Détection de la somnolence dans la voix : nouveaux marqueurs et nouvelles stratégies

Vincent P. Martin* — **Jean-Luc Rouas*** — **Pierre Philip****

* *{vincent.martin, rouas}@labri.fr*
LaBRI – Univ. Bordeaux – Bordeaux INP – CNRS – UMR 5800
** *pierre.philip@u-bordeaux.fr*
SANPSY – Univ. Bordeaux – CHU Pellegrin – CNRS – USR 3413

RÉSUMÉ. *Cet article traite de la détection automatique de la somnolence dans la voix en vue de l'amélioration du suivi des patients souffrant de maladies neuropsychiatriques chroniques. Notre première approche s'inspire des systèmes état de l'art mais en les appliquant au cas particulier des patients atteints de somnolence diurne excessive (SDE). Nous basons notre étude sur le corpus TILE, qui diffère des autres corpus existants par le fait que les sujets enregistrés souffrent de SDE et que leur niveau de somnolence est mesuré de manière subjective mais aussi objective. Le système proposé permet de détecter la somnolence objective grâce à des paramètres vocaux simples et explicables à des non-spécialistes. Par la suite, nous avons développé une approche originale basée sur les erreurs de lecture que nous 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.*

ABSTRACT. *This paper deals with automatic sleepiness state estimation using speech applied to the follow-up of patients suffering from chronical neuropsychiatric diseases. Our first approach draws from state-of-the-art systems to estimate sleepiness level from voice, for the specific case of patients suffering from Excessive Daytime Sleepiness (EDS). We base our study on the MSLT corpus, that differs from other existing corpus by the fact that recorded subjects suffer from EDS and that their sleepiness level is measured by both subjective and objective means. The proposed system allows to detect objective sleepiness with simple vocal markers that are explainable to non-specialists. Furthermore, we devised a new method based on the reading errors and investigate their links with sleepiness measurements. We show that evaluating these reading errors can be useful to elaborate robust markers of objective sleepiness.*

MOTS-CLÉS : *détection de la somnolence, marqueurs vocaux, erreurs de lecture.*

KEYWORDS: *sleepiness detection, vocal markers, reading mistakes.*

1. Introduction

Dans un contexte de désertification médicale et d'augmentation de la demande médicale dans le domaine des pathologies neuropsychiatriques, la capacité de service offerte par les structures spécialisées n'est plus suffisante pour le suivi correct des patients. En effet, de nombreuses pathologies neuropsychiatriques chroniques nécessitent un suivi continu des patients afin de quantifier les symptômes et prévenir les rechutes précoces. Cependant, les nombreux entretiens nécessaires à une bonne prise en charge sont souvent irréguliers et ne permettent pas de mesurer les variations des différents symptômes en réponse au traitement lorsque les patients sont à leur domicile, dans des conditions écologiques. Dans ces conditions écologiques, différentes caractéristiques physiques peuvent toutefois être mesurées grâce à des dispositifs médicaux connectés tels que le poids, la pression sanguine ou encore l'activité physique. En revanche, des informations cruciales pour le suivi de ces patients comme la fatigue, la somnolence ou l'humeur des patients ne peuvent être mesurées par ces dispositifs.

Ainsi est née l'idée de créer un dispositif de suivi des patients à leur domicile sous la forme d'un agent conversationnel, élaboré en priorisant son acceptabilité lors d'entretiens réguliers et répétés (Philip *et al.*, 2020 ; Philip *et al.*, 2017). Cet agent, implémenté dans un smartphone que le patient ramène à son domicile, permet de régulariser la mesure des symptômes par le biais de questions posés par l'agent, tout en mesurant leur manifestation dans des conditions écologiques. Durant l'entretien avec l'agent virtuel, le patient a la possibilité de répondre sous la forme d'une interaction vocale. Notre but est de compléter l'analyse des réponses aux différents questionnaires proposés par le médecin virtuel en y ajoutant l'analyse de paramètres vocaux : il est désormais possible de détecter des indices permettant d'évaluer l'état du locuteur directement dans sa voix (Cummins *et al.*, 2018). Les avantages de cette méthode sont nombreux : elle ne nécessite pas de matériel ou de calibration complexe, ne repose pas sur des capteurs spécifiques et elle peut être mise en place dans des environnements variés, permettant ainsi un suivi régulier et non invasif des patients. La détection de la somnolence dans la voix en particulier est un sujet ayant déjà fait l'objet de nombreuses études, avec un pic d'intérêt lors de la compétition ajointe à la conférence Interspeech 2011 sur la détection d'état de locuteur (Schuller *et al.*, 2011) ou plus récemment celle associée à la conférence Interspeech 2019 sur l'estimation continue de la somnolence (Schuller *et al.*, 2019).

Les présents travaux, basés sur l'exploitation du corpus TILE (Martin *et al.*, 2020) enregistré au centre hospitalier universitaire de Bordeaux, se distinguent des précédents par trois aspects principaux.

Tout d'abord, les sujets enregistrés dans les bases de données associées aux compétitions (resp. le *Sleepy Language Corpus* et le corpus SLEEP pour Interspeech 2011 et Interspeech 2019) sont tous des sujets sains placés en privation de sommeil. Le médecin virtuel est destiné à une population se plaignant de somnolence diurne excessive (SDE) ayant potentiellement pour origine une maladie du sommeil : non seulement ces patients ont une perception de leur somnolence qui est différente de celle des

sujets sains, mais ils souffrent généralement de facteurs de comorbidité tels que la fatigue ou des troubles de l'humeur, susceptibles de venir interférer avec la mesure de la somnolence dans la voix.

Ensuite, l'élaboration du médecin virtuel nécessite une collaboration étroite avec le milieu médical, dont le but n'est pas tant l'implémentation d'un classificateur que la compréhension des phénomènes liés à la somnolence qui s'expriment dans la voix. La plupart des systèmes de l'état de l'art utilisent les marqueurs vocaux fournis lors des compétitions (calculés grâce à la boîte à outils openSMILE (Eyben et Schuller, 2015)), qui sont non seulement très nombreux (4 368) mais ne sont interprétables que pour des spécialistes en traitement du signal vocal. Nous souhaitons donc proposer un ensemble de marqueurs vocaux et une stratégie de classification permettant de conserver le sens des marqueurs vocaux pour pouvoir les lier, dans le cas de performances de classification acceptables, à des processus neuromoteurs ou cognitifs.

Enfin, nous désirons suivre la somnolence des patients lors de leur utilisation du médecin virtuel. Cela peut se faire selon deux modalités temporelles : soit une estimation de la somnolence à court terme, c'est-à-dire l'état du sujet sur des courtes périodes de temps, soit un suivi de la somnolence à plus long terme, qui dénote l'état habituel de somnolence du locuteur sur des échelles temporelles plus grandes et qui est un marqueur de maladies neuropsychiatriques (facteur « trait » du patient). Les deux échelles temporelles de somnolence peuvent être mesurées selon deux modalités : de manière objective, par électroencéphalogramme (EEG), ou de manière subjective, par le biais de questionnaires que remplissent les patients. À notre connaissance toutes les études menées jusqu'alors – à de rares exceptions près – se sont concentrées sur la détection de la somnolence subjective à court terme, souvent mesurée au moyen d'un questionnaire médical subjectif de somnolence comme le questionnaire de somnolence de Karolinska – KSS (Åkerstedt et Gillberg, 1990). Cette mesure de la somnolence est à la fois subjective et instantanée : l'échelle mesure l'état ressenti de somnolence du locuteur sur une période d'une dizaine de minutes. Pour un suivi à plus long terme, au contraire, nous cherchons à estimer un marqueur « trait » du locuteur, valable sur une longue durée. Pour cela, le seul corpus proposant de telles mesures est la base TILE, décrite à la section 2.

Notre objectif est donc double : d'une part, proposer une approche basée sur des marqueurs vocaux interprétables permettant la détection de la somnolence subjective à court terme dans la voix, pour des sujets sains (SLC) et des sujets atteints de SDE (base TILE). D'autre part, mettre au point une méthodologie pour la détection de la somnolence à long terme dans la voix, chez les patients souffrant de SDE (base TILE). Pour cela, nous proposons deux approches : l'une basée sur les marqueurs vocaux liés à la qualité de la voix du locuteur, l'autre sur les erreurs de lecture.

Cet article est organisé de la façon suivante. Les corpus utilisés dans cette étude sont présentés dans la section 2 tandis que les marqueurs vocaux sont introduits dans la section 3. La section 4 propose une taille minimale des échantillons audio pour la détection de la somnolence avec ces marqueurs audio. Les sections 5 et 6 présentent les méthodologies, résultats et discussions sur la détection de la somnolence à court

terme et à long terme grâce à des marqueurs basés sur la qualité de la voix, tandis que la section 7 introduit un nouveau paradigme de détection de la somnolence à long terme grâce aux erreurs de lecture des locuteurs. La conclusion et nos futurs travaux sur le sujet sont évoqués dans la section 8.

2. Corpus

2.1. *Corpus TILE*

Le corpus TILE (*Multiple Sleep Latency Test – MSLT – database* en anglais) est un corpus contenant les enregistrements de 106 patients ayant des plaintes concernant leur sommeil et venant passer un examen médical pour un suivi ou un diagnostic à la clinique du sommeil du centre hospitalier universitaire de Bordeaux. Durant cet examen, le test itératif de latence d'endormissement – TILE, il est demandé aux patients de faire une sieste de maximum 35 minutes toutes les 2 heures à partir de 9 heures du matin (Littner *et al.*, 2005). Une fois le test lancé, les patients ont 20 minutes pour s'endormir. S'ils y parviennent, le test continue et ils sont réveillés 15 minutes plus tard. Dans le cas contraire, l'itération du test est arrêtée. Les valeurs de TILE, qui correspondent aux latences d'endormissement des patients à chaque sieste, sont donc comprises entre 0 et 20 minutes.

Entre les siestes, les patients sont libres d'effectuer l'activité de leur choix (hors activité physique et consommation de stimulants tels que café ou thé), mais doivent arrêter de fumer au moins 30 minutes avant chaque sieste. Les enregistrements vocaux sont effectués une dizaine de minutes avant chaque sieste, lors de la lecture de textes d'environ 200 mots issus du *Petit Prince* d'Antoine de Saint-Exupéry (de Saint-Exupéry, 1943). Les textes sont différents pour chaque sieste mais identiques pour tous les patients à sieste identique. Tous les enregistrements ont une durée supérieure à 50 secondes. Ce texte a été choisi pour sa simplicité de vocabulaire et de grammaire, tout en ayant un contenu qui ne soit ni stimulant, ni relaxant, afin de ne pas interférer avec le test en cours.

Le corpus TILE permet l'étude de deux types de mesures de somnolence à des échelles temporelles différentes. Les échantillons audio sont étiquetés à la fois avec un questionnaire médical subjectif de somnolence instantanée et une valeur de somnolence objective rendant compte d'un facteur trait du locuteur.

2.1.1. *Somnolence subjective instantanée*

La somnolence subjective instantanée est mesurée dans la base TILE grâce à la version française du questionnaire médical KSS (*Karolinska Sleepiness Scale* (Åkerstedt et Gillberg, 1990)). Son échelle de notation va de 1 – « très éveillé » – à 9 – « très somnolent, avec de grands efforts pour rester éveillé, luttant contre le sommeil ». Il a une précision temporelle de l'ordre d'une dizaine de minutes (la consigne en français du questionnaire précise « au cours des dix dernières minutes »). Dans l'optique d'une classification binaire, il faut définir une limite sur le KSS permettant de distinguer des

enregistrements de voix somnolentes de celles non somnolentes. Si la limite la plus utilisée dans l'état de l'art est de 7,5 (Schuller *et al.*, 2011), nous préférons prendre celle de 7, pour deux raisons : non seulement cette limite correspond alors à un intitulé du questionnaire (« somnolent, mais sans effort pour rester éveillé »), mais une précédente étude sur le SLC a également montré qu'elle permet de meilleurs scores de classification (Martin *et al.*, 2019). Les échantillons associés à un KSS inférieur à 7 seront considérés comme produit par un locuteur non somnolent tandis que ceux associés à un KSS supérieur ou égal à 7 seront considérés comme produits par un locuteur somnolent.

2.1.2. *Détection de facteurs traits dans la voix*

En plus de permettre l'association entre des échantillons isolés et la somnolence subjective instantanée, ce corpus fournit des informations permettant l'estimation de traits propres aux locuteurs. En effet, la somnolence objective est mesurée par le temps d'endormissement à chaque sieste, mesure médicale liée aux variations de performances (Carskadon *et al.*, 1981) et appelée ici « valeur de TILE ». Le TILE est un test médical pour la détection de la narcolepsie lorsque la latence moyenne d'endormissement est inférieure à 8 minutes (Aldrich *et al.*, 1997). Cette mesure donne des informations sur un trait du locuteur, sur sa propension à la somnolence durant une longue période de temps. Nous réutilisons cette limite de 8 minutes pour notre tâche de classification de la somnolence. Par ailleurs, même si les seules mesures utilisées ici sont le KSS et la valeur de TILE, les patients de ce corpus sont largement phénotypés sur leur pathologie à travers de nombreux questionnaires subjectifs de somnolence, fatigue, anxiété, dépression, insomnie, addiction, etc. Ces mesures sont complétées de données physiques telles que la taille, le poids ou encore le tour de cou.

Un bref aperçu de ce corpus est présenté dans le tableau 1. Pour plus d'informations sur ce corpus, sa méthodologie de conception et les différentes mesures collectées, nous redirigeons le lecteur vers l'article le présentant (Martin *et al.*, 2020).

2.2. Sleepy Language Corpus (SLC)

Le SLC (*Sleepy Language Corpus*) est le corpus le plus utilisé dans l'état de l'art pour l'élaboration de systèmes de détection de la somnolence dans la voix (Cummins *et al.*, 2018). Élaboré pour la compétition Interspeech 2011 portant sur la détection de l'état instantané du locuteur dans la voix (Schuller *et al.*, 2011), ce corpus est constitué d'enregistrements de participants effectuant différentes tâches vocales, elles-mêmes conduites en parallèle d'autres études médicales induisant une privation de sommeil des sujets. Les sujets sont des volontaires germanophones et tous les échantillons sont soit en allemand, soit en anglais.

Les informations associées aux enregistrements vocaux sont la tâche effectuée lors de la lecture, le sexe du locuteur, l'affectation de l'échantillon dans la base d'entraînement, de développement ou de test, et la valeur de somnolence correspondante.

Donnée	Femmes	Hommes	Total
Nombre de sujets	63	43	106
Nombre d'échantillons	315	215	530
Âge moyen (écart-type)	33,9 (11,5)	38,7 (16,9)	35,9 (14,1)
Niveau social moyen (écart-type)	6,0 (2,5)	4,6 (2,3)	5,4 (2,5)
KSS moyen (écart-type)	4,6 (1,3)	4,3 (1,2)	4,4 (1,3)
Nombre d'échantillons S (KSS)	72	36	108
Durée totale S (KSS)	1 h 36 m 4 s	49 m 57 s	2 h 26 m
Nombre d'échantillons NS (KSS)	243	179	422
Durée totale NS (KSS)	4 h 58 m 22 s	4 h 31 s	8 h 58 m 53 s
TILE moyenne (écart-type) en minutes	11,8 (4,6)	10,4 (5,1)	11,2 (4,8)
Nombre de sujets S (TILE)	13	15	28
Durée totale d'enregistrement S (TILE)	1 h 20 m 55 s	1 h 39 m 37 s	3 h 32 s
Nombre de sujets NS (TILE)	50	28	78
Durée totale d'enregistrement NS (TILE)	5 h 13 m 30 s	3 h 10 m 52 s	8 h 24 m 22 s

Tableau 1. *Statistiques du corpus TILE. S : somnolent ; NS : non-somnolent*

La valeur de somnolence est la moyenne de trois KSS, un rempli par le patient lui-même, et deux remplis par des annotateurs externes. Nous redirigeons le lecteur vers (Krajewski *et al.*, 2009 ; Golz *et al.*, 2007) pour un descriptif détaillé des conditions expérimentales d'enregistrement et vers (Schuller *et al.*, 2013) pour la liste exhaustive des différents sous-corpus agrégés pour former le SLC.

Afin d'assurer une comparaison valide entre le corpus TILE et le SLC, nous sélectionnons uniquement les tâches de lecture de ce dernier. De plus, après l'étude menée dans la section 4, nous sélectionnons seulement les tâches de lecture dont la taille moyenne des échantillons est supérieure à 8 secondes : la lecture de la fable *Nordwind und Sonne* (version en allemand de la fable *La bise et le soleil*) dont la durée moyenne est de 36,5 secondes ; la lecture de deux simulations de communication de trafic aérien (« flight1 » et « flight2 » de durées moyennes respectives de 9,7 secondes et 13,8 secondes) et la lecture d'une simulation de discours d'un contrôleur de trafic aérien « roger1 » (durée moyenne : 8,5 secondes). Les statistiques de ce sous-corpus sont présentées dans le tableau 2.

3. Marqueurs vocaux

La grande majorité des systèmes de l'état de l'art ayant pour but de détecter la somnolence subjective dans la voix utilisent des marqueurs vocaux calculés avec la boîte à outils openSMILE (Eyben et Schuller, 2015). Les 4 368 marqueurs correspondant à la compétition Interspeech 2011 sur l'état de somnolence du locuteur ne sont malheureusement pas tous interprétables par des non-spécialistes de la voix. Or, l'élaboration d'un outil de détection de la somnolence dans la voix nécessite une collaboration étroite avec des médecins, qui ont besoin de pouvoir relier les marqueurs vocaux à des mécanismes neuromoteurs ou de performances cognitives. Nous avons

Sexe	Classe	**Ent.**	**Dev.**	**Test**	**Total**
Femmes	NS	10	8	9	27
		109 éch.	88 éch.	73 éch.	270 éch.
		4,15 (1,4)	4,0 (1,6)	4,18 (1,4)	4,11 (1,5)
		24 min 23 s	18 m 43 s	17 m 48 s	1 h 54 s
	S	5	6	6	17
		106 éch.	76 éch.	86 éch.	268 éch.
		8,12 (0,5)	8,13 (0,7)	8,21 (0,9)	8,15 (0,7)
		18 m 23 s	15 m 18 s	17 m 59 s	51 m 40 s
Hommes	NS	10	7	9	26
		54 éch.	27 éch.	52 éch.	133 éch.
		4,8 (1,2)	3,9 (1,6)	3,5 (1,8)	4,1 (1,6)
		17min 4s	8 m 17s	15 m 31 s	40 m 53s
	S	4	7	3	14
		33 éch.	56 éch.	30 éch.	119 éch.
		8,7 (0,9)	8,7 (1,0)	8,14(0,9)	8,6 (1,0)
		7 m 8 s	14 m 4s	6 m 41 s	27 m 53 s
Total	NS	20	15	18	53
		164 éch.	115 éch.	125 éch.	404 éch.
		4,3 (1,4)	4,0 (1,6)	3,9 (1,6)	4,1 (1,5)
		41 m 38 s	26 m 59 s	33 m 19 s	1 h 41 m 57s
	S	9	13	9	31
		139 éch.	132 éch.	116 éch.	387 éch.
		8,3 (0,7)	8,4 (0,9)	8,2 (0,9)	8,3 (0,8)
		25 m 31 s	29 m 21 s	24 m 40 s	1 h 19 m 33s

Tableau 2. *Nombre de locuteurs, nombre d'échantillons, KSS moyen (écart-type) et durée cumulée d'enregistrements du sous-corpus de la base SLC contenant uniquement des tâches de lecture. S : somnolent ; NS : non-somnolent ; Ent. : entraînement ; Dev. : développement*

donc élaboré notre propre ensemble de marqueurs vocaux, contenant exclusivement des marqueurs dont l'explicabilité a été mise à l'épreuve avec des médecins et qui peuvent être reliés à des mécanismes physiologiques.

3.1. *Statistiques concernant les parties voisées*

Les marqueurs vocaux sont calculés en deux temps. Tout d'abord, nous extrayons les segments voisés grâce à l'extraction de la fréquence fondamentale par l'algorithme ESPS (Sjölander, 2004), ainsi que l'extraction automatique de segments vocaliques (Pellegrino et Andre-Obrecht, 2000). Le premier sous-groupe de marqueurs est composé de statistiques sur ces segments, tandis que le second sous-groupe contient des marqueurs caractérisant la régularité de la production d'harmoniques sur les seg-

ments voisés. L'ensemble de ces marqueurs est ensuite moyenné pour obtenir un seul groupe de marqueurs audio par échantillon.

Les statistiques obtenues sur les parties voisées et les parties vocaliques reflètent le comportement global du locuteur et sont les suivantes :

– la durée totale des parties voisées (en secondes) ;

– le pourcentage en durée des parties voisées ;

– la durée totale des segments vocaliques (en secondes) ;

– le pourcentage en durée des segments vocaliques.

3.2. *Régularité de la production d'harmoniques sur les segments voisés*

Une fois les parties voisées et les parties vocaliques extraites, nous mesurons la régularité de la production d'harmoniques sur ces segments grâce à des mesures de fréquence fondamentale et de courbes d'intensité :

– F_0MEAN : la moyenne de la fréquence fondamentale sur les segments voisés ;

– F_0VAR : la variance de la fréquence fondamentale sur les segments voisés ;

– F_0SLOPE : le coefficient directeur de l'approximation linéaire de la fréquence fondamentale sur un segment voisé ;

– F_0MAX : le maximum de la fréquence fondamentale sur un segment voisé ;

– F_0MIN : le minimum de la fréquence fondamentale sur un segment voisé ;

– F_0EXTEND : l'amplitude de la fréquence fondamentale sur un segment voisé.

Les mêmes paramètres sont calculés sur les courbes d'intensité (NRJMEAN, NRJVAR, NRJMAX, NRJMIN, NRJEXTEND). Il en résulte 12 paramètres vocaux supplémentaires (6 sur la fréquence fondamentale F_0, 6 sur l'intensité). Nous avons également calculé les équivalents de F_0MEAN, F_0VAR, NRJMEAN et NRJVAR sur les segments vocaliques, ajoutant ainsi 4 paramètres vocaux.

Cet ensemble de paramètres est complété par des paramètres qui ont notamment été utilisés pour caractériser la classification d'attitudes sociales (Rouas *et al.*, 2019) et que nous avons calculés avec la boîte à outils Matlab Covarep (Degottex *et al.*, 2014) que nous avons modifiée pour les calculer seulement sur les segments voisés. Nous complétons ainsi notre ensemble de paramètres avec l'amplitude des harmoniques (H1, H2, H4), l'amplitude des formants (A1, A2, A3), leur fréquence (F1, F2, F3, F4) et leur bande passante (B1, B2, B3, B4), la différence entre les amplitudes des harmoniques (H1-H2, H2-H4), la différence d'amplitude entre les harmoniques et les formants (H1-A1, H1-A2, H1-A3), la *Cepstral Peak Prominence* (CPP) et les rapports harmoniques sur bruit dans différentes plages de fréquences (HNR05, HNR15, HNR25, HNR35). Tous ces paramètres sont moyennés sur chaque enregistrement, ce qui ajoute un total de 24 paramètres à notre ensemble de paramètres vocaux.

Cet ensemble de marqueurs contient ainsi un total de 44 paramètres vocaux.

4. Longueur minimale des échantillons pour la détection de la somnolence grâce à des marqueurs vocaux

Lors de la sélection d'un sous-corpus du SLC, une question jamais soulevée à notre connaissance dans l'état de l'art s'est imposée : quelle est la longueur d'enregistrement audio nécessaire pour permettre la détection de la somnolence dans la voix ?

Pour répondre à cette question, nous avons découpé tous les échantillons audio des deux corpus en tronçons contenant uniquement la première seconde de l'échantillon, uniquement les deux premières secondes de l'échantillon, uniquement les trois premières secondes de l'échantillon, etc. On obtient ainsi des échantillons de taille croissante, sur lesquels on calcule les marqueurs vocaux présentés dans la section 3. Pour éviter un biais qui serait propre à nos marqueurs, nous extrayons également les marqueurs de la conférence Interspeech 2011 grâce à la boîte à outils openSMILE (Eyben et Schuller, 2015) pour comparaison.

Ensuite, nous calculons pour chaque échantillon la similarité cosinus entre le marqueur correspondant au tronçon de taille i secondes et celui correspondant au tronçon de taille $i+1$ secondes, issus du même fichier audio :

$$s_{i,i+1} = \frac{|X_i| \cdot |X_{i+1}|}{||X_i|| \cdot ||X_{i+1}||}$$

Ainsi, quand $s_{i,i+1}$ est proche de 1, X_i est proche de X_{i+1} : l'information supplémentaire apportée par la seconde supplémentaire entre les échantillons i et $i+1$ est faible. Nous calculons la moyenne et l'écart-type des $s_{i,i+1}$ pour tous les échantillons, et nous obtenons le graphe présenté dans la figure 1. Il représente l'information supplémentaire apportée par chaque seconde supplémentaire dans l'échantillon, à partir d'un échantillon vide.

Une première remarque concerne la différence de valeurs entre l'évolution des marqueurs personnalisés et ceux extraits avec openSMILE. En effet, les valeurs de moyenne et d'écart-type de $s_{i,i+1}$ sont très proches respectivement de 1 et de 0 pour les marqueurs extraits avec openSMILE, et ce, quel que soit i. Nous faisons l'hypothèse que cela provient de la différence de taille des ensembles de marqueurs. En effet, dans le cas des marqueurs IS11, une différence franche sur un nombre réduit de marqueurs aura peu d'impact sur la similarité cosinus calculée sur les 4 368 marqueurs, contrairement aux marqueurs personnalisés, au nombre de 44. Cependant, cette différence ne change pas l'interprétation faite de l'évolution de $s_{i,i+1}$. En effet, quel que soit le set de marqueurs ou le corpus, pour une durée d'environ 8 secondes, la moyenne et l'écart-type de $s_{i,i+1}$ commencent à devenir stationnaires : toute information audio supplémentaire n'apporte plus d'information vis-à-vis des marqueurs audio calculés sur une durée plus courte. Nous prenons ainsi cette limite comme limite minimale requise pour la détection de la somnolence dans la voix grâce aux marqueurs vocaux.

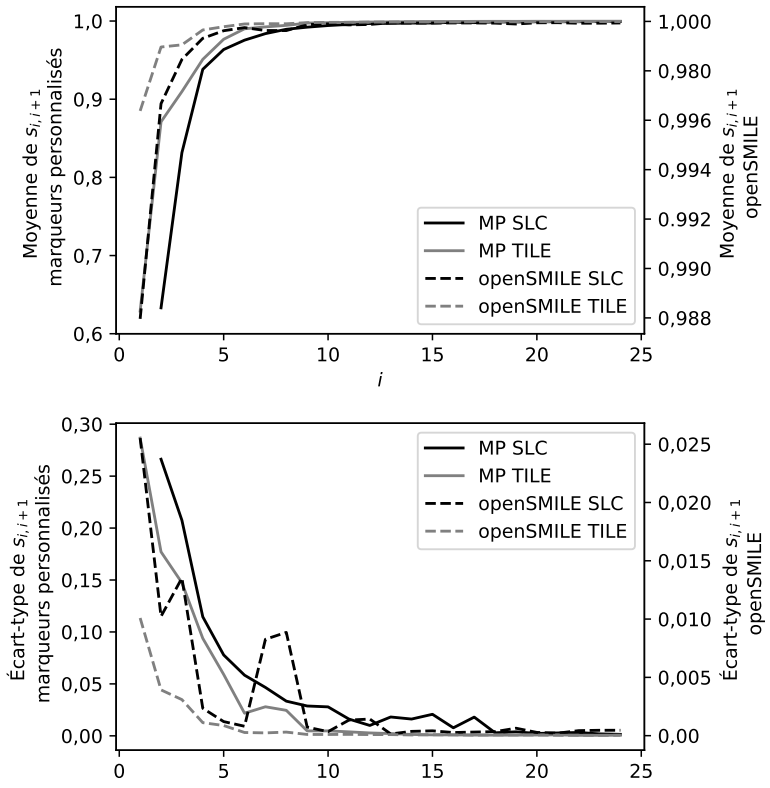


Figure 1. *Moyenne et écart-type de $s_{i,i+1}$ en fonction de i. MP : marqueurs personnalisés*

5. Détection de la somnolence subjective grâce à des marqueurs vocaux

Nous proposons dans cette partie d'estimer la somnolence subjective à court terme grâce à des marqueurs vocaux. La détection de la somnolence à court terme est utile pour anticiper les baisses de performance à court terme, qui peuvent avoir des conséquences dramatiques, par exemple lors de la conduite d'une voiture, le pilotage d'un avion ou à un poste dans lequel l'attention est critique (aiguilleur du ciel ou responsable d'une centrale nucléaire, ar exemple). À la fois dans le SLC et dans le corpus TILE, cette somnolence à court terme est mesurée grâce au questionnaire subjectif KSS. Il existe cependant une différence notable entre l'annotation des deux corpus : si dans le corpus TILE le KSS est rempli uniquement par le sujet avant l'enregistrement audio, celui du SLC est la moyenne entre un questionnaire rempli par le sujet lui-même et de deux annotateurs externes (assistants médicaux) entraînés auparavant à évaluer la somnolence.

Ce problème a été introduit lors de la compétition proposée au sein de la conférence Interspeech 2011 sur la classification d'état du locuteur, dont le meilleur système achevait une performance de 76,4 % (Huang *et al.*, 2014).

5.1. *Méthodologie*

La méthodologie employée pour calculer les performances sur les deux corpus est représentée dans la figure 2 et se décompose de la manière suivante :

– centrage des paramètres vocaux par locuteur. En soustrayant la moyenne des marqueurs vocaux d'un locuteur à tous les marqueurs de ce sujet, on élimine les facteurs propres au locuteur (sexe, âge, physionomie des voies respiratoires...) et on garde uniquement les variations instantanées des paramètres vocaux, qui ne sont plus pollués par des marqueurs traits s'exprimant dans la voix. Cette méthodologie semble d'autant plus pertinente du fait que l'on cherche à estimer la somnolence subjective à court terme et non un état général de somnolence sur le long terme du locuteur ;

– calcul pour chaque marqueur vocal de la corrélation (ρ de Spearman) entre le marqueur et la mesure de somnolence (KSS). Cela permet d'ordonner les marqueurs vocaux du plus corrélé au moins corrélé avec la mesure de somnolence. Par ailleurs, travailler avec des méthodes statistiques permet, contrairement à des techniques de réduction « classiques », comme l'analyse en composantes principales ou l'analyse en composantes indépendantes, de conserver le sens associé aux marqueurs vocaux et ainsi, postérieurement, de lier somnolence et manifestions physiologiques par l'intermédiaire de ces marqueurs. Ce calcul se fait sur l'ensemble entraînement et développement ;

– sélection du nombre de marqueurs et des paramètres optimaux du classificateur. Pour cela, nous calculons les performances du système (sur la base d'entraînement *vs* la base de développement) pour les 1, 2, ..., 44 marqueurs vocaux précédemment triés, et nous conservons le nombre de marqueurs vocaux et les paramètres du classificateur fournissant les meilleures performances. Le classificateur utilisé est un séparateur à vastes marges (SVM), dont les paramètres sont le type de noyau (linéaire ou gaussien), et les paramètres C et γ. Durant cette phase, les performances sur le corpus TILE sont mesurées avec le score F1 (moyenne géométrique de la précision et du rappel) en raison de la validation croisée qui laisse trop peu d'échantillons dans la base de développement pour que le score de rappel non biaisé (SRN), utilisé pour calculer les performances dans la suite, soit pertinent ;

– les paramètres C et γ obtenus lors de l'étape 3 sont utilisés pour entraîner le SVM sur le sous-corpus entraînement et développement et nous obtenons ainsi les classes de somnolence estimées de chaque échantillon du sous-corpus de test.

Le SLC est déjà divisé en sous-corpus d'entraînement, de développement et de test, mais ce n'est pas le cas pour la base TILE. Nous utilisons donc une validation croisée qui exclut à chaque itération un locuteur qui servira de test, puis les locuteurs restants sont divisés en bases d'entraînement et de développement (respectivement

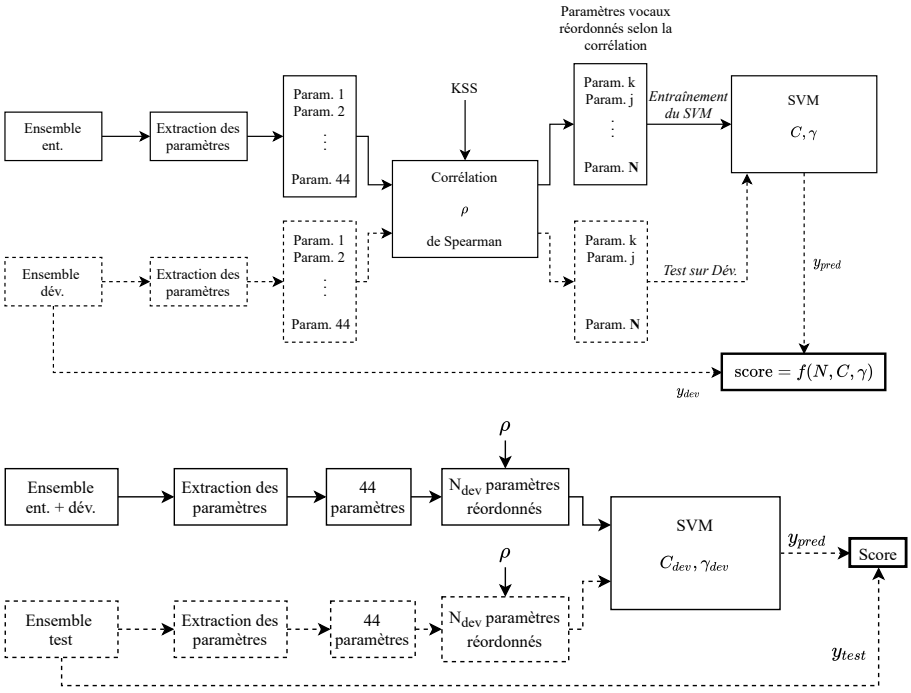


Figure 2. *Schéma du système proposé. Ent. : entraînement ; Dev. : développement*

quatre cinquièmes et un cinquième des locuteurs restants), ayant les mêmes distributions en termes de sexe, d'âge et de somnolence.

À chaque itération de la validation croisée, les classes estimées des échantillons du locuteur précédemment exclus pour le test sont ajoutées dans une matrice de confusion globale, sur laquelle le score global est calculé. Par ailleurs, en raison du faible nombre d'échantillons et de leur déséquilibre entre les deux classes S et NS, nous appliquons un suréchantillonnage de la classe minoritaire grâce au *Synthetic Minority Over-sampling Technique*, SMOTE (Chawla *et al.*, 2002) implémenté dans la boîte à outils Python Sklearn (Pedregosa *et al.*, 2011). Les résultats sont présentés dans le tableau 3.

5.2. *Résultats et discussion*

Une première analyse de ces résultats montre que les performances calculées sur la base TILE sont largement inférieures à celles calculées sur le SLC, avec ou sans centrage des marqueurs par locuteur. Par ailleurs, il est intéressant de remarquer que les performances sur la base TILE sont largement inférieures lorsque les marqueurs sont centrés que lorsqu'ils ne le sont pas (avec une différence de presque 8 %). Le

Réf.	Corpus	Système	SRN
(Huang *et al.*, 2014)	SLC	-	76,4 %
(1a)	SLC	avec centrage	**77,6 %**
(1b)	SLC	sans centrage	66,8 %
(1c)	SLC (locuteurs filtrés)	avec centrage	72,4 %
(1d)	base TILE	avec centrage	48,7 %
(1e)	base TILE	sans centrage	55,6 %
(1f)	base TILE	limite KSS = 5	56,3 %

Tableau 3. *Résultats des systèmes de classification pour la classification de somnolence subjective instantanée*

phénomène inverse est observé sur la base SLC : le système avec centrage par locuteur donne des performances supérieures de plus de 10 % au système sans centrage.

Nous formulons trois hypothèses pour expliquer ces résultats. Premièrement, chaque locuteur du SLC produit en moyenne 13 enregistrements correspondant aux tâches de lecture précédemment sélectionnées (nombre moyen d'enregistrements de tâches de lecture par patient : 13,4 ; écart-type : 19,4), ce qui est plus que ceux du corpus TILE qui sont limités à 5 enregistrements par les conditions expérimentales. Il est intéressant de noter que 5 patients sur les 94 du corpus ont produit à eux seuls 359 des 791 échantillons. Deux de ces locuteurs sont dans la base d'entraînement (n° 38 et n° 39), comptant respectivement 56 et 95 échantillons. Deux autres comptant 36 et 75 échantillons sont dans la base de développement (les n° 40 et n° 41) tandis que le dernier locuteur (n° 42) est dans la base de test et compte 96 échantillons. Nous pensons que le très grand nombre d'échantillons par locuteur dans la base SLC permet une meilleure estimation des traits vocaux lors du centrage des marqueurs par locuteur, ce qui induit un meilleur centrage et donc de meilleures performances. Pour vérifier cette hypothèse, nous avons réappliqué la procédure décrite à la section 5.1, mais sans les locuteurs précédemment ciblés. Cela conduit à un score de 72,4 % (1c), ce qui est à peine plus que 5 % de moins que le système (1a). De plus, ce score reste très supérieur à celui obtenu sur la base TILE avec la même méthodologie : si elle introduit un biais, la présence de locuteurs produisant de très nombreux échantillons dans la base SLC n'explique pas toutes les différences entre les deux systèmes.

Ces observations conduisent à une deuxième hypothèse qui concerne le ratio entre somnolents et non somnolents dans la base TILE, qui est très faible (à peine un cinquième des échantillons correspondent à un KSS supérieur à 7). En effet, un fort déséquilibre entre les classes, malgré l'augmentation de données par suréchantillonnage de la classe minoritaire, empêche une généralisation correcte des classificateurs. En abaissant la limite pour séparer les deux classes à 5 (« ni éveillé, ni somnolent »), on obtient une répartition plus équilibrée de 251 échantillons S contre 279 NS. En réappliquant la même méthodologie que précédemment avec ce nouvel étiquetage, les performances du système augmentent de manière anecdotique (1f) : à peine 0,7 de plus

que le score avec une limite pour le KSS de 7 (1d). Le déséquilibre entre les classes du KSS ne semble donc pas être la source majoritaire des erreurs du classificateur.

Nous formulons donc une troisième hypothèse qui concerne la validité de la mesure de somnolence dans la base TILE. En effet, si le score au KSS est corrélé à l'activité électroencéphalographique des sujets sains (Kaida *et al.*, 2006) comme c'est le cas dans le SLC, les patients souffrant de maladie du sommeil ont une mauvaise perception de leur somnolence subjective (Sangal, 1999). Une observation semblable avait été faite dans l'article présentant le corpus (Martin *et al.*, 2020). Ainsi, le KSS relevé dans la base TILE ne mesure pas les mêmes phénomènes sur les patients de la base TILE que sur les sujets sains du corpus SLC. Les bonnes performances des marqueurs sur le SLC tendent donc à confirmer que ceux-ci sont pertinents pour la détection de la somnolence subjective, mais aussi que ces marqueurs ne sont pas adaptés pour la détection du phénomène mesuré par le KSS sur les patients de la base TILE.

6. Estimation de la somnolence objective sur le long terme grâce à des marqueurs vocaux

Le suivi médical des patients souffrant de SDE peut tirer bénéfice de la détection de la somnolence à court terme, mais aussi à long terme, pour permettre aux médecins de suivre sur de longues plages de temps les variations des marqueurs de traits de somnolence des locuteurs. La détection d'une telle somnolence s'appuie sur le fait que dans le corpus TILE, chaque locuteur est enregistré cinq fois, à des moments différents de la journée. Cette partie a donc pour objectif de classer non plus les échantillons indépendamment les uns des autres mais les locuteurs entre somnolents (TILE moyenne inférieure ou égale à 8 minutes) et non-somnolents (TILE moyenne supérieure à 8 minutes) grâce aux enregistrements de leurs cinq siestes. La limite de 8 minutes sur la moyenne des latences d'endormissement est une limite médicale utilisée dans le diagnostic de nombreuses maladies telles que la narcolepsie par exemple (Aldrich *et al.*, 1997).

6.1. *Méthodologie et résultats*

La première intuition pour estimer la classe de somnolence des locuteurs déterminée par leur valeur moyenne des latences d'endormissement au TILE est de faire la moyenne des cinq jeux de marqueurs de chaque locuteur et d'effectuer la classification directement grâce à un unique ensemble de marqueurs moyens par locuteur. En utilisant la même validation croisée isolant un locuteur pour le test et la même procédure de sélection des marqueurs grâce à la corrélation de Spearman, on obtient un score final d'à peine 50 % (2a). Ce paradigme divisant le nombre d'échantillons par 5, la réduction drastique du nombre d'échantillons pourrait être la cause de ce faible résultat.

Une autre méthode s'appuie sur le fait que l'on a cinq enregistrements pour chaque locuteur et reprend la méthodologie précédemment détaillée dans la section 5.1. Pour chaque itération de la validation croisée, une fois le classificateur entraîné sur les itérations prises de manière indépendante, nous calculons les probabilités d'appartenance à chaque classe de somnolence des cinq enregistrements du locuteur de test. Nous moyennons ensuite ces cinq probabilités pour estimer la classe de somnolence du locuteur de test, que nous rajoutons dans une matrice de confusion globale.

Réf.	Sélection des marqueurs	limite TILE	SRN
(2a)	Moyenne des paramètres (Spearman)	8	50,2 %
(2b)	Moyenne des paramètres (Mann-Whitney)	8	54,8 %
(2c)	Spearman	8	45,6 %
(2d)	Mann-Whitney	8	53,6 %
(2e)	Mann-Whitney	13	**63,8 %**

Tableau 4. *Résultats des systèmes de classification pour la détection de la somnolence à long terme objective sur la base TILE*

L'application de cette méthodologie conduit à un SRN de 45,6 % (2c), ce qui est en dessous des performances qui seraient obtenues en tirant la classe de somnolence au hasard. Pour tenter d'améliorer ces résultats, nous conservons la méthodologie précédente et nous testons une autre méthode de sélection des marqueurs, basée sur le test statistique de Mann-Whitney au lieu de la corrélation de Spearman : au lieu de classer les marqueurs par leur corrélation à la mesure de somnolence, nous les classons par leur pouvoir discriminant entre les deux classes. En effet, plus le U du test de Mann-Whitney est faible, plus les distributions S et NS du marqueur étudié sont différentes (2d). Cette nouvelle approche permet un score de classification atteignant 53,6 %, ce qui représente une augmentation de plus de 8 % du score de classification.

De même que dans la partie 5, nous retestons notre système avec une autre limite pour séparer les deux classes de somnolence selon la valeur de TILE moyenne des patients. Afin d'avoir un meilleur équilibre entre les classes, nous proposons la limite de 13 minutes. En réappliquant les systèmes précédents avec cette nouvelle limite, nous obtenons un score de presque 64 % (2e), ce qui représente une amélioration de presque 10 % par rapport au système précédent.

Ce score reste malgré tout trop faible pour une utilisation en situation réelle, qui nécessiterait 80 % ou 85 % de performances pour une utilisation clinique. Nous faisons deux hypothèses pour expliquer ces résultats. D'une part, de même que le stress ou les émotions peuvent influencer l'expression de la somnolence immédiate dans la voix, l'anxiété, la dépression, et une multitude d'autres facteurs propres au locuteur peuvent également polluer les marqueurs vocaux utilisés pour la détection de la somnolence. Le corpus étant composé d'enregistrements de patients souffrant de SDE, la plupart ont des facteurs de comorbidité qui pourraient influencer leur voix.

D'autre part, du point de vue de la détection de la somnolence du locuteur, la base de données se réduit à 106 patients, ce qui est relativement faible, à la fois pour

l'entraînement et le calcul des performances. Par ailleurs, le fait de moyenner les probabilités des cinq échantillons de manière égale masque l'éventuelle importance que pourraient avoir certaines siestes par rapport à d'autres. Une étude plus approfondie de cette question semble nécessaire pour permettre l'estimation de la somnolence du locuteur grâce aux marqueurs vocaux de manière plus fine, ce qui pourrait mener à une meilleure compréhension des phénomènes mis en jeu et de meilleures performances.

7. Estimation de marqueurs de traits grâce aux erreurs de lecture

Une nouvelle approche pour la détection de la somnolence concerne l'utilisation des erreurs effectuées lors de la lecture des textes de la base TILE. En effet, si les marqueurs vocaux peuvent être liés à des processus neuromusculaires (Krajewski et Kroger, 2007), nous faisons l'hypothèse que les erreurs de lecture sont des marqueurs pertinents de l'influence de la somnolence sur les performances cognitives nécessaires à la lecture. Cette partie traitant de la détection de marqueurs de traits des locuteurs sur le corpus TILE, les patients seront dits « somnolents » si la valeur moyenne de leur latence d'endormissement au TILE est inférieure ou égale à 8 minutes.

7.1. *Liste des erreurs de lecture*

Afin de différencier différents comportements de lecture nous avons retenu quatre catégories d'erreurs, que nous avons voulues relativement générales afin d'obtenir un nombre suffisant d'observations dans chaque catégorie. Les erreurs prises en considération sont les suivantes :

– les achoppements (Ach) : « hésitation, coupure, dans le rythme de la parole » (Brin *et al.*, 2018). Ces erreurs sont un reflet de la capacité d'assemblage du lecteur, c'est-à-dire sa capacité de mettre bout à bout des syllabes pour former un mot. Ainsi, lorsque le lecteur commence la lecture d'un mot, s'arrête, et se reprend, le processus d'assemblage a été interrompu, causant un achoppement. Nous n'avons pas pris en compte les arrêts entre les mots mais seulement les arrêts qui se produisent au milieu d'un mot, ou les allongements artificiels de certaines voyelles, qui témoignent d'une hésitation. Dans le cas de la reprise d'une phrase ou d'un bout de phrase, un seul achoppement est compté, quelle que soit la longueur de la reprise ;

– 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 achoppements, les paralexies reflètent les erreurs d'adressage du lecteur. La capacité d'adressage est le fait de lire un mot dans sa globalité, sans le découper en syllabes ou le déchiffrer, dont les paralexies sont des erreurs symptomatiques. Nous avons généralisé cette catégorie à toute prononciation d'un mot, existant ou non, qui est lu à la place du mot correct. Les télescopages (oublis d'une ou plusieurs syllabes dans un mot) sont donc inclus dans cette catégorie ;

– les oublis de mots (O) : cette erreur est comptée lorsque le lecteur oublie de lire un mot et passe directement au début du mot suivant ;

– les additions de mots (Add) : cette erreur est comptée lorsque le lecteur ajoute un mot qui n'était pas dans le texte original.

Si un locuteur se reprend après une paralexie, un oubli ou une addition, aucun achoppement supplémentaire n'est compté, sauf s'il se trompe lors de la reprise.

7.2. *Sensibilité des erreurs de lecture à la somnolence*

Afin de mesurer si les erreurs élaborées précédemment varient avec la somnolence, les distributions du nombre total de chaque type d'erreur par locuteur chez les patients somnolents et non somnolents sont représentées dans la figure 3 (moyenne $\pm$ SEM – erreur standard de la moyenne). Sur tous les types d'erreurs, les patients somnolents font plus d'erreurs que leurs homologues non somnolents (tests de Mann-Whitney. Ach : $U = 873, p = 8,1 \times 10^{-2}$; O : $U = 738, p = 7,8 \times 10^{-3}$; Add : $U = 847, p = 5,0 \times 10^{-2}$; Plx : $U = 759, p = 1,2 \times 10^{-2}$; total : $U = 765, p = 1,4 \times 10^{-2}$).

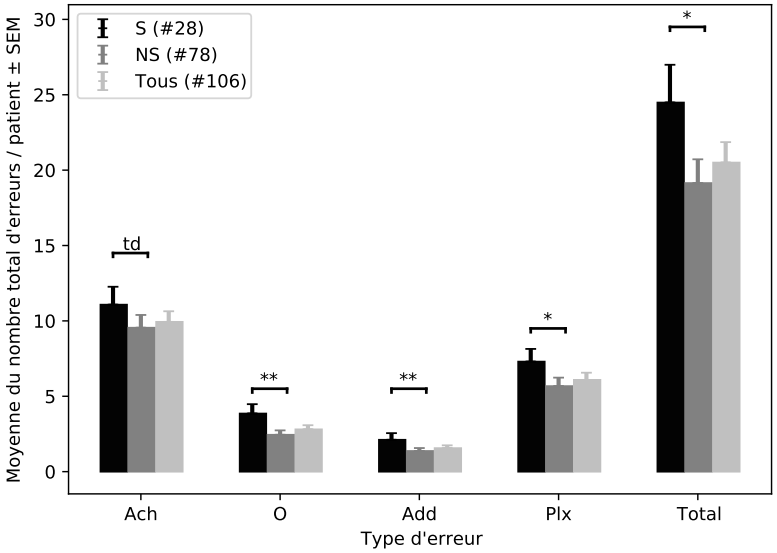


Figure 3. *Distribution du nombre total d'erreurs par locuteur (moyenne $\pm$ SEM). Ach : achoppements, O : oublis, Add : additions, Plx : paralexies. Tests de Mann-Whitney (td : $p < 10^{-1}$, * : $p < 5 \times 10^{-2}$, ** : $p < 10^{-2}$)*

7.3. *Étude des sources d'influence de production d'erreurs*

Il est nécessaire de pouvoir séparer l'influence de la somnolence des facteurs extérieurs pouvant provoquer ces erreurs. Ces facteurs peuvent être les différences entre les textes (différence de taille, quantité de dialogues, difficulté) ou les différents facteurs temporels tels que la prise de repas ou la fatigue accumulée de la journée. Dans la suite, « influence de l'itération » désignera l'influence de tels facteurs sur les erreurs produites par le locuteur. Afin de séparer la contribution de la somnolence de celle de l'itération, nous avons appliqué à nos données une ANOVA multivariée à mesures répétées avec R (R Core Team, 2017).

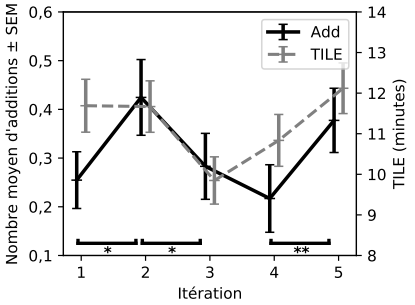
Les additions, oublis et paralexies sont représentés avec les valeurs de TILE et le KSS en fonction des itérations du TILE dans la figure 4. Les achoppements avec les valeurs de TILE et le KSS sont représentés dans la figure 5.

7.3.1. *Additions de mots*

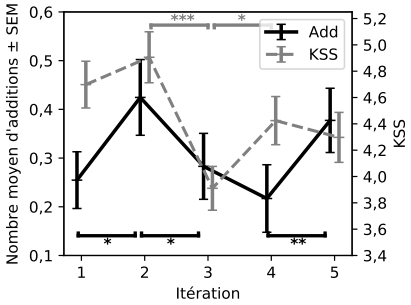
Une ANOVA prenant en compte le nombre d'additions, les différentes échelles de somnolence, et l'influence de l'itération montre que la somnolence objective a une influence quasiment significative sur les variations inter-sujets du nombre d'additions (influence de la valeur de TILE sur les variations inter-sujets : $F = 3,5 ; p = 6,6 \times 10^{-2}$) et que la somnolence subjective a un effet quasiment significatif sur les variations inter-sujets du nombre d'additions (influence du KSS sur les variations inter-sujets : $F = 3,8 ; p = 5,2 \times 10^{-2}$). 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 sur chaque sujet au cours du temps (influences conjointes 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 faisons donc l'hypothèse que les variations du nombre d'additions sont principalement dues à celles des somnolences objectives et subjectives, et qu'elles sont donc indépendantes du texte et des autres effets d'itération.

7.3.2. *Oublis de mots*

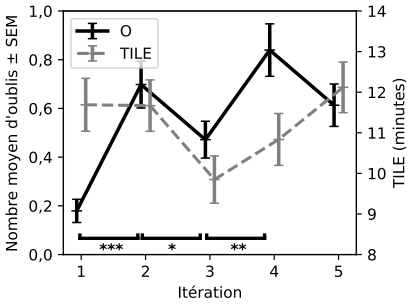
De même, la somnolence objective a une influence sur les variations de nombre d'oublis de mots (influence de la valeur de TILE sur les variations inter-sujets : $F = 3,2 ; p = 7,5 \times 10^{-2}$) tandis que la somnolence subjective a une influence sur les variations du nombre d'oublis de mots (influence du KSS sur les variations intra-sujets : $F = 3,1 ; p = 8,1 \times 10^{-2}$). En revanche, contrairement aux additions, ces erreurs subissent également les effets de l'itération (effet de l'itération sur les variations intra-sujets : $F = 12,0 ; p = 3,0 \times 10^{-9}$).



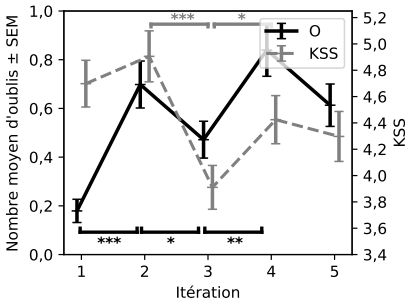
(a) *Additions et TILE en fonction des itérations*



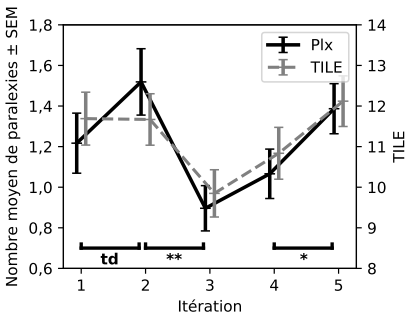
(b) *Additions et KSS en fonction des itérations*



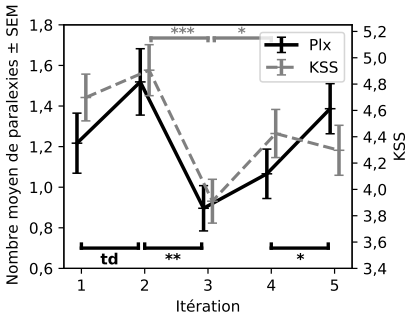
(c) *Oublis et TILE en fonction des itérations*



(d) *Oublis et KSS en fonction des itérations*



(e) *Paralexies et TILE en fonction des itérations*

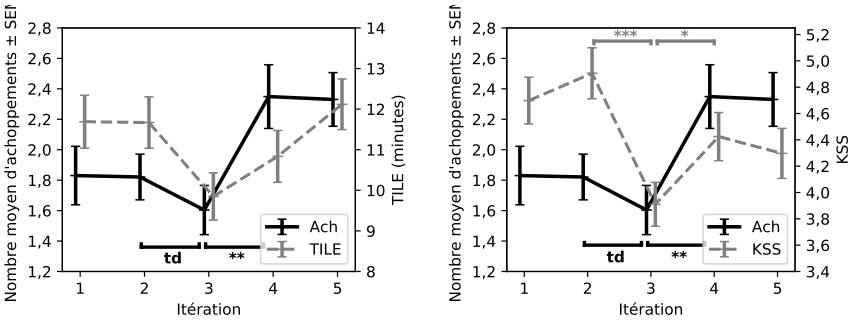


(f) *Paralexies et KSS en fonction des itérations*

Figure 4. *Additions (a, b), oublis (c, d) et paralexies (e, f) comparées au TILE et au KSS (moyenne ± SEM). Tests de Mann-Whitney*
*(td : $p < 1 \times 10^{-1}$, * : $p < 5 \times 10^{-2}$,** : $p < 10^{-2}$, *** : $p < 10^{-3}$)*

7.3.3. *Achoppements*

De même que pour les oublis de mots, l'étude des divers effets ayant une influence sur les variations du nombre d'achoppements permet de mettre en évidence



(a) *Achoppements et TILE en fonction des itérations*

(b) *Achoppements et KSS en fonction des itérations*

Figure 5. *Achoppements comparés au TILE et au KSS (moyenne ± SEM). Tests de Mann-Whitney*
(td : $p < 1 \times 10^{-1}$*, * :* $p < 5 \times 10^{-2}$*, ** :* $p < 10^{-2}$*)*

une influence significative du KSS ($F = 4,2\ ;\ p = 4,2 \times 10^{-2}$) et de l'itération ($F = 7,4\ ; p = 9,4 \times 10^{-6}$) sur les variations intra-sujets de ce type d'erreurs. En revanche, la valeur de TILE ne semble avoir aucune influence sur ce type d'erreurs.

7.3.4. *Paralexies*

Les variations de paralexies au cours des itérations ne semblent être influencées que par les facteurs d'itération (effet de l'itération sur les variations intra-sujets : $F = 6,0\,; p = 1,1 \times 10^{-4}$).

7.4. *Estimation de la somnolence du locuteur grâce aux erreurs de lecture*

Nous utilisons les erreurs de lecture précédemment élaborées ayant comme facteur d'influence la somnolence comme marqueurs pour un système de classification d'état du locuteur. Pour cela, nous concaténons les erreurs de lecture des cinq siestes pour chaque locuteur, dont nous nous servons comme entrée à un classificateur (SVM) selon la même procédure que dans la section 5.1 pour estimer le niveau de somnolence objective. Les matrices de confusion correspondantes sont représentées dans le tableau 5 (gauche) dont le score de rappel non pondéré atteint 78,7 %. Par ailleurs, les paralexies n'étant pas liées à la somnolence mais uniquement à des effets d'itération, la même procédure sans considérer ce type d'erreurs conduit aux matrices de confusion présentées dans le tableau 5 (droite). Le SRN associé est de 82,6 %.

TILE (8)	S_{pred}	NS_{pred}
S_{th}	23	4
NS_{th}	22	57

SRN = 78,7 %

TILE (8)	S_{pred}	NS_{pred}
S_{th}	20	7
NS_{th}	7	72

SRN = 82,6 %

Tableau 5. *Matrices de confusion et scores de rappel non pondérés des classificateurs utilisant les erreurs de lecture comme marqueurs de la somnolence*

7.5. *Discussion*

Les erreurs de lecture semblent de bons marqueurs de l'état de somnolence objective des locuteurs. Cependant, en raison de leur définition ou du texte incitant ou non ces erreurs, elles n'ont pas toutes la même importance dans la détection de la somnolence. En effet, en moyennant les coefficients attribués par le SVM aux différentes erreurs selon les différentes itérations de la validation croisée, on obtient les quatre marqueurs suivant les plus importants dans la classification : les additions des premières et cinquièmes siestes (ayant des coefficients respectifs $c = 8,0 \times 10^{-2}$ et $c = -1,5 \times 10^{-1}$), les achoppements de la troisième sieste ($c = 9,7 \times 10^{-2}$) et les oublis de la quatrième sieste ($c = -1,3 \times 10^{-1}$).

Ces coefficients sont cohérents avec les résultats de la partie précédente. En effet, les additions, qui ont ici le plus de poids dans la prise de décision du niveau de somnolence, avaient été identifiées comme ne dépendant que de la somnolence objective concernant les variations inter-sujets et ne dépendant pas des effets d'itération. De même, le deuxième marqueur le plus important dans la prise de décision est les oublis, qui malgré les effets d'itération variaient avec la somnolence objective. Enfin, même si l'étude des paralexies n'avait pas mis en valeur d'influence de la valeur de TILE sur la production de ce type d'erreurs, leur contribution dans la prise de décision n'est pas négligeable.

La répartition des coefficients de manière inégale sur les différentes siestes du test pose la question de l'importance relative des itérations pour l'estimation du niveau global de somnolence des locuteurs. En effet, si les additions sont les marqueurs ayant le plus de poids sur la première et la dernière sieste, leur contribution pour la détection de l'état du locuteur lors de la troisième sieste est très faible ($c = 8,8 \times 10^{-3}$), alors que celle des achoppements est la plus importante. Une cause probable de ces disparités est l'inégalité de contenu des textes. En effet, de nombreuses erreurs du corpus se répètent et certains mots sont systématiquement la cible d'une erreur spécifique. Par exemple, « méditatif » est très souvent prononcé « médiatif », causant de nombreuses paralexies à la cinquième sieste, ou encore « Il me répéta alors » est souvent lu à la place de « Et il me répéta alors », causant de nombreux oublis à la troisième sieste. Cela souligne l'aspect capital du choix des textes lus pour l'utilisation des erreurs de lecture en tant que marqueurs de la somnolence.

Par ailleurs, la définition des erreurs a également une influence sur leur robustesse. Nous faisons effectivement l'hypothèse que notre définition des achoppements ne prenant en compte que les interruptions au sein des mots et non entre les mots induit un biais qui empêche le marqueur de refléter l'état de somnolence du locuteur. De même, la fusion des paralexies et des télescopages dans la même catégorie pourrait induire des biais qui réduisent leur intérêt comme marqueurs de la somnolence.

8. Conclusion et perspectives

Pour conclure, après avoir étudié la question de la longueur minimale des échantillons pour la détection de la somnolence, nous avons proposé trois systèmes pour répondre à trois problématiques différentes en relation avec la détection de la somnolence dans la voix. La détection de la somnolence subjective à court terme grâce à des marqueurs vocaux simples donne des résultats satisfaisants lorsqu'il s'agit de sujets sains (sous-corpus de lecture du SLC) mais ne donne pas de bonnes performances lorsque la même méthodologie est appliquée pour des patients souffrant de SDE (base TILE). Nous supposons que cela vient du manque de validité de l'auto-évaluation effectuée par les patients composant le corpus TILE. De même, la méthodologie pour estimer la somnolence à long terme avec des marqueurs vocaux semble souffrir des biais apportés par les comorbidités de la SDE, empêchant le système d'atteindre des performances satisfaisantes. L'estimation de la somnolence à long terme de ces mêmes patients grâce à la valeur de TILE est en revanche très efficace lorsque l'on utilise les erreurs de lecture comme marqueurs de la somnolence : ces marqueurs semblent en effet robustes aux effets parasites qui pourraient s'exprimer dans la voix.

Nos futurs travaux comprendront l'étude approfondie de l'importance relative des siestes pour la détection de la somnolence à long terme. Par ailleurs, les erreurs de lecture étant actuellement annotées manuellement, nous travaillons à l'élaboration d'une détection automatique de ces erreurs de lecture grâce à un système de transcription automatique de la parole utilisant les caractères comme unités de reconnaissance.

Remerciements

Cette étude a été réalisée dans le cadre du projet IS-OSA, financé par la région Nouvelle-Aquitaine, et du projet SOMVOICE, financé par le Labex BRAIN (université de Bordeaux). Nous remercions également le Pr Jarek Krajewski pour nous avoir donné accès au *Sleepy Language Corpus*.

9. Bibliographie

Åkerstedt T., Gillberg M., « Subjective and objective sleepiness in the active individual. », *Int J Neurosci*, vol. 52, p. 29-37, 1990.

Aldrich M. S., Chervin R. D., Malow B. A., « Value of the multiple sleep latency test (MSLT) for the diagnosis of narcolepsy », *Sleep*, vol. 20, n° 8, p. 620-629, 1997.

Brin F., Courrier C., Lederle E., Masy V., *Dictionnaire d'orthophonie - 4ème édition*, orthoedition edn, September, 2018.

Carskadon M. A., Harvey K., Dement W. C., « Sleep Loss in Young Adolescents », *Sleep*, vol. 4, n° 3, p. 299-312, September, 1981.

Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P., « SMOTE : Synthetic Minority Over-sampling Technique », *Journal of Artificial Intelligence Research*, vol. 16, p. 321-357, June, 2002.

Cummins N., Baird A., Schuller B., « Speech analysis for health : Current state-of-the-art and the increasing impact of deep learning », *Health Informatics and Translational Data Analytics*, vol. 151, p. 1-54, 2018.

de Saint-Exupéry A., *Le Petit Prince*, gallimard edn, 1943.

Degottex G., Kane J., Drugman T., Raitio T., Scherer S., « COVAREP — A collaborative voice analysis repository for speech technologies », *IEEE - ICASSP*, p. 960-964, 2014.

Eyben F., Schuller B., « Opensmile », *ACM SIGMultimedia Records*, vol. 6, p. 4-13, 2015.

Golz M., Sommer D., Chen M., Mandic D., Trutschel U., « Feature Fusion for the Detection of Microsleep Events », *Journal of VLSI Signal Processing*, vol. 49, p. 329-342, 2007.

Huang D.-Y., Zhang Z., Ge S. S., « Speaker State Classification Based on Fusion of Asymmetric Simple Partial Least Squares (SIMPLS) and Support Vector Machines », *Comput. Speech Lang.*, vol. 28, n° 2, p. 392-419, 2014.

Kaida K., Takahashi M., Åkerstedt T., Nakata A., Otsuka Y., Haratani T., Fukasawa K., « Validation of the Karolinska sleepiness scale against performance and EEG variables », *Clinical Neurophysiology*, vol. 117, n° 7, p. 1574-1581, 2006.

Krajewski J., Batliner A., Golz M., « Acoustic sleepiness detection : Framework and validation of a speech-adapted pattern recognition approach », *Behavior Research Methods*, vol. 41, n° 3, p. 795-804, 2009.

Krajewski J., Kroger B., « Using prosodic and spectral characteristics for sleepiness detection », *Interspeech 2007*, p. 1841-1845, 2007.

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., « Practice Parameters for Clinical Use of the Multiple Sleep Latency Test and the Maintenance of Wakefulness Test », *Sleep*, vol. 28, n° 1, p. 113-121, 2005.

Martin V. P., Rouas J.-L., Micoulaud-Franchi J.-A., Philip P., « The Objective and Subjective Sleepiness Voice corpora », *12th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, p. 6525-6533, 2020.

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

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., « Scikit-learn : Machine Learning in Python », *Journal of Machine Learning Research*, vol. 12, p. 2825-2830, 2011.

Pellegrino F., Andre-Obrecht R., « Automatic language identification : an alternative approach to phonetic modelling », *Signal Processing*, vol. 80, nº 7, p. 1231-1244, 2000.

Philip P., Dupuy L., Auriacombe M., Serre F., de Sevin E., Sauteraud A., Micoulaud-Franchi J.-A., « Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients », *npj Digital Medicine*, vol. 3, nº 1, p. 2, 2020.

Philip P., Micoulaud-Franchi J.-A., Sagaspe P., De Sevin E., Olive J., Bioulac S., Sauteraud A., « Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders », *Scientific Reports*, vol. 7, nº 1, p. 426-456, 2017.

R Core Team, *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017.

Rouas J.-L., Shochi T., Guerry M., Rilliard A., « Categorisation of spoken social affects in Japanese : human vs. machine », *ICPhS*, 2019.

Sangal R., « 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*, vol. 110, nº 12, p. 2131-2135, 1999.

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., « The INTERSPEECH 2019 Computational Paralinguistics Challenge : Styrian Dialects, Continuous Sleepiness, Baby Sounds & Orca Activity », *Interspeech 2019*, 2019.

Schuller B., Steidl S., Batliner A., Schiel F., Krajewski J., « The INTERSPEECH 2011 Speaker State Challenge », *Interspeech 2011*, p. 3201-3204, 2011.

Schuller B., Steidl S., Batliner A., Schiel F., Krajewski J., Weninger F., Eyben F., « Medium-term speaker states-A review on intoxication, sleepiness and the first challenge », *Comput. Speech Lang.*, vol. 28, nº 2, p. 346-374, 2013.

Sjölander K., The Snack Sound Toolkit, Technical report, 2004.