Détection automatique de la déglutition dans les signaux d'auscultation cervicale à haute résolution

Lila GRAVELLIER[1,2] Maxime LE COZ[2] Jérôme FARINAS[1] Julien PINQUIER[1]

[1]IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France

[2]Swallis Medical, 55 avenue Louis Breguet, 31400 Toulouse, France

Résumé – L'auscultation cervicale à haute résolution est une alternative aux examens de référence visant à diagnostiquer les troubles de la déglutition. Elle consiste à placer des capteurs au niveau du cou du patient et à en analyser les signaux. Cet article présente une méthode de détection automatique de la déglutition dans les signaux issus d'un microphone et d'un accéléromètre. En combinant un détecteur d'activité et un réseau de neurones convolutionnel, cette méthode atteint 87,2% de F_{score} sur un corpus de 42 sujets sains effectuant diverses activités pharyngolaryngées.

Abstract – High resolution cervical auscultation is an alternative to reference examinations for the diagnosis of swallowing disorders: signals are captured from the patient's neck to characterize swallowing. This paper presents a method for automatic detection of swallowing in signals from a microphone and an accelerometer. By combining an activity detector based on signal energy and a convolutional neural network, this method achieves 87.2% Fscore on a corpus of 42 healthy subjects performing various pharyngolaryngeal activities.

1 Introduction

La déglutition est un mécanisme répété des milliers de fois par jour [2] permettant de guider les aliments de la bouche à l'estomac tout en protégeant les voies aériennes. Elle peut se découper en trois phases principales : orale, pharyngée et œsophagienne [10]. Si la phase orale, qui permet la préparation des aliments, est consciente, les deux autres phases impliquent quant à elles de nombreux mécanismes réflexes synchronisés pour assurer un transport des aliments efficace et sécuritaire vers l'estomac. Tout au long de cette chaîne complexe, des dysfonctionnements peuvent apparaître ; on parle alors de dysphagie. La dysphagie peut entraîner des fausses routes (entrée d'aliments dans les voies aériennes), et avoir des conséquences graves telles que la dénutrition et la pneumonie par aspiration [9]. L'examen de référence pour son diagnostic est la vidéofluoroscopie pour laquelle le patient est filmé effectuant une ou plusieurs déglutitions sous rayons X. La nasofibroscopie est aussi un examen courant consistant à introduire une caméra par le nez du patient pour observer le pharynx. Cependant, ces examens, contraignants et invasifs, peuvent difficilement servir à dépister massivement la dysphagie [1].

Depuis quelques années, de nouvelles méthodes non-invasives ont été développées pour proposer des alternatives à ces deux examens. L'auscultation cervicale à haute résolution (ACHR), consiste à placer des capteurs sur le cou du patient pour en analyser les signaux. De nombreuses études ont montré l'intérêt du microphone et de l'accéléromètre pour extraire des informations pertinentes pour le diagnostic de la dysphagie [16, 14, 3, 12, 6, 4]

La première étape pour analyser un signal de déglutition est de la détecter dans le flux enregistré sur le cou du patient. Cette tâche restant très fastidieuse manuellement, certains auteurs ont cherché à automatiser ce processus. Des études récentes ont atteint de hautes performances sur les patients [5, 7]. Pour la population saine cependant, So et al [15] ont mis en lumière les lacunes des méthodes de détection de la littérature : le nombre insuffisant de sujets (inférieur à 20), le manque d'information sur les textures ou encore l'imprécision dans les métriques d'évaluation ne permettent pas de déterminer une méthode performante.

Cet article présente une méthode répondant à ces critères pour détecter automatiquement les déglutitions de sujets sains dans les signaux vibroacoustiques captés grâce à un nouveau dispositif médical d'ACHR, le Swallis DSA.

2 Présentation du matériel et des données

Le Swallis DSA (Swallis Medical, Toulouse, France), utilisé dans cette étude pour l'enregistrement des signaux vibroacoustiques, inclut une caméra et un collier équipé d'un accéléromètre trois-axes et d'un microphone. Il permet l'acquisition synchrone de la vidéo, du son et des signaux d'accélérométrie. Les deux capteurs du collier ont une fréquence d'échantillonnage de 20 kHz.

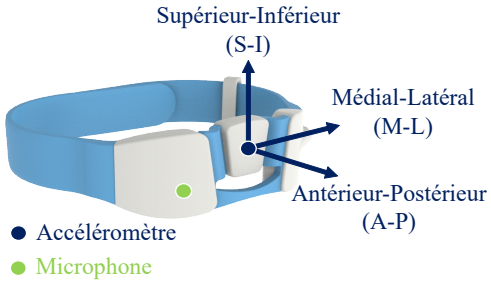

FIGURE 1 : Dispositif Swallis DSA

Pour cette étude, 42 sujets sains (19 hommes et 23 femmes), âgés entre 22 et 57 ans (34 ans en moyenne) ont été recrutés. Un protocole d'enregistrement a été établi pour acquérir de nombreux types de déglutitions, chacune caractérisée par son volume et sa texture suivant l'échelle de l'*International Dysphagia Diet Standardisation Initiative* (IDDSI). Les différentes tâches demandées aux sujets sont présentées dans la Table 1. Un seul enregistrement continu a été fait pour le protocole entier. Ainsi, de nombreuses activités pharyngolaryngées (APL) spontanées (déglutitions, parole, toux…) ont aussi été enregistrées. Au total 880 déglutitions contrôlées et 824 déglutitions spontanées ont été enregistrées dans ce corpus d'une durée totale de 5 heures 24 minutes.

TABLE 1 : Tâches supervisées du protocole

APL	Texture	Volume	Répétition
Déglutitions	Salive	-	3
	Eau (IDDSI 0)	10 mL	3
	Eau gélifiée (IDDSI 3)	2,5 mL	3
		5 mL	3
	Eau gélifiée (IDDSI 4)	2,5 mL	3
		5 mL	3
	Biscuits (IDDSI 7)	3g	3
	Eau (IDDSI 0)	10 cL	1
Hemmage	-	-	1
Toux			1
Phonation			1 phrase

3 Détection automatique des déglutitions

La fréquence de déglutition au repos d'un sujet sain est d'environ une déglutition par minute [2]. Même si cette fréquence augmente lors d'une prise alimentaire, les signaux d'intérêt restent très rares dans le flux d'enregistrement. Pour détecter efficacement la déglutition, nous avons développé deux algorithmes distincts : un détecteur d'activité pour pré-séléctionner les données à analyser ; puis un classifieur pour déterminer si l'activité est une déglutition.

3.1 Détecteur d'activité

La phase pharyngée de la déglutition est composée de trois sous-phases : l'ascension du larynx, l'ouverture du sphincter supérieur œsophagien pour le passage du bolus (aliments, salive) et enfin le retour en position initiale par le relâchement du larynx [10]. D'après Morinière, si la seconde sous-phase est toujours présente dans le signal acoustique de la déglutition, les deux autres ne sont pas forcément détectées par ce capteur [11]. Le signal sonore doit donc nécessairement comporter au moins l'une des phases de la partie pharyngée de la déglutition. Nous avons donc développé le détecteur d'activité uniquement sur le signal issu du microphone. Après calcul de l'énergie du signal acoustique avec une fenêtre de 250 ms et un pas 25 ms, ainsi qu'un lissage par une fenêtre moyennant sur 8 échantillons, un détecteur de pics (maxima locaux) a été appliqué au signal résultant. Les pics sont sélectionnés avec une distance minimale de 250 ms. Les segments d'activité correspondent à une seconde de signal centrée sur chacun des pics d'activité.

3.2 Identification de la déglutition

Pour chaque segment d'activité, les spectrogrammes des canaux sources (microphone et axes A-P, S-I et M-L de l'accéléromètre) sont calculés. Les quatre images résultantes servent de données d'entrée à un réseau de neurones convolutionnel (CNN) composé de 2 couches convolutionnelles puis d'une couche entièrement connectée. Le réseau global est détaillé dans la figure 3. Le nombre de données étant relativement limité, une méthode d'augmentation classique, le masquage temporel, a été appliquée aux données pour doubler le corpus d'entraînement. Pour cela, une zone aléatoire de 1 à 10 intervalles temporels du spectrogramme est artificiellement mise à zéro pour retirer une partie de l'information. Lorsque cette augmentation a été appliquée, chaque donnée a été mise à la fois dans sa version originale et dans sa version augmentée dans le corpus d'entraînement du modèle. En sortie, le CNN est entraîné pour attribuer au segment d'activité l'une des classes suivantes :

- 0 pour les segments ne recouvrant aucune annotation,
- 1 pour les segments ayant un recouvrement non nul avec une déglutition,
- 2 pour les segments recouvrant d'autres évènements d'intérêt hors déglutition.

Les évènements d'intérêts de la classe 2 représentent des APL telles que le rire, la phonation, la toux et les raclements de gorge. Ces APL pourraient être étudiées dans de futurs travaux pour ajouter du contexte dans la déglutition et avoir d'autres informations pertinentes dans le diagnostic de la dysphagie. Dans notre cas, la classe 2 permet au réseau de distinguer les évènements provenant de la gorge d'autres évènements sans intérêt comme des bruits d'ustensiles ou des voix d'autres locuteurs.

Une validation croisée à 7 blocs a été réalisée pour entraîner et tester le modèle. Chaque entraînement a été réalisé sur 36 enregistrements pendant 500 époques avec un taux d'apprentissage à 0,0001. Le modèle obtenu a été ensuite testé sur les 6 enregistrements restants, puis ce schéma a été réitéré 7 fois. Ce découpage permet de ne jamais avoir un même sujet dans l'entraînement et le test, afin d'estimer les performances de l'algorithme dans le cadre d'une consultation d'un nouveau sujet.

3.3 Résultats

Pour le calcul des performances, une déglutition est considérée comme détectée (D) si elle obtient un recouvrement non-nul avec au moins un segment prédit comme déglutition. Dans le cas inverse, elle est manquée (M). Un segment prédit comme déglutition en dehors de toute annotation de déglutition est quant à lui considéré comme une fausse alarme (FA). Il est ainsi possible de calculer le rappel, la précision et le F_{score} de cette méthode automatique avec les formules suivantes :

$$\text{Précision} = \frac{D}{D + FA} \quad (1)$$

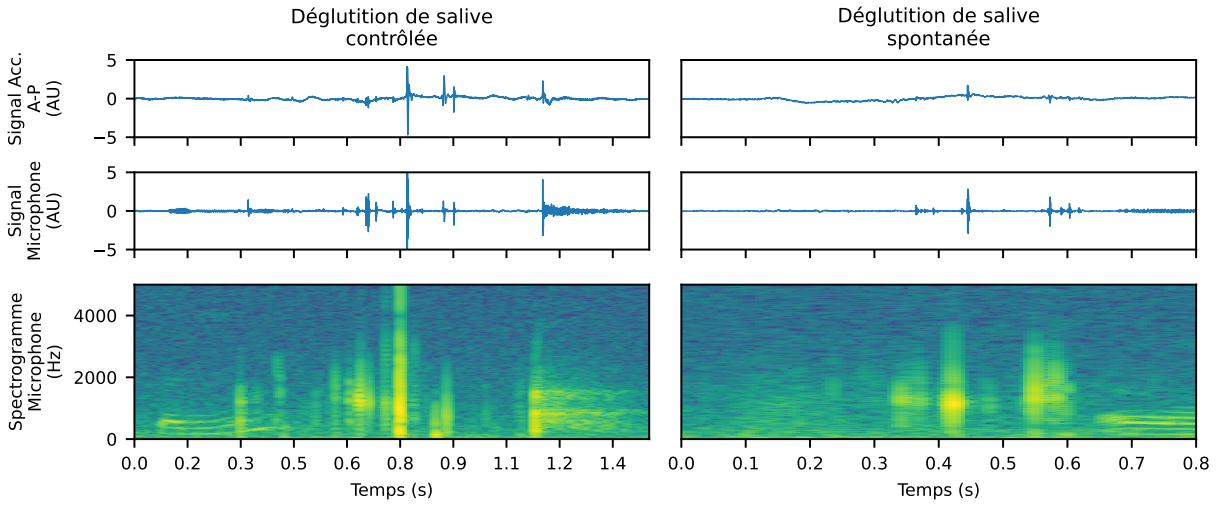

FIGURE 2 : Signaux vibroacoustiques de déglutitions de salive volontaire et spontanée effectuées par le même sujet. Les signaux temporels sont présentés avec la même échelle en unité arbitraire (AU).

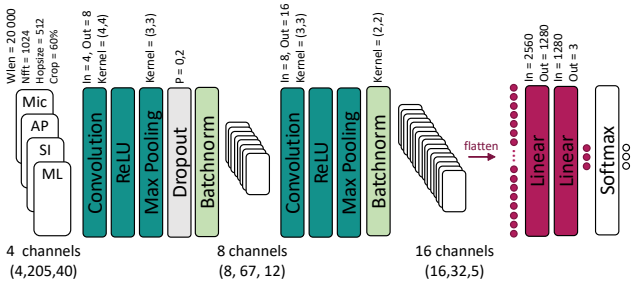

FIGURE 3 : Modèle utilisé pour la classification des segments d'activité.

$$\text{Rappel} = \frac{\text{D}}{\text{D} + \text{M}} \tag{2}$$

$$\text{F}_{\text{score}} = \frac{2 * \text{Précision} * \text{Rappel}}{\text{Précision} + \text{Rappel}} \tag{3}$$

Le détail des résultats obtenus avec et sans augmentation est présenté dans la Table 2. Le meilleur résultat est atteint en utilisant le masquage temporel, avec un F_{score} maximal de 87,2%.

TABLE 2 : Résultats de la détection de déglutition. Les erreurs du détecteur d'activité sont inclus dans les résultats (14 déglutitions manquées

Score	Sans augmentation	Masquage temporel
D	1502	1529
M	202	175
FA	265	272
Précision	85,0%	84,9%
Rappel	88,1%	89,7%
F_{score}	86,5%	**87,2%**
D spontanées	81,3%	84%
D contrôlées	94,5%	95,11%

4 Discussion

Cette méthode de détection, combinant un détecteur d'activité et un classifieur, permet de détecter 95,11% des déglutitions contrôlées, demandées par le protocole. Cependant, 16% des déglutitions spontanées ne sont pas détectées. Ces résultats semblent indiquer que l'aspect volontaire ou spontané de la déglutition a un impact sur son motif vibroacoustique. Ceci est cohérent avec la littérature qui montre que la déglutition présente une grande variabilité [8], inter-individuelle mais aussi intra-individuelle, ce qui peut rendre la tâche de détection plus difficile. La figure 2 illustre cette variabilité en présentant deux déglutitions de salive effectuées par la même personne, volontairement et spontanément. La déglutition spontanée est d'une part beaucoup moins énergétique, avec une amplitude environ 10 fois plus petite pour le son et l'accélérométrie A-P, mais aussi deux fois plus courte. Ces grandes différences entre les signaux peuvent expliquer la difficulté du modèle à être performant dans les deux cas. On retrouve cet aspect dans la littérature : les performances sont très hautes dans le cas de déglutitions enregistrées en simultané avec une vidéofluorocopie [5], qui impose une position fixe et un contrôle conscient de la déglutition. En effet le patient doit déglutir au moment de l'activation des rayons X et ne doit pas bouger pour rendre l'imagerie la plus nette possible. Seuls Sazonov et al. [13] ont développé une méthode de détection sur plus de 20 sujets sains dans des conditions écologiques et ont atteint 84,7% d'accuracy pondérée. Il sera intéressant de s'attarder sur ces différences de conditions et de spécialiser un modèle pour chaque type de déglutition. Les déglutitions contrôlées les moins bien détectées sont celles de biscuit et de salive (respectivement 8,6% et 7,5% non détectées). Ces bolus plus pâteux mènent à des déglutitions avec effort, ce qui peut modifier le motif de la déglutition dans les signaux.

Dans un contexte d'analyse de la déglutition, il est important de maximiser le nombre de détections. Cependant, pour des tâches comme le calcul de fréquence de déglutition, les fausses alarmes peuvent changer le diagnostic. La tendance de l'algorithme actuel à la sur-prédiction ne le rend pas, en l'état, adapté à ce type de tâche. On observe néanmoins, un tiers des fausses alarmes (88/272) dans des périodes de mastication avant la déglutition du biscuit. Ces périodes étant bruitées simultanément dans les deux capteurs, une détection au préalable de la mastication permettrait de pré-traiter les signaux afin d'améliorer les résultats. Enfin, des caractéristiques propres à certains enregistrements (bruits de barbe sur le micro, collier trop lâche) ont mené à d'importantes erreurs : 45% des erreurs totales sont contenues dans seulement 5 enregistrements. Ces caractéristiques pouvant survenir lors d'une auscultation, il sera intéressant de travailler sur ces cas particuliers pour améliorer la robustesse de cet algorithme. En l'état, cet algorithme pourrait permettre de structurer les signaux et aider un praticien à repérer les instants d'intérêt dans le flux d'un enregistrement.

5 Conclusion

La combinaison d'un détecteur d'activité et d'un réseau CNN permet de détecter les déglutitions de 42 sujets sains dans des signaux ACHR avec un F_{score} de 87,2%. Une meilleure compréhension des motifs de déglutitions permettra de mettre en place des actions spécifiques pour améliorer les performances sur les cas particuliers et aider les professionnels de santé dans le diagnostic de la dysphagie.

6 Remerciements

Ce travail est réalisé dans le cadre d'un doctorat de l'Université Toulouse III, dans le cadre d'une convention CIFRE n°2021/44 entre l'Association Nationale pour la Recherche et la Technologie et Swallis Medical.

Références

[1] O. BIRCHALL, M. BENNETTE, N. LAWSON, S. M. COTTON et A. P. VOGEL : Instrumental Swallowing Assessment in Adults in Residential Aged Care Homes : A Scoping Review. *Journal of the American Medical Directors Association*, 22(2):372–379.e6, février 2021.

[2] W. J. DODDS, E. T. STEWART et J. A. LOGEMANN : Physiology and radiology of the normal oral and pharyngeal phases of swallowing. *American Journal of Roentgenology*, 154(5):953–963, mai 1990.

[3] C. DONOHUE, S. MAO, E. SEJDIĆ et J. L. COYLE : Tracking Hyoid Bone Displacement During Swallowing Without Videofluoroscopy Using Machine Learning of Vibratory Signals. *Dysphagia*, 36(2):259–269, avril 2021.

[4] J. DUDIK, J. COYLE, A. EL-JAROUDI, M. Sun Z. MAO et E. SEJDIĆ : Deep learning for classification of normal swallows in adults. *Neurocomputing*, 285:1–9, avril 2018.

[5] Y. KHALIFA, J. L. COYLE et E. SEJDIĆ : Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings. *Scientific Reports*, 10(1):8704, décembre 2020.

[6] Y. KHALIFA, C. DONOHUE, J. L. COYLE et Ervin E. SEJDIC : Upper Esophageal Sphincter Opening Segmentation With Convolutional Recurrent Neural Networks in High Resolution Cervical Auscultation. *IEEE Journal of Biomedical and Health Informatics*, 25(2):493–503, février 2021.

[7] N. KURAMOTO, K. ICHIMURA, D. JAYATILAKE, T. SHIMOKAKIMOTO, K. HIDAKA et K. SUZUKI : Deep Learning-Based Swallowing Monitor for Realtime Detection of Swallow Duration. *In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4365–4368, Montreal, QC, Canada, juillet 2020. IEEE.

[8] G. L. LOF et J. ROBBINS : Test-retest variability in normal swallowing. *Dysphagia*, 4(4):236–242, décembre 1990.

[9] J. A. LOGEMANN : Dysphagia : Evaluation and Treatment. *Folia Phoniatrica et Logopaedica*, 47(3):140–164, 1995. Publisher : Karger Publishers.

[10] A. J. MILLER : Deglutition. *Physiological Reviews*, janvier 1982.

[11] S. MORINIÈRE, M. BOIRON, D. ALISON, P. MAKRIS et P. BEUTTER : Origin of the Sound Components During Pharyngeal Swallowing in Normal Subjects. *Dysphagia*, 23(3):267–273, septembre 2008.

[12] M. S. NIKJOO, C. M. STEELE, E. SEJDIĆ et T. CHAU : Automatic discrimination between safe and unsafe swallowing using a reputation-based classifier. *BioMedical Engineering OnLine*, 10(1):100, 2011.

[13] E. S. SAZONOV, O. MAKEYEV, S. SCHUCKERS, P. LOPEZ-MEYER, E. L. MELANSON et M. R. NEUMAN : Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behavior. *IEEE Transactions on Biomedical Engineering*, 57(3):626–633, mars 2010.

[14] E. SEJDIC, C. M. STEELE et T. CHAU : Classification of Penetration–Aspiration Versus Healthy Swallows Using Dual-Axis Swallowing Accelerometry Signals in Dysphagic Subjects. *IEEE Transactions on Biomedical Engineering*, 60(7):1859–1866, juillet 2013.

[15] B. P. SO, T. T. CHAN, L. LIU, C. C. YIP, H. LIM, W. LAM, D. W. WONG, D. S. K. CHEUNG et J. C. CHEUNG : Swallow Detection with Acoustics and Accelerometric-Based Wearable Technology : A Scoping Review. *International Journal of Environmental Research and Public Health*, 20(1):170, décembre 2022.

[16] C. M. STEELE, E. SEJDIĆ et T. CHAU : Noninvasive Detection of Thin-Liquid Aspiration Using Dual-Axis Swallowing Accelerometry. *Dysphagia*, 28(1):105–112, mars 2013.