

Détection et classification automatique d'erreurs de prononciation en L2 : approche basée sur les connaissances didactiques

Romain Contrain¹, Julien Pinquier¹, Lionel Fontan², Isabelle Ferrané¹

¹IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France
²Archean Labs, Montauban, France

Contexte

Entraînement à la Prononciation Assisté par Ordinateur

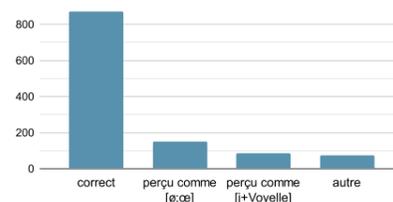
- **Enjeu de l'enseignement des langues** : travailler individuellement la prononciation avec chaque apprenant – Intérêt des outils d'EPAO
- **Création d'une plateforme d'EPAO** : expression orale, feedback et remédiation
- **Focus** : détection d'erreurs de prononciation, diagnostic et remédiation adaptée



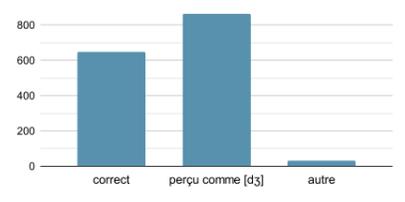
Corpus et Objectifs

- **Corpus** : productions japonais (L1) / français (L2) – Tâche de répétition
 - Enregistrements annotés et segmentés au niveau phonèmes (54 800 phones)
 - Connaissances métier (catégories didactiques fonction de l'impact sur la remédiation) : application aux instances de /y/ et /ɜ/
- **Objectif** : prédire automatiquement les catégories didactiques

Distribution des prononciations de /y/

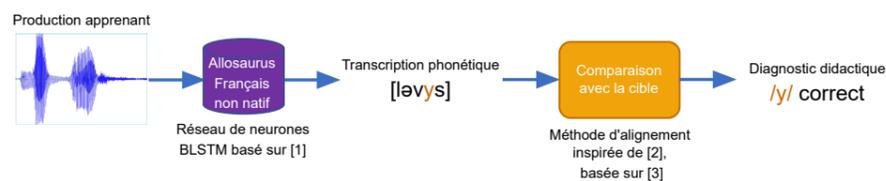


Distribution des prononciations de /ɜ/

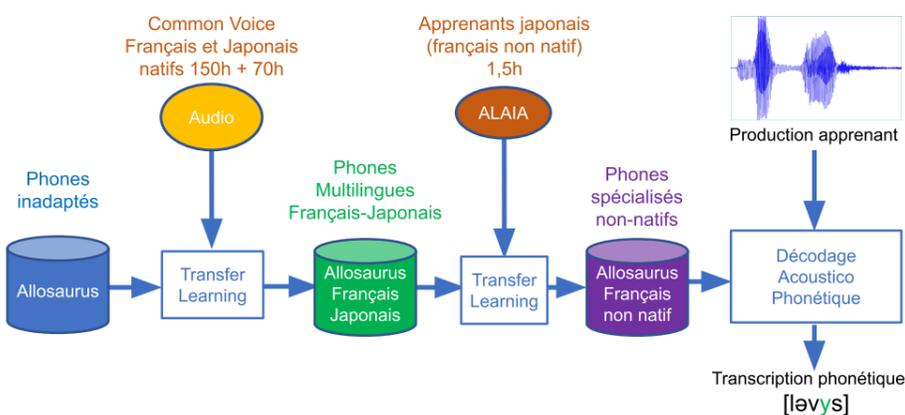


Première approche

Transcription phonétique

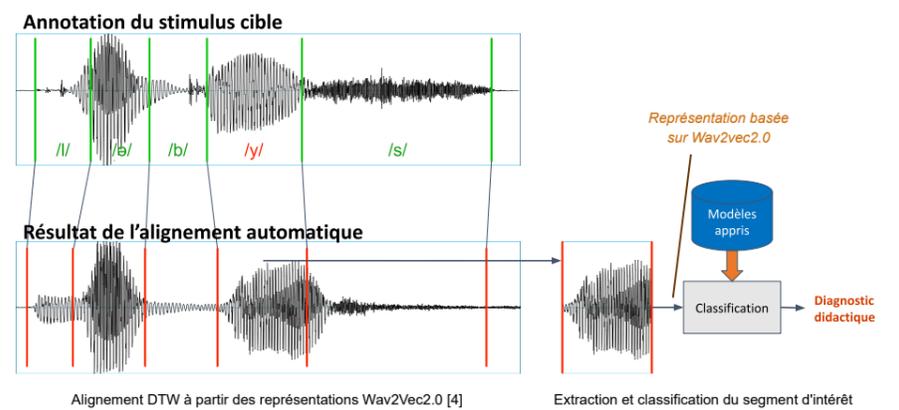


Adaptation du modèle à la parole d'apprenant

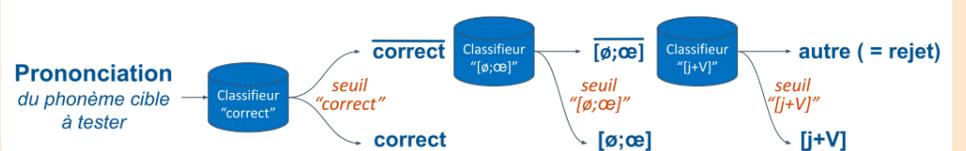


Seconde approche

Segmentation du signal et Alignement



Prise de décision



Résultats

	Précision Approche transcription	Précision Approche alignement
/y/ correct	87,9 %	94,4 %
/y/ perçu comme [ø;œ]	30,2 %	60,9 %
/y/ perçu comme [j+Voyelle]	52,4 %	80,8 %
/ɜ/ correct	83,3 %	90,3 %
/ɜ/ perçu comme [dʒ]	84,4 %	90,5 %

	Rappel Approche transcription	Rappel Approche alignement
/y/ correct	73,5 %	32,6 %
/y/ perçu comme [ø;œ]	49,0 %	9,3 %
/y/ perçu comme [j+Voyelle]	25,3 %	24,1 %
/ɜ/ correct	75,7 %	24,6 %
/ɜ/ perçu comme [dʒ]	87,6 %	31,9 %

Discussion

Performances

- Approche alignement : donne moins souvent une réponse, mais ses diagnostics sont plus souvent justes (préférable pour l'EPAO) [5]
- Précisions obtenues : industriellement insuffisantes (<85%) sur /y/, mais satisfaisantes sur /ɜ/

Perspectives

- Extension à d'autres phonèmes d'intérêt (nouvelles connaissances didactiques à intégrer)

Explicabilité

- Approche transcription : utilisation d'une architecture récurrente profonde qui gêne l'explication
- Approche alignement : basée sur une représentation Wav2vec2 [4] difficile à expliquer, car tous les aspects de la parole se retrouvent potentiellement entremêlés

Interprétabilité

- Résultats interprétables de par la connaissance des catégories didactiques
 - Remédiation adaptée à l'apprenant et au type d'erreur réalisée

Références

[1] X. Li, S. Dalmia, J. Li, M. Lee, P. Littell, J. Yao, A. Anastasopoulos, D. R. Mortensen, G. Neubig, A. W. Black et al., "Universal phone recognition with a multilingual allophone system," in *ICASSP 2020 IEEE*, pp. 8249–8253.
 [2] A. Ghio, M. Lalain, L. Giusti, G. Pouchoulin, D. Robert, M. Rebourg, C. Fredouille, I. Laaridh, and V. Woisard, "Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique," in *XXXIe Journées d'Etudes sur la Parole*. ISCA, 2019.
 [3] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the aminoacid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443–453, 1970.

[4] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0 : A framework for self-supervised learning of speech representations," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 12 449–12 460.

[5] S. M. Witt, "Automatic error detection in pronunciation training: Where we are and where we need to go," in *International Symposium on automatic detection on errors in pronunciation training*, 2012, pp. 1–8.