

CFA/VISHNO 2016

Paramètres Acoustiques de la Voix et Privation de Sommeil

S. Boyer^a, R. El-Yagoubi^a, M. Tiberge^b, R. Ruiz^c et A. Daurat^a

^aCLLE-LTC Univ. Toulouse 2 Jean Jaurès, 5 allées Antonio Machado, 31058 Toulouse Cedex 9, France

^bLaboratoire du sommeil- Service de neurologie, CHU Toulouse, place du docteur Baylac, 31059 Toulouse Cedex 9, France

^cLARA - Université Toulouse 2 Jean Jaurès, 5 allées Antonio Machado, 31058 Toulouse Cedex 9, France
stanislas.boyer@univ-tlse2.fr



LE MANS

Cette étude a pour objectif d'évaluer l'impact d'une privation de sommeil sur les paramètres acoustiques de la voix humaine et de déterminer si certains d'entre eux peuvent être utilisés pour se prononcer sur l'état de vigilance. Il s'agirait d'applications concernant la sécurité des transports notamment aéronautiques. 22 individus ont été invités à participer à deux sessions expérimentales, avant et après une nuit sans sommeil. Chaque session était composée d'un test objectif de somnolence (*Karolinska Drowsiness Test* ou KDT) ainsi que d'une tâche de lecture de texte. Pour le KDT, l'ÉlectroEncéphaloGraphie (EEG) a été enregistré pendant 9 min. Une analyse spectrale de l'EEG dans les bandes de fréquence des ondes alpha (α) et thêta (θ) a été réalisée ; représentatives de l'état de vigilance de l'individu. La voix a été enregistrée avec un microphone de proximité et un magnétophone numérique lors de la tâche de lecture d'un texte sans valence émotionnelle. Comme attendu, les résultats des puissances spectrales EEG montrent une augmentation de la somnolence après privation de sommeil. L'analyse acoustique des voyelles segmentées dans le discours met en évidence des variations significatives du rapport HNR, des coefficients MFCC, d'une fréquence de modulation des pics d'amplitude et d'un paramètre original de la longueur euclidienne du signal. Les variations de certains de ces paramètres acoustiques à la suite de la privation de sommeil sont corrélées avec les variations de l'EEG.

1 Introduction

La mise en œuvre de méthodes de détection en temps réel de l'état de somnolence d'un individu représente un enjeu majeur pour le domaine des transports (e.g., aérien, ferroviaire, etc.). Dans cette optique, l'approche acoustique à travers l'étude de la voix humaine semble particulièrement adaptée chez des populations de locuteurs dont les professions nécessitent une communication constante via une chaîne électro-acoustique (e.g., pilotes d'avion, contrôleurs aérien, etc.). Quelques études se sont intéressées aux perturbations de la voix suite à des privations totales ou partielles de sommeil. Les résultats ont montré :

- une diminution de l'intensité vocale [1] vs. aucun changement de l'intensité vocale [5] ;
- une diminution de la fréquence fondamentale (F_0) [1,2,3,4,5] vs. une diminution de F_0 [6] ;
- une diminution de l'écart type de F_0 (SD_{F_0}) [1][3] vs. une augmentation de SD_{F_0} [8] vs. aucun changement de SD_{F_0} [5] ;
- une augmentation du jitter [5,6]
- une augmentation du shimmer [5]
- une diminution de la fréquence des formants 1 [4,7], 2 et 3 [7] et 4 [7,8] vs. aucun changement de la fréquence des formants 1, 2 et 3 [8] ;
- une augmentation de la pente spectrale déterminée par le rapport alpha (énergie ≤ 1 kHz/énergie 1-8 kHz) [8] ou obtenue par régression linéaire sur la distribution d'énergie pour les fréquences > 1 kHz [5] vs. une diminution de la pente spectrale pour les fréquences > 1 kHz [9] ;
- une variation des coefficients cepstraux (MFCC)[10]
- une diminution du débit de parole [1,8] vs. aucun changement du débit de parole [5], ni de la durée des phrases, des voyelles et des pauses [6].

Ces résultats parfois contradictoires s'expliquent en partie par les nombreuses différences méthodologiques entre les travaux telles que la taille de l'échantillon de locuteurs, le type de prononciation qu'effectue le locuteur (discours naturel, lecture de texte, énonciation de mots ou de voyelles), le type de vocabulaire analysé (énoncés, phrases, mots, consonnes ou voyelles) ou encore de la durée de la privation de sommeil. Par ailleurs, aucune de ces études ne s'est focalisée sur la relation entre les perturbations acoustiques et les variations de l'état de somnolence consécutives à la privation de sommeil.

L'objectif de ce travail de recherche est d'évaluer l'impact d'une privation totale de sommeil d'une nuit (PTS) sur les paramètres acoustiques de la voix et de déterminer s'il existe une relation entre les variations des paramètres acoustiques et les variations de l'état de somnolence.

2 Matériel et Méthode

2.1 Participants

22 participants (12 femmes) de nationalité et de langue maternelle française ont été privés de sommeil durant une nuit dans le cadre d'un examen médical pour suspicion d'épilepsie réalisé au laboratoire du sommeil du centre hospitalier universitaire de Toulouse (âge moyen 27.05 ± 7.92).

Tous les participants ont rapporté ne pas souffrir de troubles psychiatriques et ne pas prendre de traitements pharmacologiques. Ils ont signé une fiche de consentement éclairé et ont été informés qu'ils étaient libres de mettre fin à l'expérimentation à n'importe quel moment.

2.2 Procédure

L'expérimentation était composée de deux sessions expérimentales d'environ 1h30, une avant et une après une nuit de privation totale de sommeil (PTS). La session expérimentale avant la PTS débutait la veille à 18h30 (pic circadien de la vigilance) et la session après la PTS débutait à 7h le lendemain matin. Chaque session était composée d'une version modifiée du test de somnolence de Karolinska (en anglais *Karolinska Drowsiness Test* ou KDT ; [11]) et d'une tâche de lecture de texte.

Les participants ont passé la nuit au centre du sommeil durant toute la PTS. Ils pouvaient regarder des films, écouter de la musique, jouer à des jeux vidéo, de carte ou de société ou interagir avec l'équipe médicale. Ils n'avaient pas le droit de pratiquer une activité physique, de prendre de la caféine ou tout autre stimulant présent dans le laboratoire.

2.3 Electroencéphalographie

L'ÉlectroEncéphaloGraphie (EEG) a été enregistré avec un système Biosemi à 34 électrodes présenté dans la Figure 1. Les électrodes étaient positionnées sur un bonnet au-dessus des aires frontale, centrale, pariétale, occipitale et temporale. Elles étaient placées en image miroir sur chaque hémisphère de l'encéphale en accord avec le système de placement international 10-20 standard : Fp1/2, AF3/4,

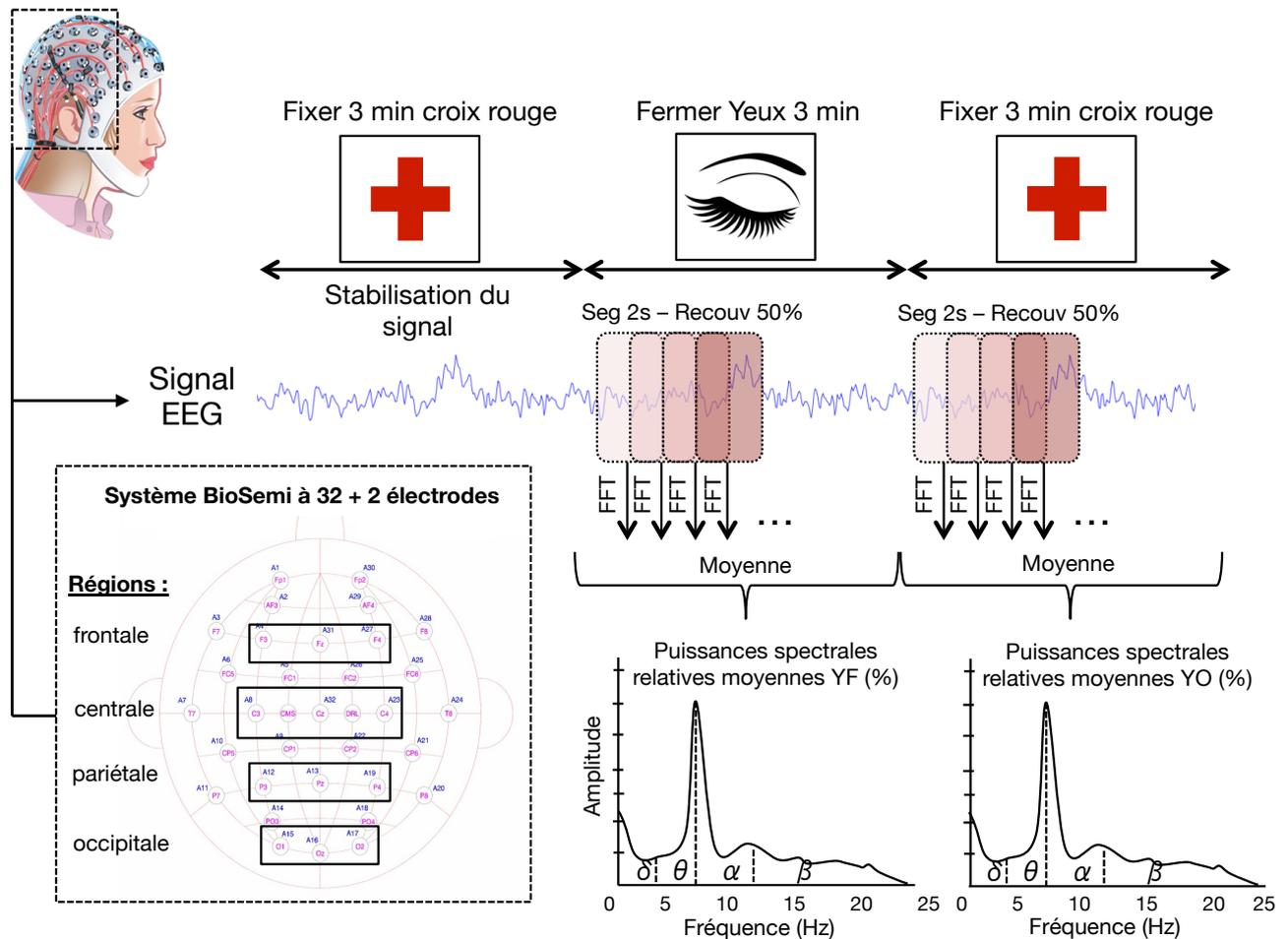


Figure 1 : Système BioSemi à 32 électrodes et version modifiée du test de somnolence de Karolinska (KDT).

Seg : segment ; Recouv : recouvrement ; YF : phase « Yeux Fermés » ; YO : phase « Yeux Ouverts » ;

FFT : transformée de Fourier

F7/8, F3/4, FT7/8, FC1/2, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, PO3/4, CMS/DRL, O1/2 et les quatre électrodes centrales Fz, Cz, Pz, Oz (*International 10/20 System* ; [12]). Deux électrodes de référence étaient placées au niveau de l'os mastoïde gauche et droit. Les données étaient référencées sur la moyenne algébrique des deux mastoïdes. L'impédance des électrodes a été maintenue en dessous de $1k\Omega$. Les données étaient transmises sans interruption par l'intermédiaire des électrodes actives, amplifiant le signal à 1024 Hz et reliées à un ordinateur d'acquisition équipé du logiciel Actiview 6.0.2. L'électro-oculogramme horizontal (HEOG) a été enregistré à l'aide de deux électrodes en étain placées à 1 cm à droite et à gauche du canthi externe et l'électro-oculogramme vertical (VEOG) à l'aide de deux électrodes placées en dessous et au dessus de l'oeil gauche, permettant ainsi de détecter les clignements et les mouvements verticaux des yeux.

2.4 Test de somnolence de Karolinska

La version modifiée du test de somnolence de Karolinska (KDT, [11]) présentée dans la Figure 1 consistait à enregistrer l'électroencéphalogramme (EEG) du participant durant 9 minutes, alors qu'il était installé sur une chaise confortable, dans une pièce calme et à une distance de 2 mètres d'un mur sur lequel était placé à la hauteur de ses yeux une croix rouge.

Durant l'enregistrement, il était demandé au participant de se relaxer, de garder ses yeux ouverts et de fixer la croix rouge durant 3 minutes en évitant le plus possible le clignement des yeux afin de maximiser la qualité du signal. Les 3 minutes suivantes, il devait fermer ses yeux et lutter contre le sommeil. Il devait les 3 dernières minutes garder ses yeux ouverts et fixer de nouveau la croix rouge. Le participant avait pour consigne de lutter contre le sommeil durant les 3 phases du test.

Les puissances spectrales relatives des ondes cérébrales alpha (α : 7.9-12.6 Hz, en %) et thêta (θ : 3.9-7.8 Hz, en %) – qui reflètent l'état de somnolence d'une personne – ont été calculées par transformée de Fourier (FFT) avec une fenêtre de *Hamming* durant les 3 minutes du KDT avec les yeux fermés (phase « Yeux Fermés » ou YF) et durant les 3 dernières minutes du test avec les yeux ouverts (phase « Yeux Ouverts » ou YO). La première phase de 3 minutes du KDT avec les yeux ouverts permettait de stabiliser le signal EEG. Cette phase est de ce fait exclue de l'analyse. Les phases YF et YO ont été divisées en segment de 2 s avec un taux de recouvrement de 50% (i.e., chaque segment de 2 s recouvrait le précédent de 1 s), ce qui représentait 178 segments de 2 s par phase. Les puissances spectrales relatives α et θ ont été calculées sur chaque segment de 2 s du signal EEG sur 12 électrodes : F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz et O2. Le filtre passe-haut était défini à

0.5 Hz et le filtre passe-bas à 30 Hz. Les segments de 2 s avec des amplitudes supérieures à $75\mu\text{V}$ ont été rejetés de l'analyse et les artefacts sur le signal liés aux clignements des yeux ont été corrigés pour l'analyse statistique [13]. Les puissances spectrales relatives α et θ obtenues pour chaque segment de 2 s ont été moyennées afin d'obtenir la puissance spectrale relative moyenne α et θ de chaque phase (YF et YO). La dernière étape a consisté à regrouper les douze électrodes en 4 aires (cf. Figure 1) : frontale (F3, Fz, F4), centrale (C3, Cz, C4), pariétale (P3, Pz, P4) et occipitale (O2, Oz, O1), puis à calculer les puissances spectrales relatives α et θ pour chaque aire (moyenne des 3 électrodes) et pour chaque phase (YF et YO).

2.5 Tâche de lecture et analyse acoustique

La tâche acoustique consistait à lire un texte. Deux textes sans valence émotionnelle (i.e., extraits des notices d'utilisation des logiciels PowerPoint et Excel) ont été sélectionnés. Afin d'éviter les effets d'apprentissage, le texte utilisé avant la PTS était différent de celui utilisé après PTS.

Durant la tâche de lecture de texte, la voix des participants a été enregistrée avec un microphone de proximité monté sur un serre-tête (AKG C555L) pour maintenir une distance constante entre les lèvres et le transducteur (i.e., prise de son indépendante des mouvements de la tête). La position du microphone est restée la même tout au long de l'expérimentation. L'enregistreur numérique était un Fostex FR-2 avec une fréquence d'échantillonnage de 44.1 kHz et une résolution de 16-bits.

Les voyelles ont été étiquetées et délimitées manuellement avec le logiciel PRAAT, en ne conservant que la partie quasi-stationnaire du signal. Un contrôle auditif a également été réalisé. L'attaque, l'extinction et la partie de transition vers les phonèmes adjacents ont été supprimées. Les monophthongues ont donc été analysées durant leur période quasi-stationnaire. Un « script » PRAAT a été développé pour extraire du signal les segments de voyelles repérés et les sauvegarder sous forme de fichier (i.e., un fichier .wav par voyelle). Pour chaque participant, environ 200 voyelles de douze types (i.e., 10 orales : [a], [ə], [i], [e], [ɛ], [œ], [o], [ɔ], [u], [y] ; 2 nasales : [ã] ; [ẽ]) ont été extraites pour chaque session de la tâche expérimentale.

Pour chaque voyelle, 45 paramètres acoustiques articulatoires et phonatoires (prosodiques, forme de l'onde acoustique et spectraux) ont été mesurés. Les caractéristiques spectrales dépendant du type de voyelle, les analyses spectrales se sont focalisées sur les voyelles [a], [ə], [i] et [e] car elles étaient les plus représentées à l'issue de la segmentation. Ces voyelles sont parmi les plus courantes, leurs fréquences formantiques couvrent un large spectre et leur nombre (> 20) dans les textes à lire était suffisant pour réaliser des analyses statistiques – ce qui n'était pas le cas pour les autres types. Un script PRAAT a été développé pour automatiser les analyses des paramètres prosodiques (excepté la fréquence de modulation d'amplitude) et des paramètres spectraux (excepté la pente spectrale et la fréquence d'équilibre énergétique) des voyelles segmentées. Ces paramètres sont désignés par leur nom provenant du manuel du logiciel PRAAT. Leurs descriptions proviennent également en majorité de ce même

logiciel. Des programmes MATLAB® ont été développés pour mesurer et automatiser le traitement des paramètres qui concernent la forme de l'onde acoustique, la fréquence de modulation d'amplitude, la pente spectrale et la fréquence d'équilibre énergétique. Des études passées ont exploré le signal vocal afin d'analyser les effets de nombreuses perturbations psychophysologiques [14,15, 16]. Certaines des caractéristiques analysées ont été sélectionnées sur la base de ces études.

- Caractéristiques prosodiques

Le logiciel PRAAT fournit la fréquence fondamentale (F_0) moyenne de chaque segment sélectionné (i.e., pour chaque voyelle) et l'écart type associé (*Standard Deviation of F_0* ou SD_{F_0}). Les calculs sont effectués par tranche d'analyses successives dans le segment considéré. Les jitters ont été calculés afin d'estimer la variabilité à court-terme de F_0 et les shimmers afin d'estimer la variabilité à court-terme de l'amplitude des pulses glottiques, avec un effet de lissage plus ou moins important selon le nombre de valeurs voisines prises en compte dans le calcul. Cinq jitters (Jitter 1, Jitter 2, RAP, PPQ5 et DDP) et six shimmers (Shimmer 1, Shimmer 2, APQ3, APQ5, APQ11 et DDA) ont été sélectionnés. Comme pour les shimmers, la fréquence de modulation d'amplitude (Fmod) est liée aux variations d'amplitudes au sein du signal de la voyelle. Plus spécifiquement, Fmod est la fréquence de modulation de l'amplitude des pulses glottiques de la voyelle. Elle est obtenue en réalisant une régression sinusoïdale sur l'amplitude des pics glottiques et sa validité est estimée en calculant le coefficient de détermination (R^2) entre l'ensemble des valeurs d'amplitudes ajustées et l'ensemble des valeurs brutes [16].

- Caractéristiques de la forme de l'onde acoustique

Le *Digital Amplitude Length* (DAL) est une estimation de la longueur euclidienne du signal, obtenue par le calcul de la somme des différences (en valeur absolue) de l'amplitude entre les échantillons successifs [13]. Lorsque cette somme est calculée entre deux pulses glottiques, DAL_{T_0} devient une estimation de la longueur d'un cycle vibratoire et DAL_{T_0}/T_0 une estimation de cette longueur par unité de temps. $\langle DAL_{T_0} \rangle$ et $\langle DAL_{T_0}/T_0 \rangle$ sont les valeurs moyennes de ces deux grandeurs sur l'ensemble des périodes de la voyelle. Les variations à court-terme de DAL_{T_0} et DAL_{T_0}/T_0 sont obtenues par le calcul du jitter de DAL_{T_0} ($jitt(DAL_{T_0})$) et de celui de DAL_{T_0}/T_0 ($jitt(DAL_{T_0}/T_0)$). L'hypothèse est que les variations du signal dues à des perturbations de la parole vont modifier le contour de la forme de l'onde acoustique sur le chronogramme et donc sa longueur.

- Caractéristiques spectrales

Les paramètres spectraux ont été choisis pour étudier principalement les mouvements d'énergie dans le spectre – on s'intéresse ici à la forme du spectre. Ces mouvements peuvent être observés par (1) les variations des fréquences des quatre formants (F_i), le centre de gravité spectral (*Spectral Center of gravity* ou SCG) et la fréquence d'équilibre énergétique (*Energy Balance Frequency* ou EBF), (2) les caractéristiques de la forme spectrale avec le coefficient d'asymétrie (*skewness* ou SKE), d'aplatissement (*kurtosis* ou KUR) et la pente spectrale (*Spectral Tilt* ou ST), et (3) par comparaison avec trois différences d'énergie (*Energy Differences* ou EDi). Un autre paramètre, le rapport signal sur bruit (*Harmonic-to-Noise Ratio* ou

HNR), joue un rôle important dans l'étude de la qualité de la voix et des pathologies. Très sensible à la présence de bruit dans la voix, il a semblé pertinent de le tester en émettant l'hypothèse que la voix pourrait être plus bruitée chez un individu somnolent. Douze coefficients cepstraux (*Mel-Frequency Cepstral Coefficients* ou MFCC) ont également été sélectionnés et mesurés en raison de leur importance dans la reconnaissance vocale et la classification automatique.

2.6 Analyses statistiques

Des analyses de variance (ANOVAs) à mesures répétées avec un risque d'erreur de 5% ont été effectuées afin de comparer les puissances spectrales relatives α et θ avant et après la PTS lors des phases YF et YO du KDT. Le modèle statistique comprenait deux facteurs à mesures répétées : région (frontale, centrale, pariétale et occipitale) et privation totale de sommeil (PTS : avant et après). Des tests-t ont été utilisés comme tests post-hoc.

Des tests-t ont été effectués afin de comparer les paramètres prosodiques et de la forme de l'onde acoustique avant et après la PTS. Les caractéristiques spectrales dépendant du type de voyelle, chaque paramètre spectral a été analysé au moyen d'une ANOVA à mesures répétées avec un risque d'erreur de 5%. Le modèle statistique comprenait deux facteurs à mesures répétées : voyelle ([a], [ə], [i] et [e]) et privation totale de sommeil (PTS : avant et après). Des tests-t ont été utilisés comme tests post-hoc.

Des corrélations de Pearson ont été réalisées entre les variations des indicateurs objectifs de la somnolence (i.e., puissances spectrales relatives α et θ) et les variations des paramètres acoustiques à la suite de la PTS.

3 Résultats

3.1 Test de somnolence de Karolinska

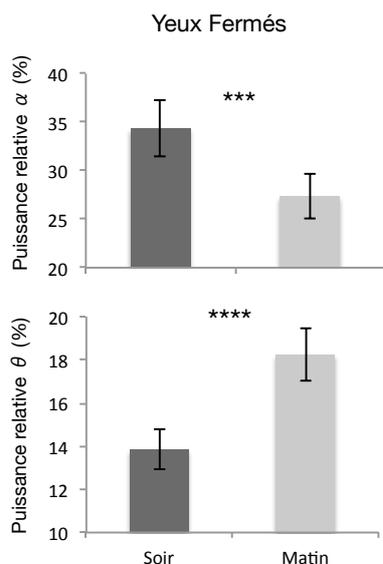


Figure 2 : Puissances spectrales relatives α et θ avant et après la PTS lors de la phase « Yeux Fermés » du KDT. *** $p < 0.001$, **** $p < 0.0001$.

L'ANOVA à mesures répétées conduite sur la puissance spectrale relative α mesurée dans le KDT lors de la phase YF a montré un effet principal de la PTS ($F(22,1) = 15.43$,

$p < 0.001$, $\eta^2 = 0.42$) : la puissance spectrale relative α a significativement diminué à la suite de la PTS ($t(21) = 3.93$, $p < 0.001$, $dz = 0.84$).

L'ANOVA à mesures répétées conduite sur la puissance spectrale relative θ mesurée dans le KDT lors de la phase YF a également montré un effet principal de la PTS ($F(22,1) = 16.96$, $p < 0.0001$, $\eta^2 = 0.45$) : la puissance spectrale relative θ a significativement augmenté à la suite de la PTS ($t(21) = -4.12$, $p < 0.0001$, $dz = -0.87$).

La diminution de la puissance spectrale relative α était de 17.03% ($\sigma = 19.14$) et l'augmentation de la puissance relative θ de 39.33% ($\sigma = 52.56$).

3.2 Paramètres acoustiques

- Caractéristiques prosodiques et de la forme de l'onde acoustique

La fréquence de modulation (Fmod), le $\langle DAL_{To} \rangle$, le $jitt(DAL_{To})$, le $\langle DAL_{To}/To \rangle$ et le $jitt(DAL_{To}/To)$ ont diminué significativement à la suite de la PTS (respectivement $t(21) = 3.64$, $p < 0.01$, $dz = 0.77$; $t(21) = 2.80$, $p < 0.05$, $dz = 0.60$; $t(21) = 2.21$, $p < 0.05$, $dz = 0.47$; $t(21) = 2.64$, $p < 0.05$, $dz = 0.56$; $t(21) = 2.23$, $p < 0.05$, $dz = 0.48$). Leur diminution était respectivement de 4.92% ($\sigma = 6.46$), 9.88% ($\sigma = 27.37$), 8.20% ($\sigma = 30.44$), 9.89% ($\sigma = 27.23$) et 5.84% ($\sigma = 34.73$).

Les autres paramètres prosodiques et de la forme de l'onde acoustique n'ont pas varié significativement à la suite de la PTS.

- Caractéristiques spectrales

Les analyses factorielles à mesures répétées ont montré un effet principal de la PTS sur le rapport harmonique sur bruit (HNR) et les coefficients MFCC3, MFCC6, MFCC7, MFCC9, MFCC10, MFCC11 et MFCC12 (respectivement $F(22,3) = 6.19$, $p < 0.5$, $\eta^2 = 0.23$; $F(22,3) = 11.26$, $p < 0.01$, $\eta^2 = 0.35$; $F(22,3) = 11.71$, $p < 0.01$, $\eta^2 = 0.36$; $F(22,3) = 5.94$, $p < 0.05$, $\eta^2 = 0.22$; $F(22,3) = 7.22$, $p < 0.05$, $\eta^2 = 0.26$; $F(22,3) = 7.06$, $p < 0.05$, $\eta^2 = 0.25$; $F(22,3) = 5.26$, $p < 0.05$, $\eta^2 = 0.20$; $F(22,3) = 5.61$, $p < 0.05$, $\eta^2 = 0.24$) : le HNR a diminué significativement ($t(21) = 2.49$, $p < 0.05$, $dz = 0.53$) tandis que les coefficients MFCC3, MFCC6, MFCC7, MFCC9, MFCC10, MFCC11 et MFCC12 ont augmenté significativement à la suite de la PTS (respectivement $t(21) = -3.36$, $p < 0.01$, $dz = -0.71$; $t(21) = -3.42$, $p < 0.01$, $dz = -0.73$; $t(21) = -2.44$, $p < 0.05$, $dz = -0.52$; $t(21) = -2.69$, $p < 0.05$, $dz = -0.58$; $t(21) = -2.66$, $p < 0.05$, $dz = -0.57$; $t(21) = -2.29$, $p < 0.05$, $dz = -0.49$; $t(21) = -2.59$, $p < 0.05$, $dz = -0.55$).

La diminution du HNR était de 1.88% ($\sigma = 3.54$) et l'augmentation des coefficients MFCC3, MFCC6, MFCC7, MFCC9, MFCC10, MFCC11 et MFCC12 était respectivement de 2.27% ($\sigma = 3.25$), 1.95% ($\sigma = 2.69$), 1.64% ($\sigma = 3.40$), 1.99% ($\sigma = 3.66$), 1.07% ($\sigma = 1.91$), 0.90% ($\sigma = 1.95$), 1.14% ($\sigma = 2.28$).

3.3 Corrélations

Des corrélations de Pearson ont été effectuées entre la variation (différence « après - avant » PTS) des puissances spectrales relatives α et θ dans le KDT lors de la phase YF et les paramètres acoustiques ayant montré une sensibilité générale (i.e., indépendamment du type de voyelle) à la PTS (i.e., Fmod, $\langle DAL_{To} \rangle$, $jitt(DAL_{To})$, $\langle DAL_{To}/To \rangle$,

jitt(DAL_{T0}/T₀), HNR, MFCC3, MFCC6, MFCC7, MFCC9, MFCC10, MFCC11, MFCC12).

Les analyses corrélationnelles ont mise en évidence une corrélation positive entre la puissance spectrale relative α et le jitt(DAL_{T0}), $r(22) = 0.50$, $p < 0.5$. Le nuage de points associé à cette corrélation significative est présenté dans la Figure 3.

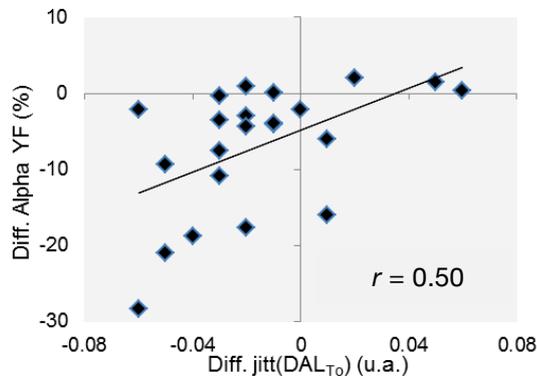


Figure 3 : Nuage de points de la corrélation significative entre la puissance spectrale relative α et le jitt(DAL_{T0}). La valeur du coefficient de corrélation de Pearson (r) est montrée.

4 Discussion

4.1 Somnolence physiologique

La privation totale de sommeil a conduit à une diminution de la puissance spectrale relative α et à une augmentation de la puissance spectrale relative θ dans le KDT lors de la phase YF. La réduction marquée du niveau d'éveil physiologique confirme que les participants étaient plus somnolents le matin après la nuit de PTS que la veille.

4.2 Paramètres acoustiques

- Caractéristiques prosodiques

Aucun changement significatif de la plupart des paramètres prosodiques n'a été observé à la suite de la PTS, notamment de la fréquence fondamentale (F₀) et de son écart type (SD_{F0}). Cette absence de variation de F₀ a déjà été rapportée par [8] et celle de SD_{F0} par [5]. Les résultats des autres études relatives aux effets de la privation de sommeil sur ces deux caractéristiques sont très contrastés. Certains auteurs [2,3,4,5] ont ainsi rapporté une diminution de F₀ après privation de sommeil tandis que [6] a observé une augmentation ; [1,3] ont montré une diminution de SD_{F0} alors qu'une augmentation a été rapportée par [8]. Par ailleurs, la réduction de la période consacrée au sommeil ou sa suppression totale a conduit dans quelques travaux à une augmentation du jitter [5,6] et du shimmer [5]. Dans cette étude, un jitter (PPQ5) et trois shimmers (Shimmer 1, Shimmer 2 et APQ11) ont eu tendance à augmenter à la suite de la PTS. Les autres jitters et shimmers (Jitter 1, Jitter 2, RAP, DDP, APQ3 et APQ5) n'ont pas varié significativement – leurs valeurs étaient toutefois plus élevées le matin à la suite de la PTS que la veille. En conséquence, la privation de sommeil pourrait conduire à une légère hausse des perturbations à court terme de F₀ et de celles de l'amplitude des pulses glottiques. La fréquence

de modulation d'amplitude (Fmod), également reliée aux pics d'amplitude de la voyelle, est la seule caractéristique prosodique ayant varié significativement à la suite de la PTS. Sa réduction indique une diminution de la fréquence de la modulation d'amplitude des pulses glottiques de la voyelle.

- Caractéristiques de la forme de l'onde acoustique

Les paramètres acoustiques relatifs à la forme du chronogramme ont été particulièrement affectés par la privation de sommeil. Ainsi, les valeurs du $\langle \text{DAL}_{T0} \rangle$, du jitt(DAL_{T0}), du $\langle \text{DAL}_{T0}/T_0 \rangle$ et du jitt(DAL_{T0}/T₀) ont diminué à la suite de la PTS, indiquant une diminution de la longueur euclidienne du signal de voyelle et de son instabilité à court-terme, et ce indépendamment de la durée de la période T₀.

- Caractéristiques spectrales

Le rapport harmonique sur bruit (HNR) a diminué à la suite de la PTS. Cette variation indique que la voix des participants pourrait être plus bruitée, avec une plus grande présence de souffle. Une interprétation possible est que l'augmentation de la somnolence consécutive à la privation de sommeil serait liée à une diminution du tonus des muscles impliqués dans la tension des cordes vocales. Ce relâchement musculaire pourrait provoquer des fuites d'air à travers la glotte lors de la phonation, conduisant à une diminution des valeurs du HNR.

Les coefficients cepstraux à échelle Mel (MFCC) ont été particulièrement sensibles à la privation de sommeil. La plupart d'entre eux (i.e., MFCC3, MFCC6, MFCC7, MFCC9, MFCC10, MFCC11 et MFCC12) ont augmenté significativement à la suite de la PTS. Les autres coefficients (i.e., MFCC1, MFCC2, MFCC4, MFCC5 et MFCC8) n'ont pas varié significativement – leurs valeurs étaient toutefois plus élevées le matin après la nuit de PTS que la veille (excepté pour le MFCC1 qui a diminué). Les coefficients MFCC offrent une représentation compacte et compressée mais abstraite du processus de production de la parole humaine. Par conséquent, contrairement aux autres caractéristiques acoustiques, aucune interprétation précise de leur variation ne peut être faite.

4.3 Corrélations

La puissance spectrale relative α dans le KDT lors de la phase yeux fermés était positivement corrélée avec le jitt(DAL_{T0}). Ceci veut dire que plus le niveau d'éveil physiologique diminuait, plus l'instabilité à court-terme des longueurs euclidiennes diminuait.

L'absence de corrélations entre les puissances relatives α et θ et les autres paramètres acoustiques peut s'expliquer par des réponses très différentes entre les individus face à la privation de sommeil. Certains participants ont montré un niveau de vigilance et/ou un signal vocal extrêmement dégradé à la suite de la PTS alors que d'autres ont été peu affectés que ce soit sur le plan de la vigilance et/ou sur le plan du signal vocal. Bien que l'expérimentation ait montré une grande variabilité interlocuteur, somme toute certainement naturelle, tant du point de vue acoustique que de celui de l'imagerie cérébrale, un paramètre acoustique a émergé des corrélations.

Par ailleurs, les mesures EEG de l'état de somnolence n'ont pas été réalisées simultanément avec la prise de son du signal vocal. Un enregistrement de l'activité corticale a été effectué durant la tâche de lecture de texte mais

l'analyse des puissances spectrales n'a pas été possible en raison de la forte perturbation du tracé EEG (i.e., signal très bruité) lors de la phonation. En outre, la production de la parole est une activité stimulante qui peut conduire à une hausse temporaire du niveau d'excitation/éveil cortical.

5 Conclusion

La fréquence de modulation d'amplitude (Fmod), la longueur euclidienne du signal et ses paramètres dérivés, le HNR et les coefficients MFCC pourraient représenter des indicateurs adaptés pour la détection d'une dette de sommeil associée à la diminution du niveau de vigilance. Ces résultats ouvrent un champ important de perspectives et d'applications notamment pour l'amélioration de la sécurité du transport aérien.

Remerciements

Les auteurs remercient chaleureusement le laboratoire du sommeil du Centre Hospitalier Universitaire de Toulouse où le protocole de privation de sommeil a été rendu possible. Nous remercions également toutes les personnes qui ont accepté de participer à l'étude.

Références

- [1] Morris, G. O., Williams, H. L., & Lubin, A. (1960). Misperception and disorientation during sleep deprivation. *A.M.A. Archives of General Psychiatry*, 2(3), 247–254.
- [2] Johannes, B., Salnitski, V. P., Gunga, H. C., & Kirsch, K. (2000). Voice stress monitoring in space – Possibilities and limits. *Aviation, Space, and Environmental Medicine*, 71(9 Suppl), A58–65.
- [3] Nwe, T., Li, H., & Dong, M. (2006). Analysis and detection of speech under sleep deprivation. In *Proceedings of Interspeech* (Pittsburgh, USA), Vol. 9, pp. 17–21.
- [4] Krajewski, J., Batliner, A., & Golz, M. (2009). Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods*, 41(3), 795–804.
- [5] McGlinchey, E. L., Talbot, L. S., Chang, K., Kaplan, K. A., Dahl, R. E., & Harvey, A. G. (2011). The Effect of Sleep Deprivation on Vocal Expression of Emotion in Adolescents and Adults. *SLEEP*, 34(9), 1233–41.
- [6] Ruiz, R., Plantin de Hugues, P., & Legros, C. (2009). Voice analysis as a significant parameter of car driver's fatigue. Proceedings of the ACOUSTICS High Tatras 2009 “34th International Acoustical Conference - EAA Symposium,” (record 10), 7–10.
- [7] Greeley, H. P., Friets, E., Wilson, J. P., Raghavan, S., Picone, J., & Berg, J. (2006). Detecting fatigue from voice using speech recognition. In *IEEE International Symposium Signal Processing and Information Technology* (Vancouver, Canada), pp. 567–571.
- [8] Vogel, A. P., Fletcher, J., & Maruff, P. (2010). Acoustic analysis of the effects of sustained wakefulness on speech. *The Journal of the Acoustical Society of America*, 128(6), 3747–56.
- [9] Krajewski, J., Schnieder, S., Sommer, D., Batliner, A., & Schuller, B. (2012). Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech. *Neurocomputing*, 84, 65–75.
- [10] Greeley, H. P., Berg, J., Friets, E., Wilson, J., Greenough, G., Picone, J., Nesthus, T. (2007). Fatigue estimation using voice analysis. *Behavior Research Methods*, 39(3), 610–619.
- [11] Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, 52(1-2), 29–37.
- [12] Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375.
- [13] Gratton, G., Coles, M. G. ., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4), 468–484.
- [14] Ruiz, R., Absil, E., Harmegnies, B., Legros, C., & Poch, D. (1996). Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions. *Speech Communication*, 20(1-2), 111–129.
- [15] Ruiz, R., Plantin de Hugues, P., & Legros, C. (2010). Advanced Voice Analysis of Pilots to Detect Fatigue and Sleep Inertia. *Acta Acustica United with Acustica*, 96(3), 567–579.
- [16] Ruiz, R., Plantin de Hugues, P., & Legros, C. (2012). Amplitude modulation of vowel glottal pulses: application to sleep inertia. In 11ème Congrès Français d'Acoustique. Nantes, 335–339.