



HAL
open science

Apprentissage profond pour l'estimation du quotient ouvert à partir du signal électroglottographique

Minh-Châu Nguyễn, Maximin Coavoux, Solange Rossato

► To cite this version:

Minh-Châu Nguyễn, Maximin Coavoux, Solange Rossato. Apprentissage profond pour l'estimation du quotient ouvert à partir du signal électroglottographique. Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL), GDR LIFT; GDR TAL, Nov 2022, Marseille, France. pp.29-38. hal-03846833

HAL Id: hal-03846833

<https://hal.science/hal-03846833v1>

Submitted on 14 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apprentissage profond pour l'estimation du quotient ouvert à partir du signal électroglottographique

Minh-Châu Nguyễn Maximin Coavoux Solange Rossato

Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG

minhchau.ntm@gmail.com, maximin.coavoux@univ-grenoble-alpes.fr,
Solange.Rossato@univ-grenoble-alpes.fr

MOTS-CLÉS : Électroglottographie, quotient ouvert glottique, apprentissage profond.

KEYWORDS: Electroglottography, glottal open quotient, deep learning.

1 Introduction

Contexte Les outils du Traitement Automatique des Langues (TAL) peuvent venir en appui à l'étude des langues rares pour les tâches de documentation fondamentale (telles que la transcription automatique de la parole, Adams *et al.*, 2018; Foley *et al.*, 2018). Dans cet article, nous cherchons à déterminer dans quelle mesure l'apprentissage machine peut également faciliter le traitement d'autres types de données expérimentales collectées sur ces langues (au sujet de l'éventail des techniques utilisées en phonétique, voir Vaissière *et al.* 2010). Spécifiquement, notre travail porte sur les enregistrements électroglottographiques (EGG). L'électroglottographie est une méthode non invasive pour estimer l'évolution de la surface d'accolement des plis vocaux au cours de la parole (Fabre, 1957; Fourcin *et al.*, 1995). Nous présentons ici les résultats d'expériences qui visent à automatiser l'estimation du quotient d'ouverture glottique (O_q) à partir du signal électroglottographique. Cette tâche est relativement incertaine en l'absence de vérification manuelle (Orlikoff, 1998; Herbst, 2020), et chronophage dans l'approche semi-automatique pratiquée dans plusieurs travaux (Michaud, 2004a; Recasens & Mira, 2013; Michaud *et al.*, 2015; Gao, 2015).

Objet de l'étude Le quotient d'ouverture glottique (O_q , unité : %) est le rapport entre la durée de la phase ouverte (entre une ouverture et la fermeture suivante) et celle du cycle glottique entier : $O_q = (\text{phase ouverte}) / (\text{phase ouverte} + \text{phase fermée})$. O_q est couramment considéré comme un paramètre lié au type de phonation : un O_q bas indique une phonation *pressée* ; un O_q moyen s'observe en voix modale ; et un O_q élevé est le signe d'une phonation fluide, à haut débit d'air (voix murmurée, voix soufflée). Ce paramètre peut être estimé au moyen du signal électroglottographique (Henrich *et al.*, 2004). La détection de pics positifs dans la dérivée première du signal EGG permet d'estimer l'instant de fermeture glottique, et un pic négatif unique et bien marqué entre deux pics positifs est considéré comme l'indice de l'instant d'ouverture de la glotte (figure 1).

L'estimation de l' O_q nécessite donc la détection de l'instant d'ouverture de la glotte. Une difficulté bien identifiée est qu'il est relativement courant que les pics d'ouverture ne soient pas uniques et bien marqués. Là où la mesure de la fréquence fondamentale (f_0) s'appuie sur la détection des pics de fermeture, qui dans la grande majorité des cas sont bien marqués, la détection des pics d'ouverture se heurte souvent à des difficultés dues à des pics imprécis. Parfois, aucun pic ne se détache clairement. Parfois, deux pics

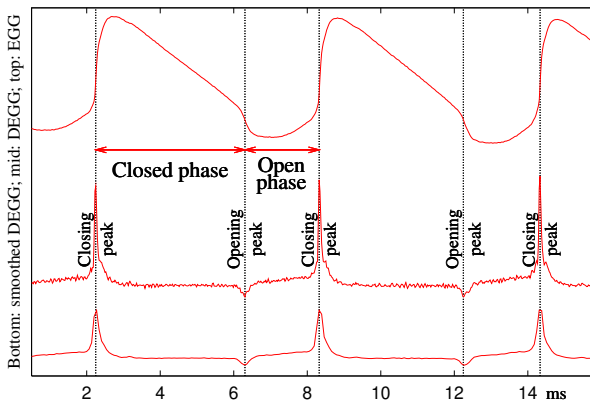


FIGURE 1 – Exemple de signaux EGG et dEGG avec indication de la fermeture et de l’ouverture de la glotte. Reproduit avec la permission de l’auteur, Alexis Michaud.

ou plus sont présents (pics multiples). La recherche des pics d’ouverture est particulièrement délicate dans le cas d’une phonation non modale, par exemple lorsque le voisement passe à la voix craquée (comme en figure 2). La vérification de l’ O_q nécessite de poser des critères pour les situations particulières. Quand l’œil exercé peut-il, avec un bon degré de confiance, estimer la position de l’instant d’ouverture ? Quand faut-il renoncer à estimer le quotient ouvert ? La présente étude aborde cette question en entraînant un modèle neuronal sur un corpus annoté manuellement (par la première autrice du présent travail).

Corpus Le corpus utilisé provient d’un dialecte du muong (Vietnam). Cette langue possède un système tonal complexe au plan phonétique, combinant des contours de f_0 et des caractéristiques phonatoires. En particulier, dans le système de 5 + 2 tons¹, l’un des tons comporte un passage en voix craquée (Nguyen, 2021). Le signal EGG a été enregistré simultanément avec le signal acoustique pour 20 locuteurs (10 hommes, 10 femmes). Le corpus audio entier est en libre accès dans la collection Pangloss² (pour l’accès aux fichiers électroglottographiques, contacter la première autrice du présent travail). Le jeu de données utilisé ici est constitué de 12 ensembles de 5 mots monosyllabiques qui s’opposent par leur ton (“ensembles minimaux”) parmi les syllabes sans occlusive finale et 3 paires minimales tonales parmi les syllabes à occlusive finale. Le corpus comprend des syllabes cibles à la fois isolées (“forme de citation”) et dans une phrase-cadre de quatre mots (y compris le mot cible). La durée d’enregistrement de chaque locuteur est d’environ 15 minutes, donc environ 5 heures pour l’ensemble des données de 20 locuteurs. Il contient en tout près de 13 000 syllabes, ce qui représente 420 000 cycles glottiques.

Pour chaque cycle glottique, on dispose de la fréquence fondamentale (f_0) et de 5 valeurs de quotient ouvert. Les 4 premières sont le fruit d’une analyse automatique utilisant **PeakDet** (<https://github.com/alexis-michaud/egg>). Cet algorithme permet de choisir entre deux méthodes : détection des maxima sur la dérivée du signal EGG, ou méthode des barycentres, chacune s’appliquant avec ou sans lissage du signal (équivalent à un filtrage passe-bas). La 5^e valeur est le fruit d’une étape manuelle consistant à annoter chaque cycle glottique par une ou plusieurs des étiquettes suivantes :

1. Le système tonal du dialecte en question oppose 5 tons dans les syllabes sans occlusive finale (syllabes ouvertes sans consonne finale, et syllabes se terminant par une coda nasale) et 2 tons dans les syllabes à occlusive finale (/p-, -t, -k, -c/).

2. <https://pangloss.cnrs.fr/> (Michaud *et al.*, 2016).

(0) pas d' O_q calculable, (1) et (2) choix de la méthode des maxima respectivement sans et avec lissage, (3) et (4) choix de la méthode des barycentres sans et avec lissage. La vérification revient ainsi, pour chaque cycle, à indiquer une double information : si l'estimation du quotient ouvert à partir du signal EGG paraît praticable pour le cycle en question; et si tel est le cas, quelle méthode donne l'estimation la plus adéquate.

La figure 2 illustre un cas de signal d'EGG bruité même après lissage. Il a été choisi d'exclure le cycle du milieu et le dernier (les 3^e et 5^e sur la fenêtre de cinq cycles mis en relief sur la figure). Le troisième comporte deux pics négatifs (ou deux "bosses" négatives), dont aucun n'est vraiment plus marqué que l'autre. Le cinquième présente un minimum clair, mais qui n'a pas une forme nette de pic. Pour les trois autres cycles, en présence d'un pic unique qui paraît bien visible, on retient ce pic sur le signal d'EGG lissé (méthode 2).

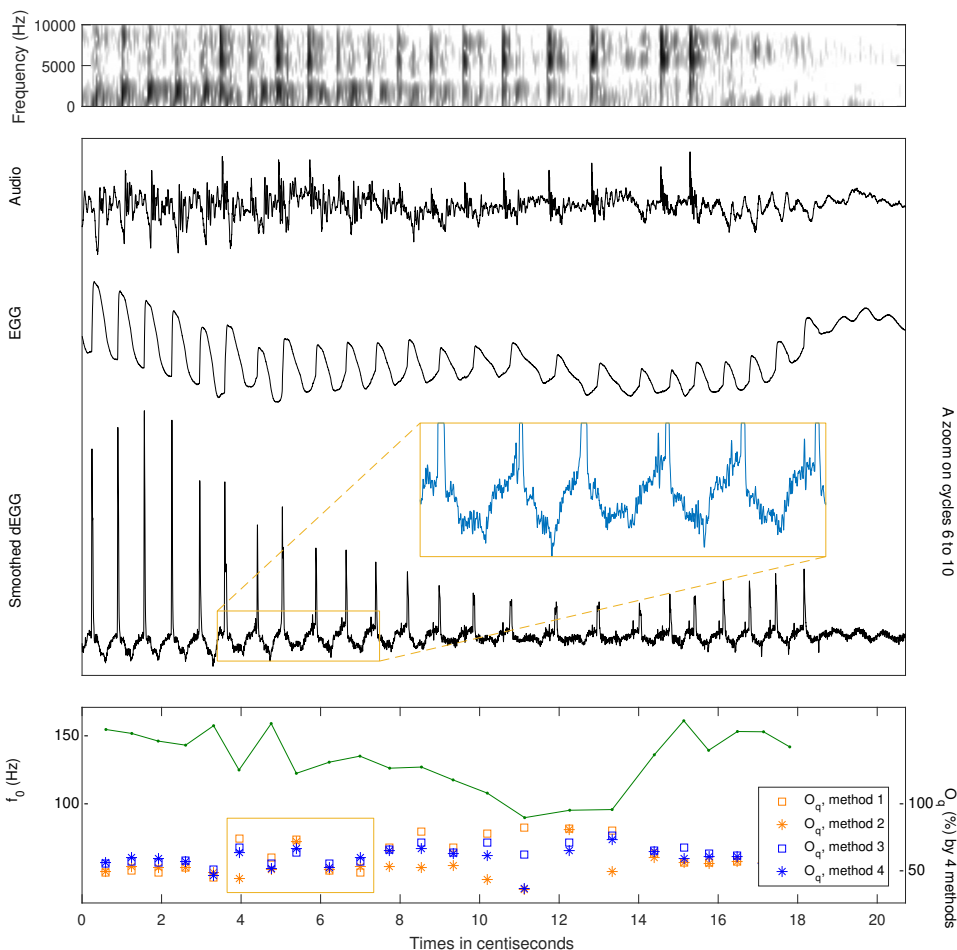


FIGURE 2 – Exemple de cas où la méthode 2 (minimum local sur le signal EGG lissé) a été retenue pour les cycles 1, 2 et 4; aucune valeur retenue pour les cycles 3 et 5. Données de locuteur F3, syllabe : /na4/ "tir à l'arc" (DOI : <https://doi.org/10.24397/pangloss-0006761#W90>).

Dans la figure 3, le signal est moins bruité, et la dérivée lissée est assez claire. Dans le 3^e des six cycles mis en valeur, un pic unique se détache. Pour les cycles 1, 2 et 4, le pic est globalement net mais bifurque

à son sommet, en forme de “fourche à deux dents”. Pour les cycles 5 et 6, il y a clairement deux pics, mais proches l’un de l’autre (la différence entre les estimations de O_q n’est que de l’ordre de 5% selon qu’on choisit le premier ou le second). Au vu de cette situation, on retient la méthode 4 : un barycentre entre les pics voisins (au sein du même cycle), pondéré par la hauteur des pics. Pour plus de détails sur l’algorithme de calcul du barycentre, nous renvoyons à la documentation du script PeakDet.

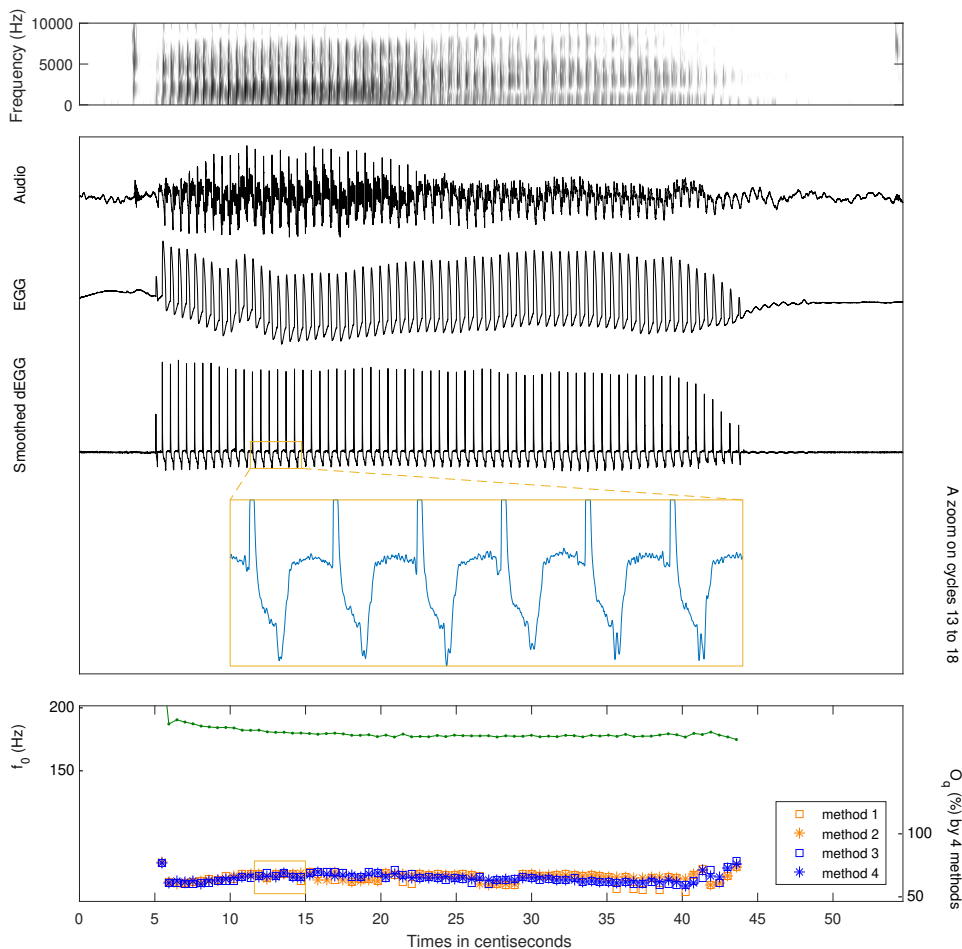


FIGURE 3 – Exemple de cas où l’option du barycentre des pics sur le signal dEGG lissé (méthode 4, représentée par des étoiles bleues) a été choisie. Données du locuteur F3, syllabe : /kajl/ “cabbage” (DOI : <https://doi.org/10.24397/pangloss-0006761#W107>).

La méthode (3) n’est choisie que pour moins d’une centaine de cycles glottiques. En effet, le recours à un barycentre entre pics (méthodes 3 et 4) a lieu dans les cas où le pic n’est pas unique et bien marqué, or dans de tels cas, il serait paradoxal de ne pas recourir au lissage du signal (donc à la méthode 4 plutôt qu’à la méthode 3).

Le fruit des décisions ainsi réalisées sur l’ensemble du corpus est considéré comme une “vérité terrain” (*gold standard*) dans les expériences d’apprentissage neuronal exposées ci-dessous.

Section	Entraînement	Développement	Test	Corpus complet
Nombre de syllabes	9050	1913	2011	12974
Nombre de cycles glottiques	295753	62350	65727	423830
Distribution des étiquettes				
0 = aucun	69237 (23.41%)	14583 (23.39%)	14670 (22.32%)	98490 (23.24%)
1 = maxima sans lissage	60527 (20.47%)	13227 (21.21%)	13461 (20.48%)	87215 (20.58%)
2 = maxima avec lissage	137461 (46.48%)	29909 (47.97%)	30386 (46.23%)	197756 (46.66%)
3 = barycentre sans lissage	42603 (14.40%)	7870 (12.62%)	10102 (15.37%)	60575 (14.29%)
4 = barycentre avec lissage	93583 (31.64%)	18775 (30.11%)	21723 (33.05%)	134081 (31.64%)

TABLE 1 – Statistiques sur le corpus. Les méthodes implémentées par PeakDet donnant parfois les mêmes valeurs d’ O_q , certains cycles ont plusieurs étiquettes correctes. Par conséquent, la somme des pourcentages des étiquettes peut dépasser 100%.

À la lumière des explications ainsi fournies au sujet du corpus de départ, nous pouvons maintenant passer à l’exposé de nos expérimentations. Elles ont pour but d’automatiser la décision manuelle concernant O_q , particulièrement chronophage. Pour le besoin des expérimentations, nous avons divisé le corpus en 3 sections : apprentissage (70%), validation (15%), test (15%). Nous présentons dans la table 1 quelques statistiques sur le corpus, dont la distribution des étiquettes. Nous observons ainsi que pour environ 23% du corpus, il n’y a pas d’ O_q calculable. L’étiquette (2) est manuellement sélectionnée, seule ou avec d’autres méthodes, pour 46% des cycles, tandis que ce taux reste à 14% pour la méthode (3), montrant une répartition déséquilibrée des étiquettes.

2 Expérimentations

Modèle Nous avons implémenté un réseau de neurones basé sur un LSTM bidirectionnel destiné à prédire pour chaque cycle glottique, le résultat de l’annotation manuelle décrite dans la section précédente. Le réseau prend en entrée le signal EGG et éventuellement des informations additionnelles pour chaque cycle glottique : les chronocodes de chaque cycle (temps de début et de fin), la f_0 , et les valeurs d’ O_q estimées par les 4 méthodes implémentées par PeakDet. Pour représenter le signal EGG, nous utilisons des MFCC, avec une fenêtre de 6 ms glissante toutes les 2 ms (ces valeurs sont faibles par rapport aux valeurs habituellement utilisées en traitement de la parole, pour tenir compte de la granularité des représentations dont nous avons besoin). Nous obtenons une matrice $\mathbf{M}^{(0)}$ de taille $N \times F$ où N est le nombre de trames MFCC (c’est-à-dire la longueur du signal en millisecondes divisée par 2) et F est le nombre de traits MFCC extraits pour chaque trame. Ensuite, cette matrice est donnée en entrée à un réseau à propagation avant :

$$\mathbf{M}^{(1)} = \tanh(\mathbf{W}^{(1)} \cdot \text{LayerNorm}(\mathbf{M}^{(0)}) + \mathbf{b}^{(1)}),$$

et contextualisée à l’aide d’un LSTM bidirectionnel :

$$\mathbf{M}^{(2)} = \text{bi-LSTM}(\mathbf{M}^{(1)}). \quad (1)$$

Pour représenter chaque cycle glottique c , nous utilisons la concaténation de 3 vecteurs $\mathbf{v}_c = [\mathbf{M}_{c_d}^{(2)}; \mathbf{M}_{c_f}^{(2)}; \mathbf{o}_c]$, où c_d et c_f sont les indicateurs temporels respectifs du début et de la fin du cycle, et $\mathbf{v}_c \in \mathbb{R}^5$ est un vecteur de traits additionnels contenant les 4 valeurs de quotient ouvert ainsi que la f_0 du

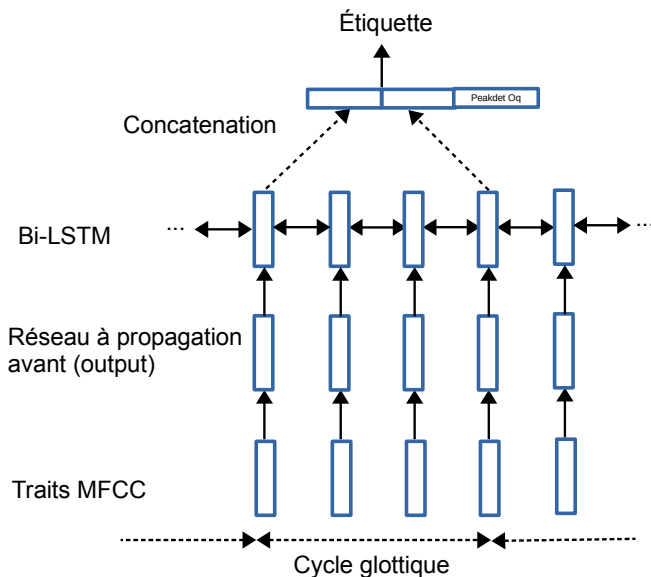


FIGURE 4 – Illustration du fonctionnement du réseau de neurones (centré sur un seul cycle glottique). En pratique, le bi-LSTM encode une syllabe complète.

cycle. Enfin, pour réaliser une prédiction pour un cycle, nous utilisons un second réseau à propagation avant :

$$\mathbf{P} = \text{Sigmoid}(\mathbf{W}^{(3)} \cdot \text{ReLU}(\mathbf{W}^{(2)} \cdot \text{LayerNorm}(\mathbf{v}_c) + \mathbf{b}^{(2)}) + \mathbf{b}^{(3)}),$$

où chaque $\mathbf{P} = [P(y_0 = 1|c), \dots, P(y_4 = 1|c)]$ donne la probabilité est la probabilité que chaque étiquette soit correcte. Le réseau de neurones est illustré en figure 4.

Nous optimisons tous les paramètres du réseau à l'aide de l'algorithme Adam (Kingma & Ba, 2015), par maximisation des probabilités des étiquettes de référence. Lors de l'évaluation du modèle, nous considérons l'étiquette de plus forte probabilité comme prédiction du réseau. Les hyperparamètres du système sont les suivants :

- pour les MFCCs : taille des fenêtres (6ms), taille du pas entre chaque fenêtre (2ms), taille du contexte (2 trames de chaque côté);
- pour le réseau : dimension des couches cachées (128 unités pour les réseaux à propagation avant et 128 pour chaque direction du LSTM bidirectionnel);
- pour l'entraînement : algorithme d'optimisation (Adam), pas d'apprentissage (0.008 pour les modèles utilisant un signal, 0.001 pour le modèle sans signal), taille des *batches* (8), nombre d'époques.

Nous avons calibré les hyperparamètres lors d'expériences préliminaires (en particulier le pas d'apprentissage) sur le corpus de validation. Pour les expériences finales dont nous rapportons les résultats plus bas, nous entraînons chaque modèle pour 100 époques et sélectionnons l'époque qui maximise l'exactitude sur le corpus de validation avant évaluation finale sur le corpus de test.

Configurations expérimentales Nous cherchons à déterminer si l'utilisation du signal EGG permet d'améliorer la prédiction des étiquettes et s'il apporte une information supplémentaire aux valeurs d' O_q

Modèle	Validation			Test		
	Exactitude-3	Exactitude-2	Fscore-2	Exactitude-3	Exactitude-2	Fscore-2
Baseline (classe la plus fréquente)	48.9	76.6	0	46.2	77.7	0
(i) EGG + O_q + f_0	63.6	86.1	91.0	58.4	85.8	91.0
(ii) EGG	56.9	80.6	87.7	53.2	78.2	86.1
(iii) O_q + f_0	58.7	83.2	89.1	56.9	83.6	89.5
(iv) Audio + O_q + f_0	63.4	85.9	91.0	59.4	85.2	90.6
(v) Audio	57.1	78.9	86.6	51.5	79.2	87.3

TABLE 2 – Résultats finaux sur les sections de développement et de test (%).

pred ↓ gold →	0	1	1/2	2	2/3	2/3/4	2/4	3/4	4
0	61.77*	6.9	6.46	9.24	4.11	1.1	1.61	3.27	13.13
1	0	0*	0*	0.01	0	0	0	0	0
2	26.79	89.66	73.2*	72.67*	76.71*	74.88*	79.03*	60.23	50.35
3	0.04	0	0.29	0.22	0*	0.37*	0	0.46*	0.08
4	11.4	3.45	20.05	17.87	19.18	23.66*	19.35*	36.04*	36.44*

TABLE 3 – Matrice de confusion (corpus de test) pour le modèle (iv). Les valeurs données sont des pourcentages calculées sur les étiquettes gold. Les prédictions correctes sont indiquées par *.

et f_0 estimés par PeakDet. Par ailleurs, nous cherchons à savoir si le signal EGG est plus informatif que le simple signal audio. Nous évaluons ainsi plusieurs types d’entrée différents pour étudier le comportement du modèle :

- (i) signal EGG + O_q estimé par PeakDet + f_0 : $[M_{c_d}^{(2)}; M_{c_f}^{(2)}; \mathbf{o}_c]$;
- (ii) signal EGG seul : $[M_{c_d}^{(2)}; M_{c_f}^{(2)}]$;
- (iii) O_q et f_0 estimés par PeakDet : $[\mathbf{o}_c]$;
- (iv) comme (i) mais en utilisant le signal audio à la place du signal EGG;
- (v) comme (ii) mais en utilisant le signal audio à la place du signal EGG.

Résultats Nous rapportons les résultats dans le tableau 2. Les métriques d’évaluation sont : l’exactitude à 3 classes (0 vs 1-2 vs 3-4), l’exactitude et le Fscore à 2 classes (étiquette 0 vs autre étiquette), pour les 5 modèles, ainsi qu’une baseline choisissant systématiquement la classe la plus fréquente. Les valeurs s’élèvent (modestement) au-dessus de la baseline, signe qu’il y a eu apprentissage. Les résultats obtenus avec le signal EGG seul en entrée (ii) montrent que celui-ci n’est pas exploité à plein, puisque les résultats sont moins bons qu’en (iii) lorsque l’on utilise uniquement les 4 méthodes de PeakDet et la f_0 . Lorsque les deux informations sont présentes (i), les performances sont meilleures. Par ailleurs, les valeurs pour les entrées (i) et (iv) sont quasi-identiques, indiquant que l’information du signal EGG et celle du signal acoustique sont sensiblement équivalentes pour le système, en complément des estimations de O_q par PeakDet.

Matrice de confusion Nous présentons la matrice de confusion du modèle (iv), qui a obtenu les meilleurs résultats dans le tableau 3. Chaque colonne représente une combinaison d’étiquettes vue dans les données de référence (par exemple, la colonne 3/4 représente les cycles glottiques où les méthodes PeakDet 3 et 4 ont donné la bonne valeur. On observe que les étiquettes 1 et 3, qui sont les moins présentes dans les données (et donnent souvent le même résultat respectivement que les méthodes 2 et 4), ne sont quasiment

jamais prédites, ce qui montre que le système ne discrimine pas les classes 1 et 2 d'une part et les classes 2 et 3 d'autre part. Les erreurs les plus fréquentes sont la difficulté à prédire les classes 3, 4 et 0, où le modèle se replie sur la classe la plus fréquente (2).

3 Perspectives

Études des types de phonation Le travail pluridisciplinaire décrit ici a des implications concernant l'étude phonétique des types de phonation. En effet, l'ensemble de résultats obtenu ici met en lumière le fait que le processus qui consiste à évaluer la fiabilité de l'estimation du quotient ouvert au vu du signal constitue une tâche non triviale. Cela fournit l'occasion de revenir sur la question fondamentale de ce que reflète le quotient ouvert glottique, et de son interprétation. Le quotient ouvert constitue une projection linéaire de phénomènes non linéaires qui entrent en jeu dans la phonation, et ne saurait donc, à l'évidence, constituer par lui-même un descripteur suffisant des divers types de phonation (voix murmurée, voix pressée, voix craquée; parmi les référentiels couramment employés, voir notamment [Laver, 1980](#)). Spécifiquement dans le cas de la voix craquée, illustrée ci-dessus par la figure 2, on observe une bonne corrélation entre la présence de voix craquée et l'information de pente spectrale reflétée dans le spectrogramme (une intensité plus forte dans la moitié supérieure du spectrogramme, de 5 à 10 kHz, que dans la moitié inférieure : cycles 2, 6, 8, 9, 13, 14, 15, 17, 18). Tandis que le quotient ouvert glottique pour les cycles en question ne montre pas de valeurs exceptionnellement basses. Le signal audio est donc ici un meilleur outil que le quotient ouvert pour détecter la voix craquée. Le signal EGG contient d'autres informations, à commencer par la fréquence fondamentale, qui fournissent des indications plus claires que O_q concernant le type phonatoire.

Au plan acoustique, il est connu que le quotient ouvert n'est ni le seul, ni le plus important parmi les paramètres de source glottique. Le fait qu'on puisse en obtenir une estimation à partir du signal EGG a sans doute amené à lui accorder une importance particulière, en comparaison par exemple du quotient de vitesse (*speed quotient*), qui, lui, n'est pas aisément accessible à une estimation. Ainsi, la hauteur du pic positif sur la dérivée (correspondant à l'instant de fermeture glottique en début et fin de cycle), DECPA (pour Derivative-Electroglottographic Closure Peak Amplitude, [Michaud, 2004b](#)), ne permet hélas pas d'estimer directement le quotient de vitesse. Clairement, O_q gagnerait donc à être intégré à des expériences d'apprentissage machine dans lesquelles il serait intégré au sein d'un ensemble élargi de paramètres acoustiques, de façon à caractériser divers types de phonation d'une façon à la fois objective et complète.

Perspectives pour la suite du travail Dans la suite de ce travail, nous prévoyons de développer et d'évaluer des modèles de régression permettant de prédire directement la variable O_q , au lieu de la prédire indirectement via une tâche de classification. Par ailleurs, nous prévoyons d'évaluer le remplacement des traits MFCC par des traits extraits par des modèles acoustiques préentraînés ([Conneau et al., 2020](#)).

Remerciements

Les travaux présentés dans cet article sont financés dans le cadre du projet franco-allemand "La documentation automatique des langues à l'horizon 2025" (Computational Language Documentation by 2025, CLD 2025, ANR-19-CE38-0015-04). Nous remercions les relecteurices anonymes pour leurs commentaires. Merci à Alexis Michaud pour de nombreuses discussions et sa relecture attentive.

Références

- ADAMS O., COHN T., NEUBIG G., CRUZ H., BIRD S. & MICHAUD A. (2018). Evaluating phonemic transcription of low-resource tonal languages for language documentation. In *LREC 2018 (Language Resources and Evaluation Conference)*, p. 3356–3365. HAL : [halshs-01709648](https://halshs.archives-ouvertes.fr/halshs-01709648).
- CONNEAU A., BAEVSKI A., COLLOBERT R., MOHAMED A. & AULI M. (2020). Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv :2006.13979*.
- FABRE P. (1957). Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation : glottographie de haute fréquence. *Bulletin de l'Académie Nationale de Médecine*, **141**, 66–69.
- FOLEY B., ARNOLD J. T., COTO-SOLANO R., DURANTIN G., ELLISON T. M., VAN ESCH D., HEATH S., KRATOCHVIL F., MAXWELL-SMITH Z., NASH D. *et al.* (2018). Building speech recognition systems for language documentation : The CoEDL Endangered Language Pipeline and Inference System (ELPIS). In *SLTU*, p. 205–209.
- FOURCIN A., ABBERTON E., MILLER D. & HOWELLS D. (1995). Laryngograph : speech pattern element tools for therapy, training and assessment. *European Journal of Disorders of Communication*, **30**(2), 101–115.
- GAO J. (2015). *Interdependence between tones, segments and phonation types in Shanghai Chinese*. Ph.D., Université Sorbonne Nouvelle.
- HENRICH N., D'ALESSANDRO C., DOVAL B. & CASTELLENGO M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America*, **115**(3), 1321–1332.
- HERBST C. T. (2020). Electroglottography—an update. *Journal of Voice*, **34**(4), 503–526.
- KINGMA D. P. & BA J. (2015). Adam : A method for stochastic optimization. *CoRR*, **abs/1412.6980**.
- LAVER J. (1980). *The phonetic description of voice quality*. Cambridge : Cambridge University Press.
- MICHAUD A. (2004a). Final consonants and glottalization : new perspectives from Hanoi Vietnamese. *Phonetica*, **61**(2-3), 119–146.
- MICHAUD A. (2004b). A Measurement from Electroglottography : DECPA, and its Application in Prosody. In B. BEL & I. MARLIEN, Éd.s., *Speech Prosody 2004*, p. 633–636, Nara, Japan.
- MICHAUD A., GUILLAUME S., JACQUES G., MAC Đ.-K., JACOBSON M., PHAM T.-H. & DEO M. (2016). Contribuer au progrès solidaire des recherches et de la documentation : la Collection Pangloss et la Collection AuCo. In *Journées d'Etude de la Parole 2016*, volume 1 de *Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 1 : Journées d'Etude de la Parole*, p. 155–163, Paris, France : Association Francophone de la Communication Parlée. HAL : [halshs-01341631](https://halshs.archives-ouvertes.fr/halshs-01341631).
- MICHAUD A., VAISSIÈRE J. & NGUYÊN M.-C. (2015). Phonetic insights into a simple level-tone system : 'careful' vs. 'impatient' realizations of Naxi High, Mid and Low tones. In *ICPhS XVIII (18th International Congress of Phonetic Sciences)*. HAL : [halshs-01148765](https://halshs.archives-ouvertes.fr/halshs-01148765).
- NGUYEN M.-C. (2021). *Glottalization, tonal contrasts and intonation : an experimental study of the Kim Thuong dialect of Muong*. thèse de doctorat, Université de la Sorbonne nouvelle - Paris III. HAL : [tel-03652510](https://halshs.archives-ouvertes.fr/halshs-03652510).
- ORLIKOFF R. F. (1998). Scrambled EGG : The uses and abuses of electroglottography. *Phonoscope*, **1**(1), 37–53.
- RECASENS D. & MIRA M. (2013). Voicing assimilation in Catalan three-consonant clusters. *Journal of Phonetics*, **41**(3-4), 264–280.

VAISSIÈRE J., HONDA K., AMELOT A., MAEDA S. & CREVIER-BUCHMAN L. (2010). Multisensor platform for speech physiology research in a phonetics laboratory. *Journal of the Phonetic Society of Japan*, **14**(2), 65–77.