

APPRENTISSAGE PROFOND POUR L'ESTIMATION DU QUOTIENT OUVERT À PARTIR DU SIGNAL ÉLECTROGLOTTOGRAPHIQUE

Minh-Châu Nguyễn, Maximin Coavoux, Solange Rossato

Université Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France



MOTIVATIONS

- Déterminer dans quelle mesure l'apprentissage machine peut faciliter le traitement des **enregistrements électroglottographiques (EGG)**
- Mener des expériences qui visent à automatiser l'estimation du **quotient d'ouverture glottique (O_q)** à partir du signal EGG
- Objectif à moyen terme : Explorer quelles autres informations peuvent être extraites du signal EGG

CORPUS

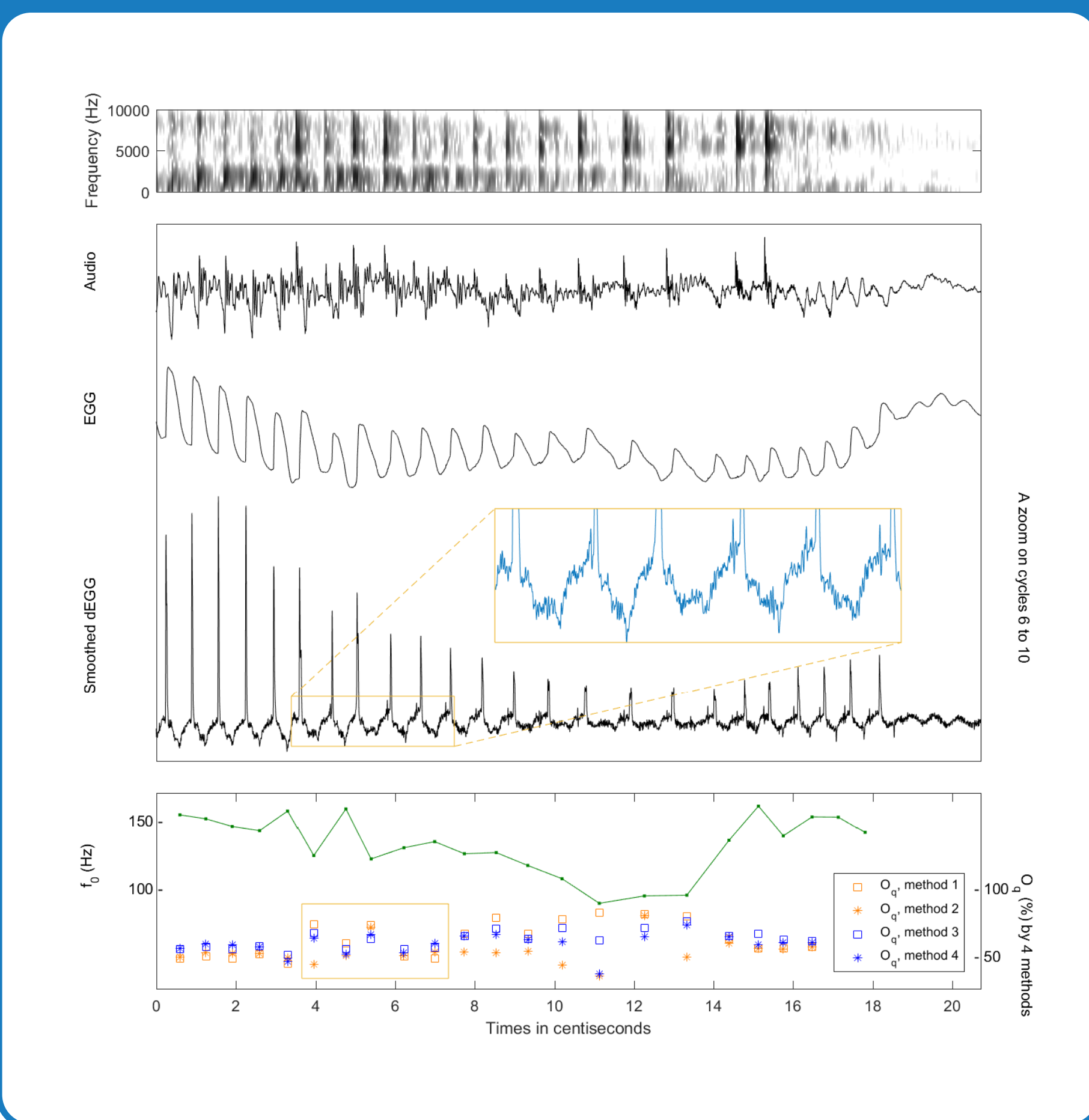
- Langue : un dialecte du **muong** (Vietnam)
- Signaux : signal EGG + signal acoustique
- Jeu de données : 12 ensembles minimaux tonaux (syllabes sans occlusive finale) + 3 paires minimales tonales (syllabes à occlusive finale)
- 20 locuteurs (10 hommes, 10 femmes). Environ 5 heures en tout

DONNÉES NUMÉRIQUES : MESURES DE L' O_q PAR PEAKDET[3]

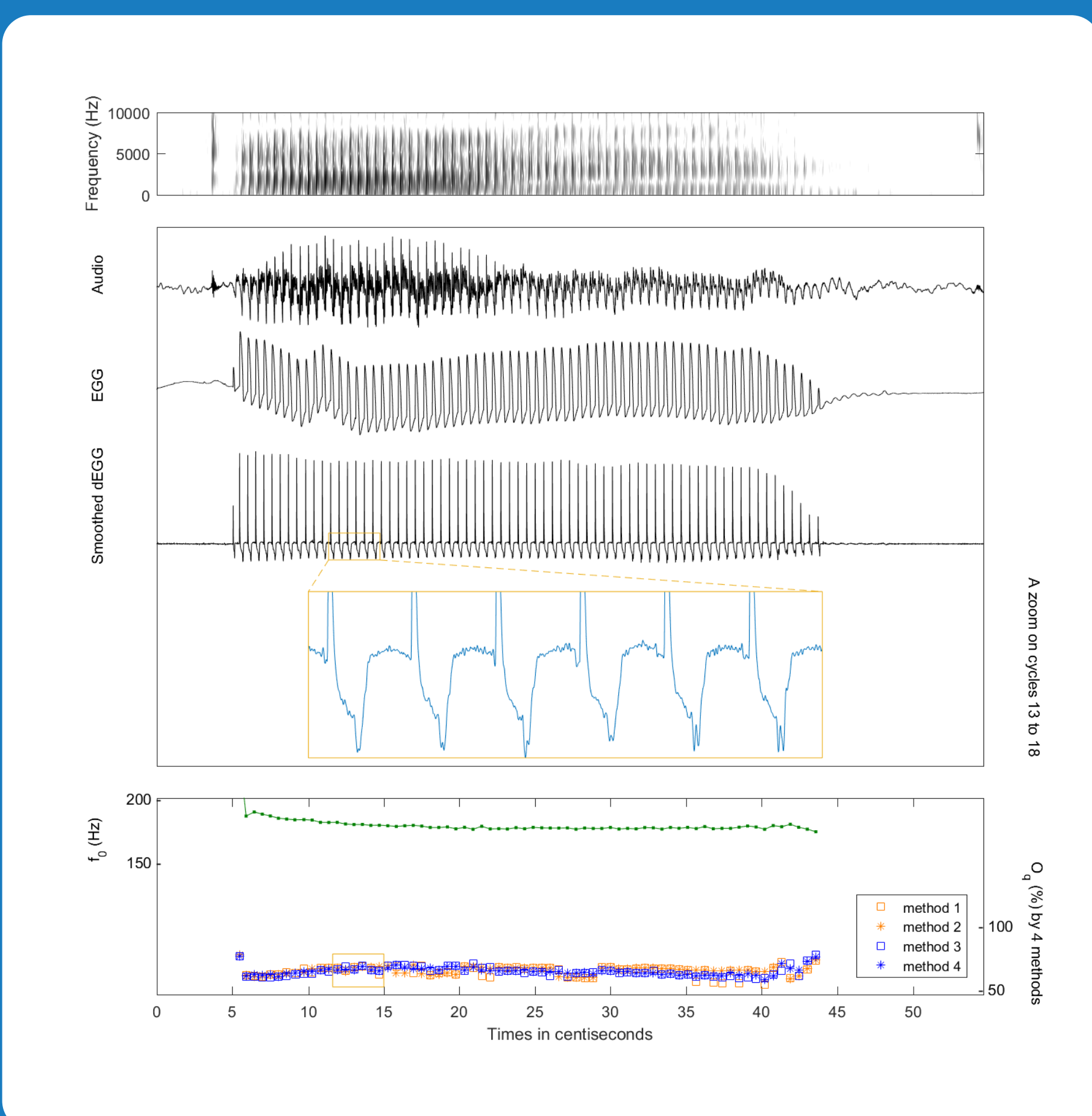
- La méthode pour l'analyse du signal EGG : utilise son signal dérivé (dEGG). [2]
- L'équation de $O_q(\%)$:

$$O_q = (\text{phase ouverte}) / (\text{phase ouverte} + \text{phase fermée})$$
- Le processus de vérification de O_q :

Cas où la méthode 2 (minimum local sur le signal dEGG lissé) a été choisie



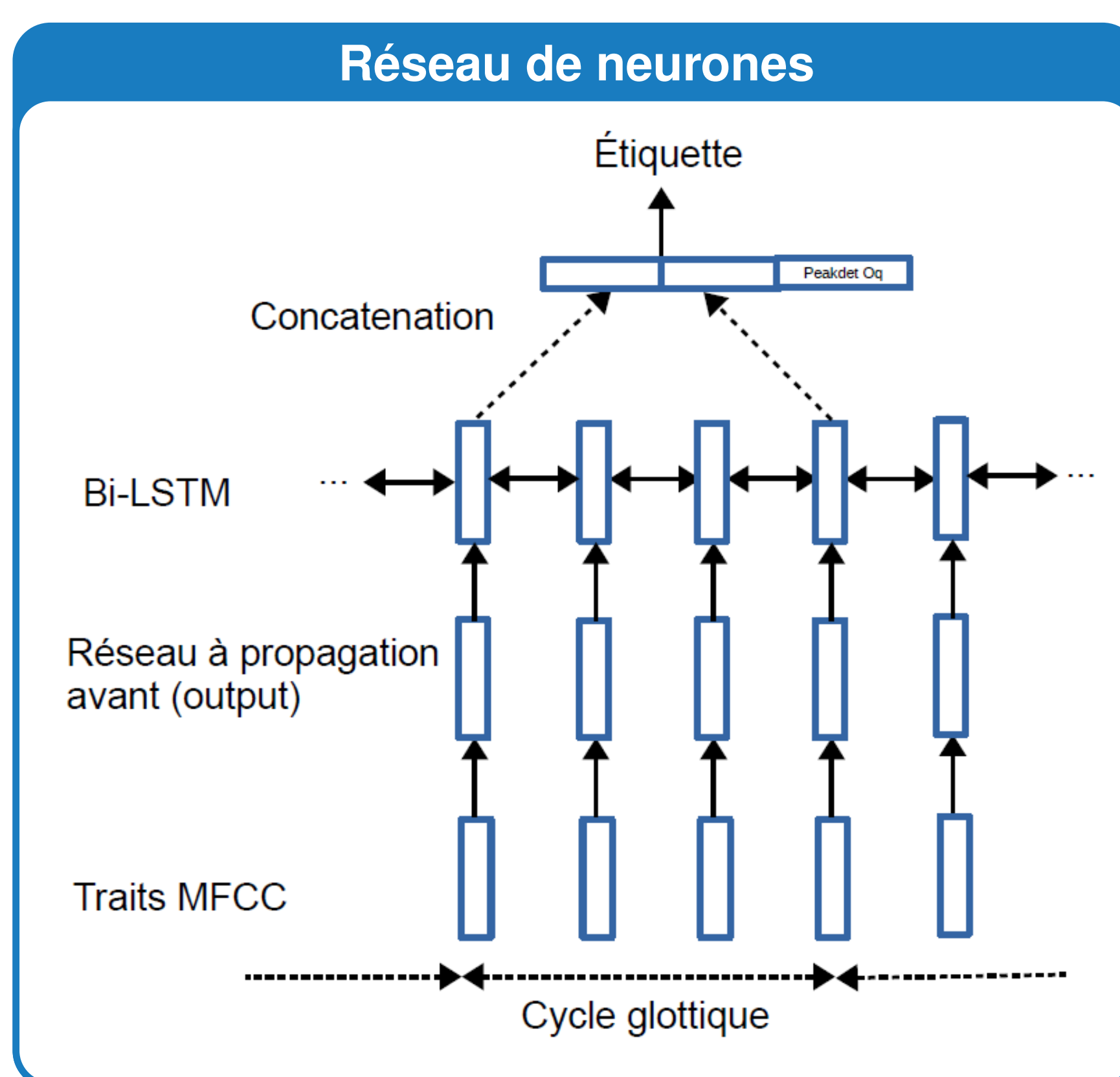
Cas où la méthode 4 (barycentre des pics sur le signal dEGG lissé) a été choisie



Section	Statistiques sur le corpus			
	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%)

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%

EXPÉRIMENTATIONS



TÂCHE:

- Reproduire le travail d'annotation manuelle
- Pour chaque cycle glottique, prédire laquelle des 4 méthodes (1-4) donne la meilleure estimation d' O_q ou si aucun O_q n'est calculable (étiquette 0).

MODÈLE:

- Input:
 - signal EGG
 - time codes pour chaque cycle
 - $F_0 + O_q$ calculés par peakdet
- Réseau illustré ci-contre

CONFIGURATIONS EXPÉRIMENTALES:

- Signal EGG + $O_q + f_0$
- Signal EGG
- $O_q + f_0$
- Signal Audio + $O_q + f_0$
- Signal Audio

RÉSULTAT

Résultats finaux sur les sections de développement et de test (%)

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

- Évaluation à 3 classes (0 vs 1-2 vs 3-4) ou à 2 classes (0 vs 1-2-3-4)
 - Résultats des modèles (modestement) au-dessus de la baseline
 - Modèles (i) et (iv) quasi-identiques: 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.
- ⇒ L'information du signal EGG n'est quasiment pas exploitée par notre modèle (résultat négatif)

Remerciements:

Les travaux présentés ici sont financés dans le cadre du projet "La documentation automatique des langues à l'horizon 2025" (Computational Language Documentation by 2025, ANR-19-CE38-0015-04). Un grand merci à Alexis Michaud pour de nombreuses discussions et sa relecture attentive.

EGG: Électroglottographiques • DEGG: La dérivé du signal électroglottographiques • O_q : Quotient d'ouverture glottale

CONCLUSION ET PERSPECTIVES

Contributions:

- Premiers résultats expérimentaux pour l'estimation du quotient ouvert, en se basant sur l'estimation de la fiabilité de méthodes sans apprentissage (Peakdet).
- Résultats empiriques plutôt négatifs.

⇒ tâche difficile

Perspectives pour la suite du travail:

- Prédiction directe d' O_q à l'aide de modèles de régression.
- Traits extraits de modèles acoustiques préentraînés [1].

References

[1] Alexis Conneau et al. "Unsupervised cross-lingual representation learning for speech recognition". In: *arXiv preprint arXiv:2006.13979* (2020).

[2] Nathalie Henrich et al. "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation". In: *The Journal of the Acoustical Society of America* 115.3 (2004), pp. 1321–1332.

[3] Alexis Michaud. *Peakdet: a script for the semi-automatic analysis of the electroglottographic (EGG) signal*. https://github.com/alexis-michaud/egg/tree/master/peakdet_inter. 2004–2019.