Détection de séquences de phonèmes en parole spontanée pour la caractérisation de la somnolence diurne excessive

Colleen Beaumard, Vincent P. Martin, Charles Brazier, Yaru Wu, Jean-Luc Rouas

Détection de séquences de phonèmes en parole spontanée pour la caractérisation de la somnolence diurne excessive

La Somnolence Diurne Excessive (SDE) est un trouble courant se manifestant par une difficulté à rester éveillé ou attentif durant la journée. Son diagnostic précoce et précis est crucial pour améliorer la qualité de vie des patients et prévenir des accidents liés à la somnolence, notamment au volant ou sur le lieu de travail.

Parmi les approches pour détecter ce trouble, nous avions montré l'intérêt de l'analyse de la voix à partir de tâches de lecture [1]. Afin de proposer des mesures pouvant être effectuées sans impact sur la vie quotidienne, nous proposons d'analyser les réalisations de séquences de phonèmes en parole spontanée, ce qui à notre connaissance n'a pas été étudié. Nous espérons ainsi évaluer des erreurs dans l'articulation de séquences de phonèmes pouvant être liées à la SDE.

Dans un premier temps, nous avons évalué les performances d'un système de reconnaissance automatique de phonèmes (HMM-TDNN) permettant d'avoir à la fois le phonème détecté et ses frontières temporelles contrairement à d'autres modèles plus récents. Il a été appliqué au corpus Rhapsodie [2] en terme de taux d'erreur en tokens (TER, Tableau 1) et sur la qualité de la segmentation temporelle grâce à l'outil « trackeval » [3], [4] (Tableau 2). Ces évaluations sont poursuivies ici avec la détermination automatique des séquences de phonèmes les plus fréquentes via des comptages générés avec le logiciel SRILM [5]. Nous avons confirmé les bonnes évaluations du système HMM-TDNN avec une F-mesure de 0.70 pour les 10 séquences de trois phonèmes les plus fréquentes. Ces résultats sont encore améliorés lorsque l'on regroupe les unités phonétiques en fonction du mode d'articulation (i.e. occlusives, fricatives, semi-voyelles ou « glides », nasales et liquides pour les consonnes ; nasales, centrales ou antérieures arrondies/non arrondies pour les voyelles). Ainsi, en considérant les 10 séquences de trois symboles représentant les modes d'articulation (e.g. occlusive+voyelle_antérieure+liquide), la F-mesure atteint 0.80, ce qui permet d'envisager leur analyse acoustique (Tableau 3).

Nous avons ensuite appliqué la détection automatique de ces séquences au corpus « Medispeech », qui est la prolongation du corpus MSLT [6]. Il contient des tâches de lecture et de parole (semi-)spontanée et est annoté à la fois avec une mesure subjective de la somnolence (échelle de Karolinska — KSS) et une mesure physiologique (Test Itératif de Latence d'Endormissement — TILE). Ce corpus contient les enregistrements de 34 locuteurs, soit 170 enregistrements par type de parole. Nous avons caractérisé les séquences de modes articulatoires en calculant des paramètres acoustiques simples : Fréquence fondamentale moyenne et écart-type, Intensité moyenne et écart-type, pourcentage de voisement, durée. Ces paramètres sont ensuite évalués statistiquement avec des modèles linéaires mixtes. Les analyses montrent des liens forts entre les paramètres calculés sur certains types de séquences et les mesures cliniques de la somnolence (Tableau 4). Par exemple, le ressenti de la somnolence (KSS) est lié à l'intensité moyenne mesurée sur la séquence voyelle_antérieure+liquide+voyelle_antérieure (**). Ces résultats préliminaires nous encouragent à utiliser cette méthode d'analyse en vue de la classification automatique de la somnolence à partir d'enregistrements de parole spontanée.

500 mots

Cette recherche est supportée par le CNRS au moyen du MITI PRIME 80 DSM-HEALTH et de l'Agence Nationale de la Recherche (ANR) avec l'axe « Autonom-Health » du PEPR « Santé Numérique », Accord de subvention n°ANR-22-PESN-000X

Références :

[1] V. P. Martin, J.-L. Rouas, and P. Philip, 'Automatic detection of sleepiness-related symptoms and syndromes using voice and speech biomarkers', *Biomedical Signal Processing and Control*, vol. 91, p. 105989, May 2024, doi: 10.1016/j.bspc.2024.105989

[2] A. Lacheret-Dujour, S. Kahane, and P. Pietrandrea, 'Rhapsodie: A prosodic and syntactic treebank for spoken French'. John Benjamins, 2019. Available: https://www.jbe-platform.com/content/books/9789027262929,

[3] V. Martin, C. Beaumard, J.-L. Rouas, and Y. Wu, '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*, Leiden (Netherlands), Netherlands: ISCA, Jul. 2024, pp. 1120–1124. doi: 10.21437/SpeechProsody.2024-226. Available: https://hal.science/hal-04679813

[4] V. P. martin, C. Beaumard, C. Brazier, J.-L. Rouas, and Y. Wu, 'La reconnaissance automatique de phonèmes est-elle réellement adaptée pour l'analyse de la parole spontanée ?', in *35èmes journées d'Études sur la parole (JEP 2024) 31ème conférence sur le traitement automatique des langues naturelles (TALN 2024) 26ème rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues (RECITAL 2024)*, 2024, pp. 431–440.

[5] A. Stolcke, 'SRILM - An extensible language modeling toolkit', in *ICSLP*, 2002, pp. 901–904. Available: http://www.speech.sri.com/projects/srilm/

[6] V. P. Martin, J.-L. Rouas, J.-A. Micoulaud-Franchi, and P. Philip, 'The objective and subjective sleepiness voice corpora', in *12th edition of its language resources and evaluation conference.*, Marseille, France, May 2020, pp. 6525–6533. Available: https://hal.archives-ouvertes.fr/hal-02489433

Annexes :

	%TER (35 phones)	%TER (10 modes)
planned	12.8	7
semi-spontaneous	20.1	13.8
spontaneous	21.9	15.7
All	18.8	12.7

Tableau 1: Évaluation du système de transcription phonétique automatique HMM-TDNN en Taux d'Erreur de Tokens (Token Error Rate, TER) en fonction de l'unité cible (phones ou modes d'articulation), pour les trois styles de parole du corpus Rhapsodie.

	F-measure (35 phones)	F-measure (10 modes)
planned	0.6805	0.6519
semi-spontaneous	0.6148	0.7009
spontaneous	0.6082	0.632
All	0.6334	0.6269

Tableau 2: Évaluation des performances en segmentation de tokens du système de transcription phonétique automatique HMM-TDNN en fonction en fonction de l'unité cible (phones ou modes d'articulation), pour les trois styles de parole du corpus Rhapsodie.

	F-measure (10 trigrammes-phones)	F-measure (10 trigrammes-modes)
planned	0.7467	0.8479
semi-spontaneous	0.6343	0.8018
spontaneous	0.6765	0.7762
All	0.6999	0.8043

Tableau 3: Évaluation des performances en segmentation de tokens du système de transcription phonétique automatique HMM-TDNN en fonction en fonction de l'unité cible (trigrammes phones ou modes d'articulation), pour les trois styles de parole du corpus Rhapsodie.

sequence	CSto_VFrU_CLiq	VFrU_CFri_VFrU			VFrU_CLiq_VFrU
# total	805	1159			870
# per speaker (std)	23.70 (13.68)	34.11 (17.16)			25.61 (12.71)
# per recording (std)	5.23 (3.62)	7.11 (4.40)			5.37 (3.39)
feature	dur	F0v	NRJv	dur	NRJm
IMC			**		
HAD_anxiete			*		
FSS					*
TILE		*	*	*	*
TILE_moy	*	*	*		
KSS					**

*Tableau 4: Résultats des tests statistiques pour les séquences de trois symboles de modes articulatoires ayant donné des valeurs significatives. Première ligne : séquence de symboles (CSto=occlusive, CFri=fricative, CLiq=liquide, VFrU=voyelle antérieure non arrondie). Deuxième ligne : nombre total de séquences analysées. Troisième ligne : nombre de séquences détectées pour chaque locuteur sur chaque enregistrement (moyenne et écart-type). Quatrième ligne : nombre de séquences détectées par enregistrement (moyenne et écart-type). Cinquième ligne : paramètre étudié (dur=durée de la séquence, F0v=variance de fréquence fondamentale sur la séquence, NRJv=variance de l'intensité sur la séquence, NRJm=moyenne de l'intensité sur la séquence). Les lignes suivantes donnent les p-values des tests statistiques effectués avec la bibliothèque python « statsmodels », td. : * : p <0.05 ; ** : p <0.01. Seuls sont reportés ici les facteurs significatifs (IMC=indice de masse corporelle, HAD_anxiete=score au questionnaire « Hospital Anxiety and Depression scale » pour la partie Anxiété, FSS=score au questionnaire « Fatigue Severity Scale », TILE=temps de latence d'endormissement mesuré au cours du test clinique, TILE_moy=moyenne des TILE sur 5 itérations (journée), KSS=score au questionnaire « Karolinska Sleepiness Scale ».*