionnelle et

pour avoir une grammaire et un vocabulaire simple, nous avons proposé aux sujets des extraits du Petit Prince d'un peu plus de 200 mots.

3.2 Sélection des sujets (critères d'exclusion)

Donnée	Femmes	Hommes	Total
Nombre de sujets	59	40	99
Nombre d'échantillons	295	200	495
Âge moyen (écart-type)	34,2 (11,6)	39,0 (17,1)	36,1 (14,3)
Niveau Social moyen (écart-type)	4,6 (2,4)	5,9 (2,6)	5,4 (2,6)
TILE moyenne (écart-type)	11,8 (4,6)	10,3 (5,2)	11,2 (4,9)
KSS moyen (écart-type)	4,2 (1,2)	4,6 (1,2)	4,4 (1,2)
Nombre de sujets somnolents - S	12	15	27
Nombre de sujets non somnolents - NS	47	25	72

TABLE 1: Statistiques concises du corpus une fois les patients exclus

En annotant la base de données avec les erreurs décrites précédemment, nous avons pu établir des critères d'exclusion pour cette étude. Exclure des patients de la base de données va certes en réduire la taille, mais cela va en contrepartie assurer que les marqueurs vocaux calculés et les erreurs de lecture mesurées sur les patients inclus seront principalement influencés par la somnolence, excluant les effets des pathologies et des troubles de la lecture sur les variables mesurées.

Tout d'abord, nous avons exclus trois patients présentant des troubles visuo-attentionnels ou d'alexie, séquelles répandues des AVC ou des Accidents Ischémiques Transitoires (AIT). Les patients exclus avaient effectivement de tels antécédents médicaux et avaient rythme de lecture très lent accompagné de très nombreuses erreurs. De même, trois patients produisaient de nombreuses erreurs de lecture suspectées d'être liées à un manque de contrôle musculaire lors de la production vocale, dûes à une maladie neuro-musculaire (comme par exemple une dysphonie, une myotonie, une chorée d'Huntington ou de l'épilepsie). Ces patients produisant un grand nombre d'erreurs ont tous les trois été diagnostiqué de maladies neuro-musculaires.

Trois patients ont été également été exclus après avoir omis ou répété une ligne, ou après avoir omis un trop grand nombre de mots, rendant la lecture incohérente. En effet, ces erreurs d'attention peuvent être dues à la somnolence, mais ce sont aussi des marqueurs des Troubles et Déficit de l'Attention avec ou sans Hyperactivité (TDAH). Différencier l'origine de ces erreurs étant très difficile, nous choisissons d'exclure ces patients, dont un était diagnostiqué TDAH.

Concernant la fluence verbale, un patient présentait des caractéristiques de bredouillement et a été exclus de notre corpus. Les pathologies concernant la fluence verbale sont importantes à prendre compte puisque leurs effets sont difficilement différenciables des effets de la somnolence. Finalement, nous avons également exclus quatre patients qui déchiffraient le texte lors de la lecture à voix haute malgré une précédente lecture dans leur tête quelques minutes avant, symptômes d'une éventuelle dyslexie ou d'un trouble associé.

Un patient supplémentaire souffrant de différentes maladies inflammatoires sérieuses (Maladie de Crohn, de Basedown et Spondylarthrite Ankylosante) et un autre souffrant lui de troubles anxieux sévères impactant sa lecture ont également été exclus.

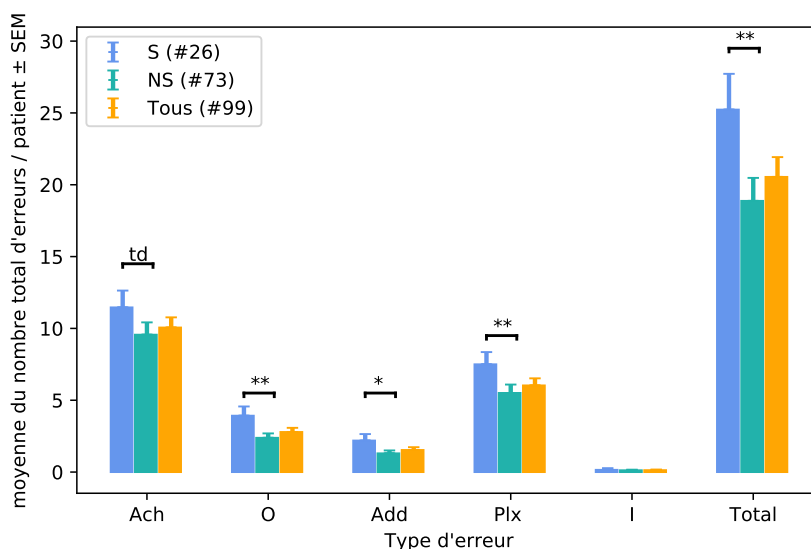


FIGURE 1: Distribution du nombre total d'erreurs par locuteur (moyenne \pm SEM). Ach : Achoppements, O : Oublis, Add : Additions, Plx : Paralexies, I : Inversions de mots. Tests de Mann-Whitney (td : $p < 6 \times 10^{-2}$, * : $p < 5 \times 10^{-2}$, ** : $p < 10^{-2}$).

Nous avons ainsi gardé un total de 99 locuteurs sur les 115 originaux. Des statistiques concises sur le corpus une fois ces patients exclus sont proposés dans le Tableau 1. Les patients sont séparés entre Somnolents (S) et Non-Somnolents (NS) grâce à la limite médicale de 8 minutes sur la moyenne des cinq valeurs de TILE qui est utilisée pour le diagnostic de la narcolepsie (Aldrich *et al.*, 1997).

4 Résultats

4.1 Caractéristiques des locuteurs

Afin de mesurer si les erreurs évaluées précédemment sont sensibles à la somnolence, nous avons tracé la distribution du nombre total de chaque type d'erreur par locuteur dans la Figure 1 (moyenne \pm SEM - *Standard Error of the Mean*).

Excepté pour les Inversions de mots, les patients Somnolents produisent plus d'erreurs que leurs homologues Non-Somnolents, et ce quelle que soit la catégorie d'erreur (Test de Mann-Whitney. Ach : $U = 746,0$; $p = 5,3 \times 10^{-3}$. O : $U = 635,0$; $p = 5,7 \times 10^{-3}$. Add : $U = 702,5$; $p = 2,1 \times 10^{-2}$. Plx : $U = 616,5$; $p = 3,9 \times 10^{-3}$. I : $U = 921,5$; $p = 0,34$. Total : $U = 631,0$; $p = 5,8 \times 10^{-3}$). En raison du faible nombre d'Inversions de mots observés, ce type d'erreur n'est pas pris en compte dans la suite de l'étude.

Nous avons ensuite étudié si le nombre total d'erreurs dans chaque catégorie corrèle (ρ de Spearman) avec les différentes données médicales et sociales disponibles dans le corpus. Contrairement à nos intuitions, le niveau social (mesuré comme le nombre d'années d'étude après le Brevet des Collèges) ou encore l'âge ne corrèlent pas avec la production d'erreurs des 99 locuteurs du corpus. Il y a cependant une corrélation entre le nombre total d'additions par locuteur et la valeur moyenne de TILE ($\rho = 0,28$; $p = 5,3 \times 10^{-3}$) : plus les patients sont affectés par une maladie du sommeil

(sommolence objective plus élevée), plus ils font d'additions dans les textes. Ces erreurs corrént également avec le score du questionnaire Index de Sévérité de l'Insomnie (Morin *et al.*, 2011) ($\rho = 0,24$; $p = 1,6 \times 10^{-2}$) : plus les patients ont de troubles du sommeil et des insomnies, plus ils produisent d'additions.

Les paralexies corrént non seulement avec la sommolence objective, i.e. la valeur moyenne de TILE ($\rho = 0,25$; $p = 1,3 \times 10^{-2}$) mais aussi avec la sommolence subjective, i.e. la valeur moyenne des KSS ($\rho = 0,25$; $p = 1,1 \times 10^{-2}$). Le fait que les paralexies corrént positivement avec à la fois les mesures de sommolence objectives et subjectives indique que la production de ce type d'erreurs augmente avec la sévérité de leur sommolence objective mais aussi avec la perception que les locuteurs en ont. Cela a l'avantage de pouvoir détecter les deux types de sommolence grâce à ce type d'erreurs mais a le principal inconvénient de ne pas pouvoir différencier les deux types de sommolence : un patient produisant un grand nombre de paralexies pourra soit avoir une sommolence objective haute ou alors seulement le ressenti qu'elle est haute. En comparaison, le nombre d'additions semble être un biomarqueur du niveau de sommolence des patients plus précis puisqu'il ne corrént qu'avec la sommolence objective.

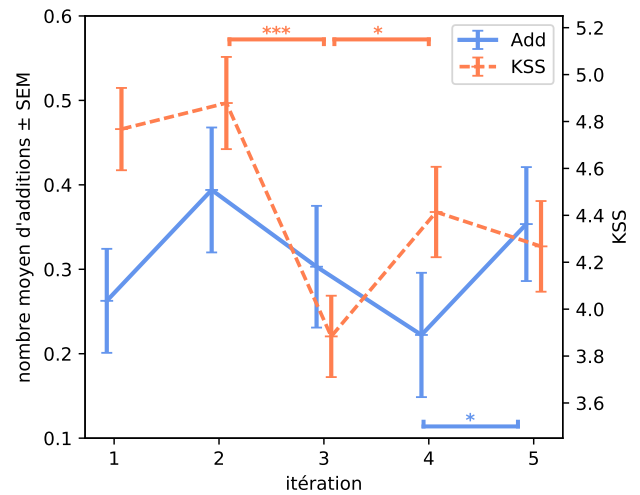
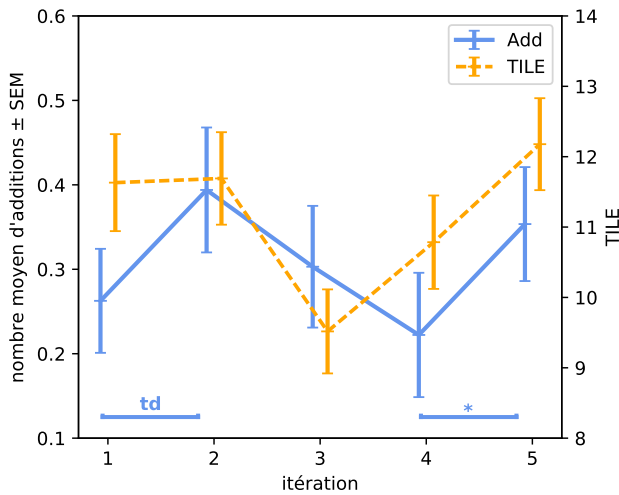
4.2 Étude des sources d'influences de production d'erreurs

Aucun des textes choisis n'a exactement la même taille, la même difficulté ou la même quantité de dialogues. Il est donc nécessaire d'étudier l'influence des textes sur cette production d'erreurs et de la séparer de l'influence de la sommolence. De plus, des variables dépendant du temps peuvent affecter à la fois le niveau de sommolence et la production. Parmi ces variables comptent le fait que les patients petit-déjeunent avant la première itération du test (à 9h du matin), qu'ils déjeunent peu avant la troisième itération (à 13h) ou qu'ils expriment généralement de la fatigue et un ennui vis à vis du test durant la dernière itération (17h). Dans la suite, "influence de l'itération" fera indépendamment référence à l'influence du texte ou des variables précédemment décrites, les deux effets n'étant pas séparables.

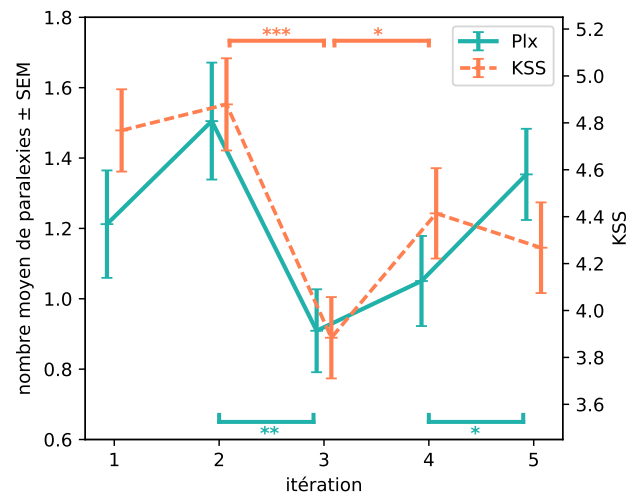
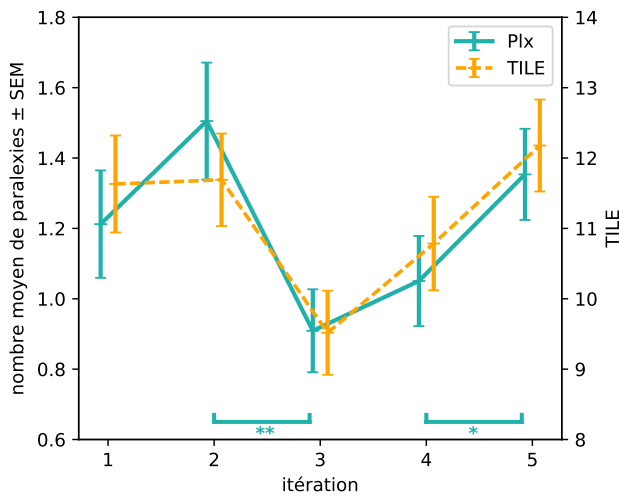
Afin de séparer la contribution de l'itération, de la valeur de TILE et du KSS sur les variations du nombre d'additions et de paralexies, nous avons appliqué une ANOVA multivariée à mesures répétées avec R (R Core Team, 2017).

En étudiant les variations des valeurs de TILE et du nombre d'additions chez les deux classes de patients (Figure 2a), la sommolence objective semble avoir une influence sur la production de ce type d'erreurs. Une ANOVA permet de mettre en évidence une influence significative du TILE sur les variations inter-sujets ($F = 6,0$; $p = 1,6 \times 10^{-2}$) et une influence presque significative du KSS sur les variations intra-sujets ($F = 3,4$; $p = 6,3 \times 10^{-2}$). Cela signifie que les différences observées entre les sujets indépendamment du temps sont principalement expliquées par leurs différences de TILE (ce qui confirme le lien entre TILE et additions) tandis que celles observées entre les sujets au cours du temps (influence conjointe de la session et du locuteur) sont principalement expliquées par les différences de variation de KSS au cours des itérations du test. La session n'a aucun effet significatif sur la production des additions. Nous émettons donc l'hypothèse que les variations du nombre d'additions sont principalement dues aux variations de sommolences objectives et subjectives, et qu'elles sont donc indépendantes du texte et des autres effets temporels.

L'influence de la sommolence sur le nombre de paralexies est plus complexe à analyser. Malgré une forte corrélation entre l'évolution des valeurs de TILE et de KSS avec le nombre de paralexies



(a) Additions et TILE en fonction des itérations du test (b) Additions et KSS en fonction des itérations du test



(c) Paralexies et TILE en fonction des itérations du test (d) Paralexies et KSS en fonction des itérations du test

FIGURE 2: Additions (a,b) et paralexies (c,d) comparées au TILE et au KSS (moyenne \pm SEM). Tests de Mann-Whitney

(td : $p < 6 \times 10^{-2}$, * : $p < 5 \times 10^{-2}$, ** : $p < 10^{-2}$, *** : $p < 10^{-3}$).

observées sur les Figures 2c et 2d, l'effet dominant lors de l'analyse par ANOVA est l'influence de l'itération ($F = 5, 2; p = 4, 2 \times 10^{-4}$). En effet, nous avons observé lors de l'annotation de la base de données que certains mots sont systématiquement mal prononcés, menant à une paralexie (par exemple le mot "méditatif" est très souvent prononcé "médiatif"). L'influence du texte induit ici une influence de l'itération sur la production des paralexies. Les corrélations entre le nombre total de paralexies et les deux mesures différentes de la somnolence restent malgré tout des indications optimistes concernant l'utilité des paralexies dans la détection de la somnolence.

Une ANOVA permet d'expliquer les variations intra-sujets des achoppements par le KSS ($F = 5, 8; p = 1, 7 \times 10^{-2}$) et l'itération ($F = 5, 9; p = 1, 2 \times 10^{-4}$), ce qui exclut ce type d'erreur des biomarqueurs robustes pour estimer la valeur de TILE. Nous avons choisi d'ignorer les arrêts entre les mots, puisqu'ils sont difficilement différenciables des accents des locuteurs ou de respirations inhabituelles, ce qui pourrait être la cause de l'absence de corrélation entre le nombre total d'achoppements et la

somnolence objective des patients.

Concernant les oublis, nous avons remarqué que les mots oubliés sont souvent les mêmes. Ce sont des petits mots de liaison qui sont en général omis à l'oral (par exemple "Il me répéta alors" à la place de "Et il me répéta alors"). Une ANOVA confirme que les variations du nombre d'oublis dépendent fortement du texte (influence de l'itération : $F = 12,6$; $p = 1,2 \times 10^{-9}$), les empêchant d'être des biomarqueurs robustes de l'état de somnolence du locuteur. Une autre version de ces erreurs pourrait toutefois avoir du potentiel en tant que biomarqueur de la somnolence puisque les variations inter-locuteurs sont expliquées par la valeur de TILE ($F = 4,1$; $p = 0,05$). Ces observations soulèvent la nécessité d'une étude en profondeur du contenu des textes, pour éviter la reproduction de ces problèmes mais également pour s'assurer qu'ils sont visuellement équivalents (les dialogues présents dans notre corpus de textes semblent impliquer des erreurs visuo-attentionnelles) et que la difficulté des textes n'est pas la source des erreurs faites par les sujets.

5 Élaboration d'un classificateur

Une première approche pour élaborer un classificateur entre sujets Somnolents et Non-Somnolents consiste à concaténer les erreurs des cinq itérations du test et d'utiliser ce vecteur comme entrée d'une Machine à Vecteur Support - SVM (noyau linéaire, $C = 1 \times 10^{-2}$). Comme la taille de notre corpus est relativement faible (99 locuteurs), nous utilisons de la validation croisée *Leave One Speaker Out Cross Validation (LOSOVCV)* : chaque locuteur est tour à tour isolé pour servir de test, tandis que les autres forment la base d'entraînement. Le résultat de la classe estimée pour le locuteur de test est ajouté dans une matrice de confusion globale, qui sert d'évaluation moyenne de l'ensemble des systèmes. Après avoir normalisé l'ensemble d'entraînement, entraînés les paramètres du SVM et évalué le système obtenu pour chaque itération de la LOSOVCV, nous obtenons une performance (*Unweighted Accuracy Recall - UAR*) calculée sur la matrice de confusion globale de 61,0% (Sensibilité : 57,7%, Spécificité : 64,4%). Même si les performances obtenues par ce système sont en dessous de celles obtenues dans l'état de l'art, il sera intéressant d'étudier leur combinaison avec des systèmes classiques en utilisant différents ensembles de biomarqueurs.

6 Conclusions et Perspectives

En conclusion, nous avons proposé de nouveaux marqueurs pour la détection de la somnolence à partir des erreurs de lecture. La mesure de ces erreurs permet à la fois de proposer des critères d'exclusion pour notre corpus (basés sur leur niveau de lecture ou sur leurs pathologies) et de mesurer la somnolence des locuteurs grâce à leur voix. Par ailleurs, nous avons aussi montré que ces nouveaux marqueurs peuvent être utilisés dans des tâches de classification d'état de somnolence de locuteur. Cela pourrait aider à améliorer les systèmes classiques de détection de la somnolence dans la voix, actuellement basés sur des marqueurs acoustiques, en les utilisant avec différents types de marqueurs.

Nos futurs travaux comprennent l'élaboration d'un système de détection automatique des erreurs présentées dans cet article grâce à des systèmes de transcription automatique. D'autre part, nous projetons de mettre au point des marqueurs équivalents pour la parole spontanée et d'étudier les possibilités de fusion des marqueurs des systèmes classiques avec les nôtres.

Références

- ALDRICH M. S., CHERVIN R. D. & MALOW B. A. (1997). Value of the multiple sleep latency test (MSLT) for the diagnosis of narcolepsy. *Sleep*, **20**(8), 620–629.
- BRIN F., COURRIER C., LEDERLE E. & MASY V. (2018). *Dictionnaire d'orthophonie - 4ème édition*. Orthoédition édition.
- CUMMINS N., BAIRD A. & SCHULLER B. (2018). Speech analysis for health : Current state-of-the-art and the increasing impact of deep learning. *Health Informatics and Translational Data Analytics*, **151**, 1–54.
- EYBEN F. & SCHULLER B. (2015). Opensmile. *ACM SIGMultimedia Records*, **6**, 4–13.
- HORNE J. & BURLEY C. (2010). We know when we are sleepy : Subjective versus objective measurements of moderate sleepiness in healthy adults. *Biological Psychology*, **83**(3), 266–268.
- 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.
- LITTNER M. R., KUSHIDA C., WISE M., DAVILA D. G., MORGENTHALER T., LEE-CHIONG T., HIRSHKOWITZ M., LOUBE D. L., BAILEY D., BERRY R. B., KAPEN S. & KRAMER M. (2005). Practice Parameters for Clinical Use of the Multiple Sleep Latency Test and the Maintenance of Wakefulness Test. *Sleep*, **28**(1), 113–121.
- MARTIN V. P., ROUAS J.-L., MICOULAUD-FRANCHI J.-A. & PHILIP P. (2020). The Objective and Subjective Sleepiness Voice Corpora. In *12th Language Resources and Evaluation Conference*.
- 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*. DOI : [10.1109/SPED.2019.8906577](https://doi.org/10.1109/SPED.2019.8906577).
- MORIN C. M., BELLEVILLE G., BÉLANGER L. & IVERS H. (2011). The Insomnia Severity Index : Psychometric Indicators to Detect Insomnia Cases and Evaluate Treatment Response. *Sleep*, **34**(5), 601–608.
- R CORE TEAM (2017). *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing.
- SANGAL R. (1999). Subjective sleepiness ratings (Epworth sleepiness scale) do not reflect the same parameter of sleepiness as objective sleepiness (maintenance of wakefulness test) in patients with narcolepsy. *Clinical Neurophysiology*, **110**(12), 2131–2135.
- 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*.
- SCHULLER B., STEIDL S., BATLINER A., SCHIEL F. & KRAJEWSKI J. (2011). The INTERSPEECH 2011 Speaker State Challenge. In *Interspeech 2011*, p. 3201–3204.
- ÅKERSTEDT T. & GILLBERG M. (1990). Subjective and objective sleepiness in the active individual. *Int J Neurosci*, **52**, 29–37.