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

Détection automatique des schwas en français - Application à la détection des troubles du sommeil

Colleen Beaumard[1,2], Vincent P. Martin[3], Yaru Wu[4], Jean-Luc Rouas[1], Pierre Philip[2]
(1) LaBRI/UMR5800, 351 cours de la Libération, 33400 Talence, France
(2) SANPSY/6033, Place Amélie Raba-Léon, 33076 Bordeaux, France
(3) DDP Research Unit, Department of Precision Health, LIH, 1 A-B Rue Thomas Edison, 1445 Strassen, Luxembourg
(4) CRISCO/UR4255, Esplanade de la Paix, 14032 Caen, France
{colleen.beaumard, jean-luc.rouas}@labri.fr, vincentp.martin@lih.lu, yaru.wu@unicaen.fr, pierre.philip@u-bordeaux.fr

RÉSUMÉ

La Somnolence Diurne Excessive affecte négativement les individus et est un problème de santé publique. L'analyse de la parole pourrait aider les cliniciens à la surveiller. Nous nous sommes concentrés sur la détection du schwa /ə/ et avons trouvé un lien entre le nombre d'occurrences annoté manuellement et le niveau de somnolence des patients hypersomnolents d'un sous-ensemble du corpus TILE. Dans un second temps, afin de pouvoir généraliser ces résultats à l'intégralité du corpus, nous avons conçu un système de détection des schwas, robuste à la somnolence. Dans un troisième temps, nous avons étendu notre analyse à deux autres phonèmes supplémentaires /ø/ et /œ/. Nous avons ainsi observé une relation significative entre /ø/ et la combinaison des trois phonèmes et la somnolence subjective à court terme.

ABSTRACT

Automatic Schwa Spotting in French - Application to Pathological Sleepiness Detection

Excessive Daytime Sleepiness negatively affects both individuals and public health. Speech analysis could help clinicians monitor it. We focused on the detection of schwa /ə/ since it is a vowel whose realization is optional in French. We found a link between its manually annotated number and the sleepiness level of hypersomnolent patients of a subset of the MSLTc. We need a schwa spotting system robust to sleepiness to attest to this link in the whole corpus. We compared two different systems outputting either phonemes (baseline) or words (proposed) and found the second system was more robust to sleepiness, using manual annotation as our reference. Two other phonemes were considered due to their closeness to /ə/. Using automatic detection, we found a significant relationship between /ø/ and their combination, and short-term subjective sleepiness.

MOTS-CLÉS : Détection de schwa, Somnolence, Reconnaissance Automatique de la Parole.

KEYWORDS: Schwa Spotting, Sleepiness, Automatic Speech Recognition.

1 Introduction

La Somnolence Diurne Excessive (SDE) est associée à une grande variété de maladies (neurologiques, cardiovasculaires...) et affecte négativement la vie quotidienne et professionnelle des personnes qui en

Actes des 35èmes Journées d'Études sur la Parole, pages 312–321.

souffrent (Barnes & Watson, 2019). Elle augmente également le risque de mortalité (pour les accidents de véhicule (Bioulac *et al.*, 2017), etc.) et de handicap (Jike *et al.*, 2018), qui sont des problèmes de santé publique. Le suivi des symptômes de la SDE est difficile étant donné que les tests sont effectués à l'hôpital pendant une journée, ce qui est à la fois chronophage pour le patient et coûteux pour l'hôpital. Ainsi, les cliniciens ont besoin d'un outil pour collecter ces symptômes régulièrement et dans des conditions écologiques. L'analyse automatique de la parole peut être employée pour surveiller la SDE. La collecte de données vocales peut également être simplifiée en étant effectuée via des smartphones en enregistrant la parole lue ou spontanée.

Plusieurs corpus ont été créés pour détecter automatiquement la somnolence en analysant la parole, chacun utilisant différents types d'annotation. La mesure de somnolence utilisée pour annoter les corpora Sleepy Language (SLC) (Schuller *et al.*, 2011) et SLEEP (Schuller *et al.*, 2019) n'est pas validée cliniquement, donc non-interprétable par les médecins du sommeil. L'annotation du dataset Voiceome (Tran *et al.*, 2022) ne permet pas de distinguer la somnolence de la fatigue (Maclean *et al.*, 1992). Le corpus TILE (Martin *et al.*, 2021b) (Test Itératif de Latence d'Endormissement) est le seul à notre connaissance à mesurer à la fois la somnolence subjective et physiologique avec des mesures validées par les cliniciens.

Bien que de nombreuses tentatives aient été faites pour étudier la détection automatique de la somnolence par l'analyse de la voix, les corpora utilisés sont presque exclusivement composés de parole lue. Ainsi, malgré les performances encourageantes obtenues par les derniers systèmes, certaines des caractéristiques utilisées sont spécifiques à ce type de parole (Martin *et al.*, 2021a, 2022).

Puisque notre objectif est de pouvoir classifier automatiquement la somnolence de manière écologique, nous avons besoin d'évaluer des caractéristiques liées à la somnolence spécifiques à l'analyse de la parole spontanée. Le contrôle neuro-moteur et la planification cognitive peuvent avoir un impact sur la parole élicitée par les sujets somnolents (Harrison & Horne, 1997), nous avons donc décidé de nous concentrer en premier lieu sur l'étude de phonèmes spécifiques qui peuvent être influencés par un tel phénomène, et notamment le schwa.

En français, le schwa /ə/ est une voyelle centrale optionnelle - ce qui signifie que sa prononciation ou son absence ne change pas le sens du mot (Bürki *et al.*, 2011; Durand, 2014). Par exemple, /dəmɛ̃/ « demain » peut être prononcé [dəmɛ̃] ou [dmɛ̃] sans changer son sens. Une pause remplie peut également être produite comme une voyelle de type schwa ([ə ː], « euh »). Sa présence ou son absence est liée à de nombreux facteurs tels que l'accent, la coarticulation, ou le type de parole. L'analyse phonétique peut être appliquée à la fois à la parole spontanée et à la lecture, ce qui pallie au manque de corpus de parole spontanée spécifique à la somnolence. Un corpus nommé Medispeech est actuellement en cours d'enregistrement au Service Universitaire de Médecine du Sommeil (SUMS) du CHU de Bordeaux et contient à la fois de la parole lue et de la parole spontanée. En attendant d'avoir un nombre suffisant d'enregistrement, nous avons utilisé un sous-ensemble du corpus TILE contenant 20 patients avec les plus importantes variations de somnolence à court terme (Martin *et al.*, 2023a) et avons montré que le nombre de schwas annotés manuellement est lié au niveau de somnolence (Beaumard *et al.*, 2023).

Notre objectif est de concevoir un système de détection automatique de schwas – c'est-à-dire détecter la présence (ou non) de schwa dans un signal de parole – sachant qu'elle ne doit pas être impactée par le niveau de somnolence des patients (robuste à la somnolence), et de le valider sur une base de données étiquetée à la main. Ce système peut ensuite être utilisé pour détecter automatiquement le schwa sur un corpus plus large (le corpus TILE) afin de confirmer nos précédents résultats obtenus.

Cet article est organisé comme suit : la Section 2 décrit le corpus utilisé, le sous-ensemble annoté manuellement à des fins d'évaluation ainsi que la méthodologie d'annotation. Dans la Section 3, nous présentons les systèmes de détection des schwas de réference et proposés. Leurs performances sont présentées et comparées dans la Section 4. Enfin, la Section 5 discute de l'utilité du système de détection de schwas conçu pour l'évaluation automatique de la somnolence.

2 Description des données

2.1 Le corpus TILE

Le corpus TILE (Martin *et al.*, 2021b) contient 660 enregistrements de 132 patients hypersomnolents du SUMS du CHU de Bordeaux. Chaque patient a effectué un Test Itératif de Latence d'Endormissement (test éponyme du corpus) consistant en 5 opportunités de sieste de 20 minutes toutes les 2 heures, de 9h à 17h. Leur latence d'endormissement, c'est-à-dire le temps entre le début du test et le moment où le patient s'endort ou la fin du test, est mesurée à chaque occurrence du test. C'est une mesure de référence de la somnolence physiologique à long terme (Arand *et al.*, 2005; Martin *et al.*, 2023b). Avant chaque opportunité de sieste, les patients sont enregistrés en train de lire à voix haute un texte différent du *Petit Prince* (A. de Saint-Exupéry) d'environ 250 mots. Les patients ont également renseigné deux échelles cliniques : l'échelle de somnolence de Karolinska (KSS) et l'échelle de somnolence d'Epworth (ESS). Le premier mesure le niveau de somnolence subjective instantanée sur une échelle de Lickert à 9 points. Les patients remplissent ce questionnaire avant chaque lecture (somnolence subjective à court terme). Le second questionnaire mesure le niveau de somnolence subjective à long terme avec des questions sur des situations quotidiennes. Les patients remplissent ce questionnaire une fois.

Nous utilisons un sous-ensemble du corpus TILE (Martin *et al.*, 2023a) contenant 100 enregistrements de 20 patients hypersomniaques ayant les plus fortes variations de somnolence à court terme (KSS) dans le but de mesurer l'impact du niveau de somnolence sur les performances de détection automatique de schwas.

Ce corpus et les descriptions de son sous-ensemble sont présentés dans la Table 1.

Corpus	**TILE**	**Sous-ensemble**
Durée	14h 10m 12s	2h 11m 47s
#Patients (hommes/femmes)	132 (81/51)	20 (10/10)
#Enregistrements	660	100
KSS moyenne (sd)	4 (2)	5 (2)
Latence d'endormissement moyenne (sd)	12 (6)	14 (6)

TABLE 1 – Description du corpus TILE et du sous-ensemble utilisé

2.2 Annotation manuelle

Nous avons tout d'abord transcrit les textes originaux en utilisant le Lexique 3.83 (New *et al.*, 2004), qui contient la prononciation française standard d'environ 140 000 mots. Le /ə/ étant phonétiquement

proche du /ø/ (« deux ») et du /œ/ (« neuf ») (Fougeron *et al.*, 2007), nous pensons que notre système de détection de schwas pourrait confondre le /ə/ avec l'un d'eux et avons donc étendu les phonèmes considérés au /ə/, /ø/, et /œ/. La Table 2 contient le nombre ainsi que la proportion de /ə/, /ø/, /œ/, et « e » (la combinaison des trois phonèmes) pour chaque texte et pour l'ensemble des textes. Nous avons ensuite annoté manuellement la présence ou l'absence de ces phonèmes sur les enregistrements audio du sous-ensemble.

Phonème	**Prononciation Réalisations**	**Texte 1**	**Texte 2**	**Texte 3**	**Texte 4**	**Texte 5**	**Tous**
Tous	Texte lu	735	726	634	689	734	3 518
/ə/	Texte lu	30 (4,1%)	53 (7,3%)	42 (6,6%)	42 (6,1%)	35 (4,8%)	202 (5,7%)
	Annotées manuellement	28 (3,8%)	48 (6,6%)	38 (5,5%)	37 (5,4%)	28 (3,8%)	179 (5,1%)
/ø/	Texte lu	4 (0,5%)	7 (1,0%)	7 (1,1%)	7 (1,0%)	6 (0,8%)	31 (0,9%)
	Annotées manuellement	4 (0,5%)	6 (0,8%)	6 (0,9%)	6 (0,9%)	8 (1,1%)	30 (0,8%)
/œ/	Texte lu	4 (0,5%)	3 (0,4%)	3 (0,5%)	2 (0,3%)	0 (0,0%)	12 (0,3%)
	Annotées manuellement	3 (0,4%)	3 (0,4%)	3 (0,5%)	2 (0,3%)	0 (0,0%)	11 (0,3%)
« e »	Texte lu	38 (5,2%)	63 (8,7%)	52 (8,2%)	51 (7,4%)	41 (5,6%)	245 (7,0%)
	Annotées manuellement	36 (4,9%)	58 (8,0%)	48 (7,6%)	46 (6,7%)	36 (4,9%)	224 (6,4%)

TABLE 2 – Nombre et proportion de tous les phonèmes (33), /ə/, /ø/, /œ/ et « e » pour chaque texte et pour l'ensemble des textes. Nombre moyen et ratio moyen pour l'annotation manuelle.

3 Systèmes de détection de schwa

3.1 Système de détection de schwas de référence

Notre système de détection de schwas de référence est un système de Reconnaissance Automatique de la Parole (RAP) basé uniquement sur un modèle TDNN-HMM entraîné avec la fonction LF-MMI (Boyer & Rouas, 2019). Le réseau neuronal est un réseau à délai temporel échantillonné avec 7 couches TDNN, chacune ayant 1024 unités. La valeur de pas temporel est réglée sur 1 pour les trois premières couches, 0 pour la quatrième, et 3 pour les suivantes. Le modèle acoustique est basé sur un vecteur MFCC de haute résolution à 40 dimensions concaténé avec un i-vecteur de 100 dimensions (Gupta *et al.*, 2014). Il a été entraîné en utilisant la boîte à outils Kaldi (Povey *et al.*, 2011) sur un sous-ensemble d'ESTER 1 et 2 (Galliano *et al.*, 2009). Nous utilisons directement la sortie phonétique du système pour les analyses. Le nombre de /ə/, /ø/ et /œ/ est ensuite extrait de cette transcription. La Figure 1 schématise le système de référence. Ce système atteint un taux d'erreur de phonèmes de 19,5% sur le corpus Rhapsodie composé de parole préparée, semi-spontanée et spontanée (Martin *et al.*, 2024).

3.2 Système de détection de schwas proposé

L'idée de notre système proposé est d'améliorer les performances de la détection de schwa en supprimant les différences individuelles en utilisant la prononciation standard des mots. Pour ce faire, nous utilisons le système de RAP avec un lexique contenant des variantes de prononciation du dictionnaire phonémique fourni par le Laboratoire d'Informatique de l'Université du Mans (LIUM). De plus, un modèle de langage en mots 3-gram prenant en compte le contexte est implémenté.

Ce dernier a été entraîné sur les corpus ESTER en utilisant la méthode de comptage n-gram de SRILM (Stolcke, 2002) avec une réduction KN et a été limité aux 50 000 mots les plus fréquents dans les textes d'entraînement et le dictionnaire. Ce système atteint un taux d'erreur de mots de 13,7% (Boyer & Rouas, 2019).

Nous utilisons ainsi la sortie en mots de ce système complet de RAP que nous transcrivons ensuite en unités phonétiques en utilisant le Lexique 3.83 (voir la Figure 2). Si un mot n'est pas inclus dans celui-ci, nous l'avons transcrit manuellement. Le nombre d'occurrences détecté de /ə/, /ø/, et /œ/ est ensuite extrait de cette transcription. Cette méthode permet d'associer l'audio à la prononciation standard, ce qui réduirait la différence dans les transcriptions entre le lexique du système complet RAP et le lexique 3.83. Par exemple, le lexique à l'intérieur du système transcrivait le mot « premier » comme /pʁəmje/ alors que la prononciation standard dans le Lexique 3.83 est /pʁømje/.

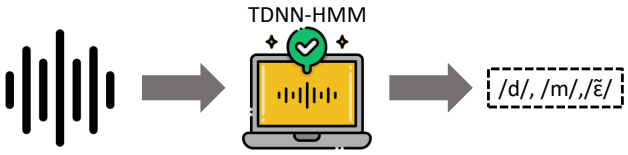

FIGURE 1 – Système de détection de schwas de référence

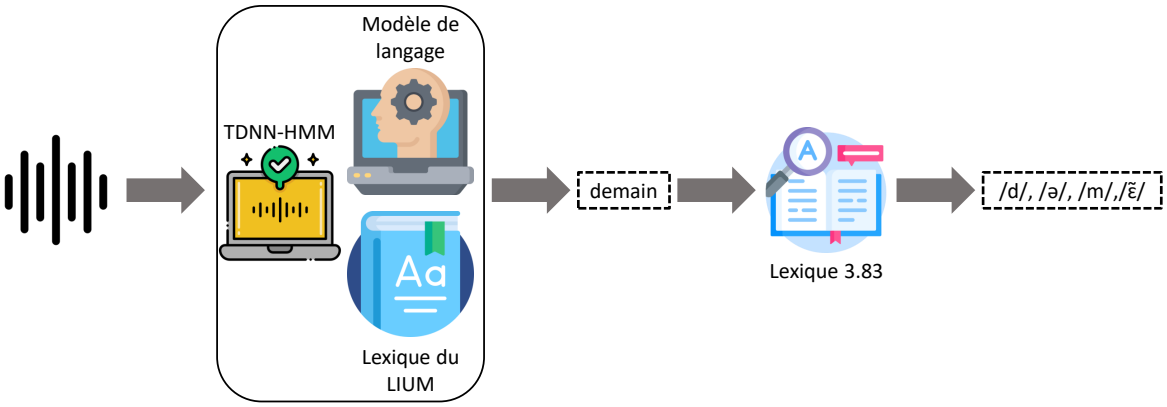

FIGURE 2 – Système de détection de schwas proposé

4 Performances de reconnaissance des phonèmes étudiés

Afin d'évaluer les performances des deux systèmes, nous avons calculé l'Erreur Absolue Moyenne normalisée (%MAE) et la Racine de l'Erreur Quadratique Moyenne (%RMSE) entre le nombre de schwas annotés manuellement et le nombre détecté par chaque système. Plus ces valeurs sont basses, meilleur est le système. La Table 3 référence le %MAE et le %RMSE pour le système initial tandis que la Table 4 référence les mêmes métriques pour le système proposé.

Dans la Table 3, le pourcentage d'erreur de détection du schwa est élevé, indiquant que ce système est peu performant par rapport à l'annotation manuelle, en détectant plus de schwas que leur nombre annoté. De plus, ces performances dépendent du texte (le système initial est plus performant sur le texte

#2 que sur le texte #5). Les performances pour /ø/ et /œ/ sont également faibles, particulièrement pour /ø/, et dépendent elles aussi du texte. La catégorie « e » montre également de faibles performances, avec les mêmes disparités entre les textes.

Phonème	Métrique	Texte 1	Texte 2	Texte 3	Texte 4	Texte 5	Tous
/ə/	%MAE	29,3	11,2	28,4	33,5	50,2	28,3
	%RMSE	35,7	13,5	30,5	36,2	52,0	32,4
/ø/	%MAE	50,0	14,5	44,8	42,6	52,5	40,0
	%RMSE	50,0	14,5	46,3	44,1	42,5	44,6
/œ/	%MAE	2,6	33,3	3,2	20,0	-	12,5
	%RMSE	5,1	36,7	9,7	30,0	-	25,0
e	%MAE	17,5	9,6	16,5	21,5	28,0	17,8
	%RMSE	23,6	11,4	18,6	23,6	30,7	20,9

TABLE 3 – %MAE et %RMSE pour /ə/, /ø/, /œ/ et « e » pour chaque texte et pour l'ensemble avec le système de détection de schwas de référence.

Phonème	Métrique	Texte 1	Texte 2	Texte 3	Texte 4	Texte 5	Tous
/ə/	%MAE	8,9	6,4	11,5	10,2	25,3	11,4
	%RMSE	12,9	9,5	13,4	13,7	30,2	15,8
/ø/	%MAE	0,4	17,1	16,7	7,4	11,3	10,8
	%RMSE	0,1	17,1	21,2	10,3	15,0	16,9
/œ/	%MAE	2,6	6,7	6,5	25,0	-	8,3
	%RMSE	7,7	16,7	12,9	35,0	-	20,8
e	%MAE	7,5	5,9	7,4	7,8	18,8	8,9
	%RMSE	10,6	8,8	9,3	11,7	22,4	12,4

TABLE 4 – %MAE et %RMSE pour /ə/, /ø/, /œ/ et « e » pour chaque texte et pour l'ensemble avec le système de détection de schwas proposé.

Les mêmes métriques d'évaluation calculées sur la sortie du système de RAP proposé sont rapportées dans la Table 4. Les %MAE et %RMSE des trois phonèmes et de leur combinaison ont grandement diminué : la pire %RMSE pour tous les textes avec le système initial était de 44,6% (/ø/) tandis qu'avec le système proposé, il est de 20,8% (/œ/). De plus, même s'il existe encore des disparités entre les textes, elles sont moindres que celles du système initial. Le système proposé est donc plus proche de l'annotation manuelle, donc plus performant, pour la détection de /ə/, /ø/, /œ/, et de leur combinaison. Nous devons maintenant évaluer sa robustesse à la somnolence avant de le considérer comme un système fiable.

4.1 Robustesse du système proposé à la somnolence

Pour mesurer l'effet des textes, de la KSS, et de la latence d'endormissement sur les performances de détection de /ə/, /ø/, et /œ/ par le système proposé, nous avons calculé quatre ANOVA multivariées à mesures répétées pour expliquer les variations intra- et inter-locuteurs des performances de détection des phonèmes avec celles de la somnolence et l'influence des textes. Les résultats de ces ANOVA sont reportés dans la Table 5.

Les différents facteurs n'ont pas d'effet significatif sur les variations inter-locuteurs des phonèmes étudiés. Seuls les textes influencent significativement les variations intra-locuteur des performances de détection de /ə/ et /ø/. Puisque la détection de chaque phonème n'est pas affectée par la somnolence,

le système de détection de schwas proposé est robuste à la somnolence et fiable pour la détection de ces phonèmes.

Nous allons maintenant mesurer si le nombre de /ə/, /ø/, et /œ/ détecté automatiquement est corrélé au niveau de somnolence des patients sur l'ensemble du corpus TILE en utilisant le système proposé.

Facteur	Perf. sur /ə/	Perf. sur /ø/	Perf. sur /œ/	Perf. sur « e »
Texte	***	***	-	***
KSS	-	-	-	-
Latence d'endormissement	-	-	-	-

TABLE 5 – Résultats des ANOVA multivariées à mesures répétées avec le système de détection de schwas proposé sur le sous-ensemble. Nous n'avons trouvé aucune influence significative du Texte, de la somnolence subjective ou objective sur les variations inter-locuteurs ; seule l'influence sur les variations intra-locuteurs est rapportée. *** : $p < .001$

5 Lien entre somnolence et détection automatique de phonèmes

Chaque /ə/, /ø/, et /œ/ extrait des textes transcrits avec le Lexique 3.83 a été comptabilisé afin d'obtenir le nombre attendu pour chacun d'entre eux. Nous avons appliqué le système proposé au corpus TILE et calculé la moyenne et l'écart-type du nombre de phonèmes détectés sur les enregistrements. La moyenne et l'écart-type de la valeur absolue de la différence entre le nombre de phonèmes attendus et détectés ont également été calculés. La Table 6 référence ces mesures.

Texte	Mesures	#/ə/	#/ø/	#/œ/	#« e »
Texte 1	Attendu	30	4	4	38
	Détecté	31 (1)	4 (1)	4 (0)	38 (2)
Texte 2	Attendu	53	7	3	63
	Détecté	50 (2)	8 (1)	3 (1)	61 (3)
Texte 3	Attendu	42	7	3	52
	Détecté	42 (3)	6 (1)	3 (1)	50 (3)
Texte 4	Attendu	42	7	2	51
	Détecté	41 (3)	7 (1)	2 (1)	50 (3)
Texte 5	Attendu	35	6	0	41
	Détecté	34 (2)	7 (1)	0 (0)	42 (2)
Tous	Attendu	202	31	12	245
	Détecté	197 (6)	32 (2)	12 (1)	242 (7)

TABLE 6 – Nombre de /ə/, /ø/, /œ/ et « e » attendus. Moyenne et écart-type (valeurs arrondies) des phonèmes détectés par le système de détection de schwas proposé.

Le nombre détecté de /œ/ est le plus proche du nombre attendu tandis que la détection de /ə/ est la plus éloignée. Cela peut s'expliquer par son nombre attendu élevé par rapport à /ø/ et /œ/ (202, 31, et 12 respectivement pour tous les textes). En moyenne, le nombre de phonèmes détectés par ce système est égal ou proche du nombre attendu, ce qui signifie que le système proposé est fiable.

Pour évaluer l'impact de la somnolence sur le nombre de phonèmes détectés par le système proposé, nous avons réalisé quatre ANOVA multivariées à mesures répétées avec les mêmes facteurs qu'au-

paravant (textes, KSS, et latence d'endormissement). Les résultats sont référencés dans la Table 7. Aucun effet significatif des variations inter-locuteurs n'a été trouvé contrairement aux variations intra-locuteurs. Un effet significatif de la somnolence subjective à court terme (KSS) sur les nombres détectés automatiquement de /ø/ et « e » a été mesuré tandis que la latence d'endormissement n'affecte la détection automatique d'aucun phonème. Le perception subjective de la somnolence à court terme impacterait donc la détection automatique du nombre de /ø/ et « e » contrairement à la somnolence physiologique mesurée par EEG.

Facteur	Nombre de /ə/	Nombre de /ø/	Nombre de /œ/	Nombre de « e »
KSS	.	**	-	*
Latence d'endormissement	-	-	-	.

TABLE 7 – Résultats de l'ANOVA multivariée à mesures répétées (variations intra-locuteur uniquement) avec le système de détection de schwas proposé sur le corpus TILE. . : $p < .1$; * : $p < .05$; ** : $p < .01$

6 Conclusion

Dans le but d'aider les cliniciens à suivre l'évolution des symptômes de la Somnolence Diurne Excessive (SDE) en analysant la parole spontanée, nous avons cherché à compléter les approches précédentes centrées sur la durée et l'emplacement des pauses en étudiant le comportement phonétique. En particulier, nous nous sommes intéressés au schwa, phonème optionnel en français, dont sa réalisation peut refléter un effort vocal de la part du locuteur. Nous émettons l'hypothèse que la somnolence altère cet effort. Du fait de l'absence de corpus de parole spontanée spécifique à la somnolence, nous avons utilisé le corpus TILE contenant des enregistrements de parole lue car l'analyse phonétique est transposable à la parole spontanée.

Pour ce faire, nous avons validé un système de détection de schwas fiable et robuste à la somnolence sur l'annotation manuelle d'un sous-ensemble du corpus TILE et l'avons étendu à d'autres phonèmes liés au schwa. Nous l'avons ensuite appliqué à l'ensemble du corpus TILE et avons trouvé une relation significative entre les phonèmes considérés détectés automatiquement par notre système et la somnolence subjective à court terme.

Notre prochaine étape consiste à concevoir un système de classification automatique en prenant en compte les résultats actuels afin d'améliorer la détection de la somnolence. De plus, nous visons à étudier le lien entre la durée et les propriétés acoustiques des trois phonèmes et la somnolence. La méthode utilisée pour le système de détection de schwas pouvant être appliquée à d'autres phonèmes, nous projetons également d'élargir les phonèmes considérés ainsi que les caractéristiques étudiées (le débit de parole, etc.).

Remerciements

CB a reçu le soutien financier de la MITI du CNRS (projet PRIME 80 DSM-HEALTH). 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).

Références

ARAND D., BONNET M., HURWITZ T., MITLER M., ROSA R. & SANGAL R. B. (2005). The Clinical Use of the MSLT and MWT. *Sleep*, **28**(1), 123–144. DOI : 10.1093/sleep/28.1.123.

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., WU Y., ROUAS J.-L. & PHILIP P. (2023). Automatic detection of schwa in French hypersomniac patients. *Plate-Forme Intelligence Artificielle (PFIA)*.

BIOULAC S., MICOULAUD-FRANCHI J.-A., ARNAUD M., SAGASPE P., MOORE N., SALVO F. & PHILIP P. (2017). Risk of Motor Vehicle Accidents Related to Sleepiness at the Wheel : A Systematic Review and Meta-Analysis. *Sleep*, **40**(10). DOI : 10.1093/sleep/zsx134.

BOYER F. & ROUAS J.-L. (2019). End-to-End Speech Recognition : A review for the French Language. *arXiv*. DOI : 10.48550/ARXIV.1910.08502.

BÜRKI A., ERNESTUS M., GENDROT C., FOUGERON C. & FRAUENFELDER U. H. (2011). What affects the presence versus absence of schwa and its duration : A corpus analysis of French connected speech. *J. Acoust. Soc. Am.*, **130**(6), 3980–3991. DOI : 10.1121/1.3658386.

DURAND J. (2014). À la recherche du schwa : données, méthodes et théories. *SHS Web of Conferences*, **8**, 23–43. DOI : 10.1051/shsconf/20140801396.

FOUGERON C., GENDROT C. & BÜRKI A. (2007). On the acoustic characteristics of French schwa. In *ICPhS 2007*, Saarbrücken.

GALLIANO S., GRAVIER G. & CHAUBARD L. (2009). The ester 2 evaluation campaign for the rich transcription of French radio broadcasts. In *Interspeech 2009*, p. 2583–2586. DOI : 10.21437/Interspeech.2009-680.

GUPTA V., KENNY P., OUELLET P. & STAFYLAKIS T. (2014). I-vector-based speaker adaptation of deep neural networks for French broadcast audio transcription. In *ICASSP*, p. 6334–6338. DOI : 10.1109/ICASSP.2014.6854823.

HARRISON Y. & HORNE J. A. (1997). Sleep Deprivation Affects Speech. *Sleep*, **20**(10), 871–877. DOI : 10.1093/sleep/20.10.871.

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.

MACLEAN A. W., FEKKEN G. C., SASKIN P. & KNOWLES J. B. (1992). Psychometric evaluation of the Stanford Sleepiness Scale. *J. Sleep Res.*, **1**(1), 35–39. DOI : 10.1111/j.1365-2869.1992.tb00006.x.

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

MARTIN V., FERRON A., ROUAS J.-L., SHOCHI T., DUPUY L. & PHILIP P. (2023a). Physiological vs. Subjective sleepiness : what can human hearing estimate better ? Insights from the French Endymion study. In *ICPhS 2023*.

MARTIN V., 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 Med. Rev.*, **67**, 101718. DOI : 10.1016/j.smrv.2022.101718.

MARTIN V., ROUAS J.-L., BOYER F. & PHILIP P. (2021a). Automatic Speech Recognition Systems Errors for Objective Sleepiness Detection Through Voice. In *Interspeech 2021*, p. 2476–2480. DOI : 10.21437/Interspeech.2021-291.

MARTIN V., ROUAS J.-L., MICOULAUD-FRANCHI J.-A., PHILIP P. & KRAJEWSKI J. (2021b). 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., BEAUMARD C., ROUAS J.-L. & WU Y. (2024). Is automatic phoneme recognition suitable for speech analysis ? Temporal and performance evaluation of an Automatic Speech Recognition model in spontaneous French. In *Speech Prosody 2024*.

NEW B., PALLIER C., BRYSBAERT M. & FERRAND L. (2004). Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, **36**(3), 516–524. DOI : 10.3758/BF03195598.

POVEY D., GHOSHAL A., BOULIANNE G., BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLÍČEK P., QIAN Y., SCHWARZ P., SILOVSKÝ J., STEMMER G. & VESELÝ K. (2011). The Kaldi Speech Recognition Toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, Hawaii : IEEE Signal Processing Society.

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.

SCHULLER B. W., BATLINER A., BERGLER C., POKORNY F. B., KRAJEWSKI J., CYCHOSZ M., VOLLMANN R., ROELEN S.-D., SCHNIEDER S., BERGELSON E., CRISTIA A., SEIDL A., WARLAUMONT A. S., 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*, p. 2378–2382. DOI : 10.21437/Interspeech.2019-1122.

STOLCKE A. (2002). SRILM - an extensible language modeling toolkit. In *7th International Conference on Spoken Language Processing (ICSLP 2002)*, p. 901–904. DOI : 10.21437/ICSLP.2002-303.

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*, p. 6937–6941. DOI : 10.1109/ICASSP43922.2022.9747000.