Proximité rythmique entre apprenants et natifs du français Évaluation d'une métrique basée sur le CEFC

Sylvain Coulange, Solange Rossato

Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

Proximité rythmique entre apprenants et natifs du français Évaluation d'une métrique basée sur le CEFC

Sylvain Coulange[1] Solange Rossato[2]
(1) LIDILEM, Université Grenoble Alpes, France
(2) LIG, Université Grenoble Alpes, France
{sylvain.coulange, solange.rossato}@univ-grenoble-alpes.fr

RÉSUMÉ

Cette étude a pour objectif de proposer une quantification de l'accent étranger se basant sur des mesures rythmiques. Nous avons utilisé le Corpus pour l'Étude du Français Contemporain, qui propose plus de 300 heures de parole aux profils de locuteurs et aux situations variés. Nous nous sommes concentrés sur 16 paramètres temporels estimés à partir des durées de voisement et de syllabes. Un mélange gaussien a été appris sur les données de 1 340 natifs du français, puis testé sur des extraits de 146 natifs tirés au hasard (NS), sur ceux des 37 non-natifs présents dans le corpus (NNS), ainsi que sur des enregistrements de 29 apprenants japonais de niveau A2 d'un autre corpus. La probabilité que les NNS aient une log-vraisemblance inférieure aux NS ne dépasse pas la tendance ($p = 0,067$), mais celle pour les apprenants japonais est beaucoup plus significative ($p < 0,0001$). L'étude de la répartition des paramètres entre les différents groupes met en avant l'importance du débit de parole et des durées de voisement. [1]

ABSTRACT

Rhythmic Proximity Between Natives And Learners Of French – Evaluation of a metric based on the CEFC corpus

This work aims to quantify foreign accent in French based on rhythmic measurements. We used the Corpus pour l'Étude du Français Contemporain, which contains +300 hours of speech from a wide variety of speaker profiles and situations. We focused on 16 temporal parameters computed from voicing and syllables intervals. A Gaussian mixture model was trained on 1,340 native speakers of French, then tested with 146 natives (NS) and 37 non-native speakers (NNS) from the same corpus, as well as 29 A2-level Japanese learners of French from another corpus. The probability of NNS having inferior log-likelihood to NS was only a tendency ($p = .067$), but a much bigger probability was obtained for Japanese learners ($p < .0001$), where all speakers were A2 level. Parameter efficiency analysis reveals the importance of speech rate and voicing duration. [2]

MOTS-CLÉS : Modélisation du rythme, Accent étranger, Prononciation de la L2, Débit de parole, séquences voisées et non voisées.

KEYWORDS: Rhythm modelling, Foreign accent, L2 pronunciation, Speech rate, Voiced and Unvoiced sequences.

1. Une version anglaise de cet article est également parue dans les actes de la conférence internationale LREC 2020.
2. An English version of this article was also published in the proceedings of LREC 2020.

Volume 1 : Journées d'Études sur la Parole, pages 118–126.
HAL : hal-02798525.

1 Introduction

La perception de l'accent étranger est principalement due aux différences de prononciation entre ce que le locuteur dit et une norme attendue et partagée par les natifs de la langue cible (Alazard, 2013). Cette différence a été largement décrite au niveau segmental, à travers les théories de l'acquisition du langage, et le rôle de la prosodie a été amplement démontré, notamment par des études de perception (De Meo *et al.*, 2012; Pellegrino, 2012). Aujourd'hui, il est reconnu que le segmental et le suprasegmental jouent tous les deux un rôle important dans la perception de cette différence par les locuteurs natifs. Parmi les paramètres prosodiques, nous nous intéressons à des paramètres rythmiques, le rythme étant défini ici comme une récurrence de patterns de marquages forts ou faibles d'éléments dans un environnement temporel (Gibbon & Gut, 2001). Ces éléments peuvent être des alternances de syllabes longues et courtes, ou de segments vocaliques et consonantiques. Beaucoup d'études tentent d'ailleurs de classifier les langues en se basant sur les durées vocaliques et consonantiques, et nécessitent donc une transcription alignée. Des études ont toutefois montré des résultats similaires à partir des durées de voisement ou des durées de syllabes (Fourcin & Dellwo, 2013; Dellwo *et al.*, 2015), paramètres qui peuvent être quant à eux détectés automatiquement directement à partir du signal audio, sans recours à une transcription. C'est ce qui nous intéresse étant donné que la transcription automatique de la parole non-native reste encore difficile.

Les catégories rythmiques des langues sont cependant peu distinctes et la plupart des études font part de limites, dues à de trop petits effectifs de locuteurs, ou à des biais d'élicitation, facteurs qui influent aussi sur les caractéristiques rythmiques (Fourcin & Dellwo, 2013; Ramus *et al.*, 1999; Gibbon & Gut, 2001; Grabe & Low, 2002). Il est donc nécessaire d'utiliser un corpus volumineux, prenant en compte la variation selon les situations et les locuteurs, et notamment lors de production spontanée (Bhat *et al.*, 2010). Cela inclue également les enregistrements provenant de différents pays francophones, et de milieux sociaux variés.

Dans cette étude, nous proposons de modéliser le rythme du français à travers le récent Corpus d'Étude pour le Français Contemporain (CEFC, Benzitoun *et al.* 2016). Afin de modéliser l'ensemble de ces variations, nous avons entraîné un modèle de mélange gaussien sur 16 paramètres rythmiques. La log-vraisemblance obtenue avec ce modèle global pour un nouvel extrait de parole est comparée pour des locuteurs natifs et locuteurs non-natifs que nous avons réservés pour le test, ainsi qu'avec un corpus indépendant d'apprenants japonais du français. Nous avons également étudié la distribution de chaque paramètre entre les extraits de parole des locuteurs natifs et non-natifs.

2 Méthodologie

2.1 Corpus

Le corpus CEFC regroupe, uniformise et complète les annotations d'un ensemble ou de partie de 13 corpus, tels que Valibel [3] ou encore Clapi [4]. Les enregistrements peuvent provenir de différentes régions de France, de Belgique ou de Suisse. Les situations d'énonciation varient de conversations entre amis aux réunions professionnelles, en passant par des repas de famille, des débats médiatiques, des lectures de contes traditionnels ou encore des conversations enregistrées dans des magasins. Les

3. Valibel : https://uclouvain.be/fr/instituts-recherche/ilc/valibel/corpora.html.
4. Clapi : http://clapi.icar.cnrs.fr

extraits de parole sont constitués de dialogues (481 enregistrements sur 900), d'interactions à plus de 2 locuteurs (277) et des monologues (144). Ce corpus totalise environ 4 million de mots, avec 300 heures d'enregistrement. Tous les enregistrements sont transcrits et alignés, ce qui nous a permis d'identifier les extraits de parole de chaque locuteur. Sur un total de 2 587 locuteurs, les femmes représentent la majorité des locuteurs avec 1 373 locutrices, contre seulement 1 048 locuteurs, et 166 dont le genre n'est pas renseigné. [5]

Le corpus contient également la parole d'une cinquantaine de locuteurs non-natifs : ils ont été réservés pour la partition de test de notre modèle et exclus en totalité des enregistrements utilisés pour construire le modèle. En effet, les seules contraintes pour que l'extrait de parole soit inclus dans les données d'apprentissage du modèle est que la langue parlée soit le français, et que le locuteur soit natif. Le corpus de test est constitué d'environ 10% de l'ensemble des enregistrements de locuteurs natifs, choisis au hasard et mis de côté comme base de comparaison avec les locuteurs non-natifs.

Le niveau de français des locuteurs non-natifs n'étant pas précisé dans les métadonnées, nous avons inclus un second corpus de parole de locuteurs non-natifs, pour lequel nous connaissons la langue maternelle et le niveau de compétence en français de chaque locuteur. Ce corpus est constitué des enregistrements d'une évaluation en production orale pour 29 étudiants de l'Université de Langues Étrangères de Kyōto, de la même classe et tous de langue maternelle japonaise. Nous disposons également de leurs résultats à l'examen de fin de semestre, évaluant les 4 habiletés langagières : compréhension de l'oral et de l'écrit, production orale et écrite, et dont les enregistrements constituent une partie de l'évaluation.

2.2 Les paramètres acoustiques

Nous avons mesuré 16 paramètres largement utilisés soit pour la classification des langues (White & Mattys, 2007; Fourcin & Dellwo, 2013; Pettorino *et al.*, 2013, entre autres), la caractérisation des locuteurs (Rossato *et al.*, 2018), ainsi que la perception de l'accent étranger (Bhat *et al.*, 2010; Fontan *et al.*, 2018). Les paramètres rythmiques sont basés sur les segments voisés détectés par Praat [6], et les noyaux syllabiques détectés grâce à un script de De Jong & Wempe (2009) :

— Le débit de parole, ratio entre le nombre de noyaux syllabiques et la durée du segment SR ;
— Le pourcentage de voisement $\%V$;
— La moyenne μV, l'écart type σV et le coefficient de variation ρV des durées des intervalles voisés V_i ;
— La moyenne μU, l'écart type σU et le coefficient de variation ρU des durées des intervalles non-voisés U_i ;
— La moyenne μP, l'écart type σP et le coefficient de variation ρP des durées de la paire $P_i = V_i + U_i$;
— L'indice de comparaison brut $rPVI$ et normalisé $nPVI$ de couples successifs d'intervalles voisés (Grabe & Low, 2002; Fourcin & Dellwo, 2013) ;
— La moyenne $\mu\Delta NV$, l'écart type $\sigma\Delta NV$ et le coefficient de variation $\rho\Delta NV$ des durées ΔNV_i entre deux noyaux syllabiques successifs.

Les paramètres rythmiques sont calculés sur des segments constitués par la concaténation d'unités

5. Des statistiques plus détaillées sur les locuteurs du CEFC sont présentées dans Coulange (2019), mémoire de master de Sciences du langage parcours industries de la langue, Université Grenoble Alpes dirigé par Solange Rossato.
6. Praat : doing phonetics by computer. Version 6.0.37, téléchargée en mars 2019 depuis http://www.praat.org/.

entre pauses (UEP) [7] consécutives d'un même locuteur, jusqu'à atteindre une durée minimale de 30 secondes. Aucune UEP n'est coupée avant sa fin, il arrive donc que certains segments soient assez longs. Nous pensons que cette durée permet d'avoir suffisamment de parole pour obtenir des mesures fiables. Nous ne gardons que les locuteurs ayant au moins un segment et pour lesquels le statut de la langue française est connu (natif ou non). Le corpus d'apprentissage contient 16 884 segments de 1 340 locuteurs natifs. Les trois partitions de test sont constituées comme suit : 146 locuteurs natifs NS, 37 locuteurs non-natifs NNS et le corpus de 29 apprenants japonais JpNNS. Le tableau 1 récapitule ce partitionnement.

Set	Training	Test NS	Test NNS	Test JpNNS
French status	native	native	non-native	non-native
Corpus	CEFC	CEFC	CEFC	Jp learners
#speakers	1,340	146	37	29
#segments	16,884	1,919	268	96

TABLE 1 – Constitution du corpus d'apprentissage et des 3 corpus de test

2.3 Modèle de mélanges gaussiens

Selon Ferrer *et al.* (2015), les modèles de mélanges gaussiens (GMM) sont réputés pour modéliser la variation. Un GMM est une densité de probabilité calculée à partir d'une somme de gaussiennes pondérées. Cette fonction suit au mieux la distribution des données. L'apprentissage du modèle revient à trouver les meilleurs paramètres de ces gaussiennes (moyennes et matrices de covariance) et leur pondération pour représenter les données grâce à l'algorithme d'espérance-maximisation EM. La probabilité d'un vecteur $\vec{x}$, étant donné un GMM de paramètres $\{w_k, \vec{\mu}_k, \Sigma_k\}_{k=1}^{K}$ est alors :

$$p(\vec{x}) = \sum_{k=1}^{K} w_i \mathcal{N}(\vec{x}|\vec{\mu}_k, \Sigma_k) \tag{1}$$

où K est le nombre de gaussiennes, w_k est le poids de la gaussienne k, tel que $\sum_{k=1}^{K} w_k = 1$, et $\mathcal{N}(\vec{x}|\vec{\mu}_k, \Sigma_k)$ la fonction normale de $\vec{x}$ de moyenne $\vec{\mu}$ et de covariance Σ de k. Nous utilisons la covariance diagonale pour alléger l'apprentissage, même si certains paramètres acoustiques sont corrélés entre eux. Nous avons également choisi de limiter notre GMM à 1 024 gaussiennes. L'apprentissage du modèle a été implémenté en Python grâce à la librairie *SciKit Learn* Gaussian Mixture [8].

Pour calculer la proximité de la parole d'un locuteur X au modèle, nous avons utilisé le produit de la vraisemblance du modèle pour chacun de ses segments de parole $\vec{x_n}$. Pour simplifier ces calculs, nous avons transformé le produit en somme en utilisant la log-vraisemblance $log\ p(X)$, et normalisé celle-ci par le nombre N de segments de chaque locuteur qui peut varier beaucoup en fonction des locuteurs :

$$log\ p(X) = \frac{1}{N} \sum_{n=1}^{N} log\ p(\vec{x}_n) \tag{2}$$

7. Toute pause supérieure à 1 seconde ou tout changement de locuteur mettant fin à une UEP.
8. https ://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html

Nous avons calculé la log-vraisemblance moyenne de chaque locuteur natif de la partition de test (NS), et l'avons comparée à celle des locuteurs des corpus non-natifs (NNS ou JpNNS) avec un test de Wilcoxon-Mann-Whitney. Pour les locuteurs non-natifs, la log-vraisemblance obtenue par le modèle gaussien est interprétée comme un score de proximité rythmique au français et elle est comparée aux performances de l'apprenant évaluées par l'enseignant de français pour les locuteurs du Test JpNNS.

2.4 Comparaison des paramètres rythmiques entre natifs et non-natifs

Nous comparons la distribution de chacun des 16 paramètres entre les locuteurs natifs et les locuteurs non-natifs du test NNS ou ceux du test JpNNS. Or, le nombre de segments sur lesquels sont extraits ces paramètres est bien supérieur parmi les locuteurs natifs (1 919) par rapport à celui des segments disponibles pour les non-natifs du CEFC (268) et plus encore à celui des japonophones (96). La comparaison est faite sur un échantillon de 96 segments de paroles par corpus, sélectionnés aléatoirement. Pour les natifs, plusieurs tirages aléatoires sont effectués. Les comparaisons entre chaque paramètre mesuré s'effectuent donc sur 96 valeurs pour chaque groupe natif et non-natif. Un test t permet de tester si la différence observée entre les deux groupes est significative. Nous avons également calculé la valeur de l'êta-carré η^2 qui rend compte de la proportion de variance du paramètre expliquée par la variable natif/non-natif.

3 Résultats

3.1 Proximité rythmique des locuteurs

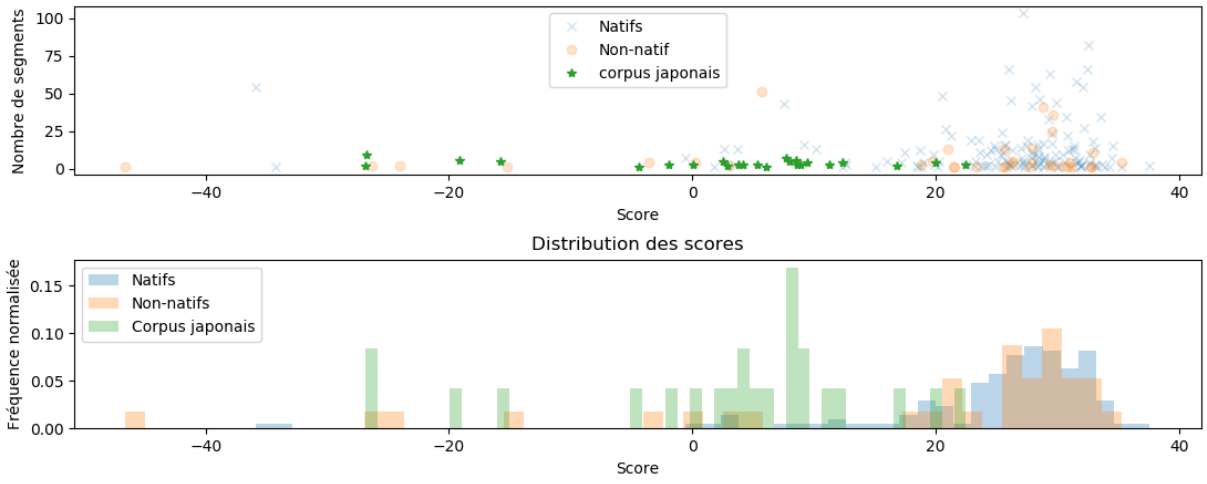


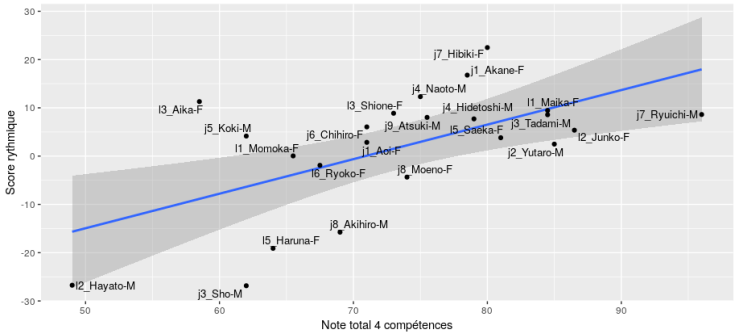
FIGURE 1 – Distribution des scores des locuteurs natifs NS (bleu), des non-natifs NNS (orange) et des apprenants japonais JpNNS (vert) (supérieurs à -50)

Nous avons d'abord comparé les scores de proximité rythmique des locuteurs natifs (NS) avec ceux des non-natifs du CEFC (NNS). Il s'avère que l'hypothèse selon laquelle les locuteurs non-natifs obtiennent un score de proximité rythmique inférieur à celui des natifs (et sont donc plus éloignés du modèle) ne dépasse pas la tendance ($p = 0.067$).

Nous avons ensuite comparé le score des natifs à celui des non-natifs du corpus d'apprenants japonais (JpNNS). Cette fois-ci, la différence est très significative ($p < 0.001$). L'écart entre les scores de proximité des locuteurs natifs et ceux des locuteurs non-natifs est bien plus net.

La figure 1 présente une distribution de ces scores, zoomée sur les scores supérieurs à -50 (ce qui correspond à 96,6% des NS, 94,6% des NNS et 79,3% des JpNNS). Les NS et les NNS sont majoritairement entre 20 et 40, tandis qu'on ne trouve aucun JpNNS avec un score supérieur à 22,48 (la majorité d'entre eux se situe en 0 et 10). La proximité moyenne de chaque population est respectivement de 25,0, 21,48 et 0,74, si on ignore les scores inférieurs à -50 qui pourraient être dus à de mauvaises détections de voisement ou de noyaux syllabiques à cause de voix trop faibles [9]. Dans cette figure, nous avons également fait apparaître le score de l'enseignant, francophone natif de la classe d'apprenants japonais, dont la voix a également été enregistrée. Son score est de 19,96.

3.2 Corrélation avec le niveau de compétence en langue



(a) Scores de proximité rythmique des apprenants en fonction de leur note globale à l'examen

	Global	Oral	Fluency
r	.598 (p = .003)	.257 (p = .237)	.410 (p = .052)
r^2	.358 (p = .003)	.066 (p = .237)	.168 (p = .052)
ρ	.478 (p = .021)	.315 (p = .144)	.228 (p = .295)

(b) Tests de corrélation entre les scores de proximité rythmique et les notes globales (gauche), de production orale (milieu) et de fluence (droite)

FIGURE 2 – Corrélation entre scores de proximité rythmique et niveau de français

Avec les enregistrements du corpus d'apprenants japonais, nous disposons également de 3 notes pour chaque étudiant : la note globale obtenue à l'examen de fin de semestre, type DELF, évaluant les 4 habiletés langagières (compréhension et production, orale et écrite), la note obtenue en production orale (PO) pour ce même examen, et le nombre de points obtenus spécifiquement pour la fluence du discours. La fluence est évaluée sur 5 points dans la partie de production orale, qui représente elle-même $\frac{1}{4}$ de la note globale sur 100 points. Les enregistrements du corpus sont ceux de la production orale des étudiants lors de l'examen, sur laquelle sont les notes de production orale et de fluence.

Nous avons mesuré la corrélation entre ces 3 notes et le score de proximité rythmique obtenu par chaque étudiant. Le tableau 2b donne le coefficient de corrélation de Spearman (r), le coefficient de détermination (r^2) et le coefficient de Pearson (ρ) pour chaque type de note : globale, production orale, et fluence ; avec les p-values associées. La note globale est assez bien corrélée avec le score de proximité rythmique ($r = .598$, $p < .005$; $r^2 = .358$, $p < .005$ et $\rho = .478$, $p < .05$). Les notes de production orale et de fluence sont quant à elles trop proches les unes des autres (17 à 24 pour la PO et 3 à 5 pour la fluence), et la corrélation n'est pas significative. La figure 2a montre les scores de proximité rythmique des 29 étudiants en fonction de leur note globale lors de l'évaluation.

9. Ces scores vont de -57,83 à -12 544,14 ; deux seulement sont compris entre -50 et -100.

3.3 Analyse des paramètres rythmiques

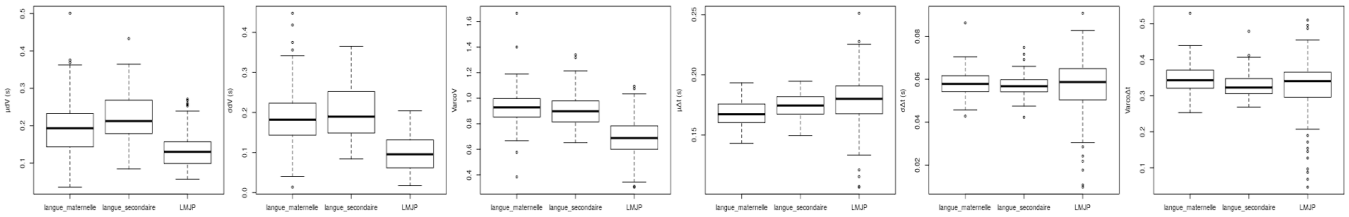


FIGURE 3 – Distribution de (μV), (σV), ρV des durées d'intervalles voisés ; et de $(\mu \Delta NV)$, $(\sigma \Delta NV)$ et $\rho \Delta NV$ des écarts entre noyaux syllabiques (respectivement NS, NNS et JpNNS).

La figure 3 illustre les distributions de certains paramètres en fonction des groupes NS, NNS et JpNNS. Pour certains paramètres, comme le débit de parole SR, on observe un comportement proche entre le groupe NS, et le groupe NNS avec respectivement 4,0 syll/s et 3,9 syll/s mais seulement 1,3 syll/s pour les JpNNS. De même, le pourcentage de voisement $\%V$ est de 57% pour les natifs, 58% pour les NNS et seulement 15% pour les JpNNS.

La durée moyenne des intervalles voisés μV est plus longue pour les locuteurs NS (190 ms) et pour les NNS (210 ms) que pour les JpNNS (130 ms). Les écarts de durée de ces intervalles sont toutefois plus importants chez les NS et les NNS que chez les apprenants japonais (σV à respectivement 180, 190 et 90 ms). ρV permet une mesure indépendante du débit de parole, il varie de 0,93 chez les natifs et 0,90 chez les NNS à 0,70 chez les JpNNS.

En ce qui concerne les métriques faisant intervenir les durées d'intervalles non-voisés, on constate des valeurs de μU d'environ 150 ms pour les NS et NNS, et de 580 ms pour JpNNS. σU reflète un même ordre de grandeur pour les NS et NNS (190 ms) et des variations très importantes pour les JpNNS (730 ms). ρU montre peu de différence entre les groupes : 1,21 pour les NS, 1,27 pour les NNS et 1,22 pour les JpNNS.

Paramètre	η^2	P-value		$\mu(\eta^2)$	$\sigma(\eta^2)$
SR	.745	$3.07e^{-58}$	<.0001	.705	.038
$\%V$	.673	$5.63e^{-48}$	<.0001	.647	.039
ρV	.415	$7.23e^{-24}$	<.0001	.391	.028
σV	.373	$4.79e^{-21}$	<.0001	.359	.029
$nPVI$	.360	$3.81e^{-20}$	<.0001	.339	.030
$rPVI$	.349	$1.99e^{-19}$	<.0001	.327	.035
μV	.281	$2.64e^{-15}$	<.0001	.246	.033
μU	.125	$4.78e^{-07}$	<.0001	.123	.003
μP	.118	$1.11e^{-06}$	<.0001	.116	.002
σU	.085	$3.94e^{-05}$	<.0001	.084	.002
$\mu \Delta NV$	.084	$4.37e^{-05}$	<.0001	.066	.017
σP	.082	$5.46e^{-05}$	<.0001	.081	.001
$\rho \Delta NV$	.046	.003	<.01	.029	.015
ρU	.012	.124	>.05	.012	.010
$\sigma \Delta NV$	.011	.157	>.05	.006	.004
ρP	.008	.206	>.05	.008	.008

FIGURE 4 – η^2 et p-value entre les NS et les JpNNS pour le 1er rééchantillonnage, et la moyenne et l'écart type des η^2 sur les 3 rééchantillonnages

Les indices de comparaisons $nPVI$ et $rPVI$ de durée entre couples successifs d'intervalles voisés présentent des valeurs très proches pour NS et NNS, et plus basses pour JpNNS (NS = 78,54% et NNS = 79,14% ; JpNNS = 64,57%), indiquant que les locuteurs JpNNS ont tendance a avoir moins d'écart de durée entre deux intervalles voisés successifs, ce qui est cohérent avec une valeur de σV plus faible. Lorsque l'on s'intéresse à l'écart temporel entre deux noyaux syllabique ΔNV, la valeur moyenne des durées séparant les noyaux syllabiques varie peu entre les groupes, avec 166 ms pour NS, 176 ms pour NNS et 180 ms pour JpNNS. Les écart-types et les coefficients de variations varient

peu entre les groupes mais montrent une dispersion plus importante pour les JpNNS (cf. les 2 derniers graphiques de la figure 3).

Nous avons donc voulu comparer NS et JpNNS afin de déterminer si, pour chaque paramètre, les différences observées étaient significatives ou non et calculé les êta-carrés. La figure 4 présente les êta-carrés (η^2) du premier rééchantillonnage, avec les p-values associées, ainsi que la moyenne des η^2 sur 3 rééchantillonnages successifs et leur écart type. On remarque que la variance du débit de parole SR est expliquée à 75% par le facteur natif/non-natif, suivi de près par le pourcentage de voisement ($\%V$, 67%). Arrivent ensuite les métriques impliquant les durées d'intervalles voisés : le coefficient de variation des durées d'intervalles voisés (ρV), l'écart type et la moyenne de durée de ces intervalles (σV, μV), comme les indices de comparaison brut et normalisé de leur paires successives ($nPVI$, $rPVI$). La variance de ces paramètres est expliquée pour 28 à 42% par le facteur natif/non-natif.

4 Discussion

Nous avons proposé une modélisation informatique du rythme du français, apprise sur un corpus volumineux de parole variée, et basée sur les mesures de 16 paramètres rythmiques détectés de manière entièrement automatique. Ce modèle a permis de mesurer un score de proximité rythmique à partir d'extraits de parole de minimum 30s. La comparaison des scores de proximité rythmique entre des locuteurs natifs et des locuteurs non-natifs du CEFC n'a pas montré de différence significative. Plusieurs raisons peuvent expliquer ce phénomène, comme l'hétérogénéité probable des niveaux de français chez les non-natifs du CEFC, la diversité de leurs langues maternelles ou encore les conditions d'enregistrement. Nous savons que les 37 locuteurs viennent d'au moins 18 pays différents, or le rythme de la langue maternelle (ou des langues maternelles) influence grandement l'acquisition des autres langues, tout comme d'autres facteurs individuels tels que la durée de séjour dans un pays où la langue cible est dominante, ou encore l'âge de première exposition à la langue (Piske *et al.*, 2001; Flege, 1988).

La comparaison d'un groupe plus homogène d'apprenants japonophones de niveau A2 a permis de montrer des différences significatives entre les scores de proximité rythmiques des apprenants et ceux des locuteurs natifs, avec une corrélation, certes peu élevée, avec le niveau global de français des étudiants japonais. Les notes de production orale et de fluence des étudiants ne nous ont pas permis de tirer de conclusion. Il nous faudrait réitérer l'expérience avec des notes plus détaillées, et plus hétérogènes entre les étudiants. Il serait également intéressant de mesurer la corrélation entre les scores rythmiques et les résultats d'un test de perception de l'accent étranger sur ces mêmes locuteurs.

Une étude plus fouillée des 16 paramètres confirme que le débit de parole est bien plus faible chez les apprenants, avec une diminution de la proportion de voisement, sans doute due à une augmentation des pauses intra UEP, inférieures à 1s. L'écart-type des intervalles voisés σV et sa version normalisée ρV ainsi que l'indice de comparaison des durées entre couples d'intervalles voisés successifs $rPVI$ et sa version normalisée $nPVI$ ne dépendent pas du débit pour leur version normalisée et montrent une tendance à uniformiser la durée des intervalles voisées chez les apprenants par rapport aux locuteurs natifs. Ces résultats corroborent les connaissances des enseignants qui savent que la parole des apprenants doit être accélérée, avec moins de pauses et plus de variation de durées de voisement pour se rapprocher de la parole native. Il serait intéressant maintenant de réitérer l'expérience avec différents niveaux de compétence en langue, et des langues maternelles appartenant à différentes familles rythmiques, pour voir comment varient les paramètres rythmiques en fonction de ces facteurs.

Références

ALAZARD C. (2013). *Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Étrangère*. Thèse de doctorat, Université Toulouse 2. Thèse de doctorat dirigée par Michel Billières et Corine Astesano.

BENZITOUN C., DEBAISIEUX J.-M. & DEULOFEU H.-J. (2016). Le projet ORFÉO : un corpus d'études pour le français contemporain. *Corpus*, **15**, 91–114.

BHAT S., HASEGAWA-JOHNSON M. & SPROAT R. (2010). Automatic fluency assessment by signal-level measurement of spontaneous speech. *Second Language Studies : Acquisition, Learning, Education and Technology*.

DE JONG N. & WEMPE T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, **41**(2), 385–390.

DE MEO A., PETTORINO M. & VITALE M. (2012). Comunicare in una lingua seconda. Il ruolo dell'intonazione nella percezione dell'interlingua di apprendenti cinesi di italiano. In *La voce nelle applicazioni. Proceedings of the 7th Congress of Italian Association of Speech Sciences AISV*, p. 117–129.

DELLWO V., LEEMAN A. & KOLLY M.-J. (2015). Rhytmic variability between speakers : Articulatory, prosodic and linguistic factors. *The Journal of the Acoustical Society of America*, **137**(3).

FERRER L., BRATT H., RICHEY C., FRANCO H., ABRASH V. & PRECODA K. (2015). Classification of lexical stress using spectral and prosodic features for computer-assisted language learning systems. *Speech Communication*, **69**(C), 31–45.

FLEGE J. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, **84**(1), 70–79.

FONTAN L., LE COZ M. & DETEY S. (2018). Automatically measuring l2 speech fluency without the need of ASR : A proof-of-concept study with Japanese learners of French. In *Interspeech 2018*, p. 2544–2548.

FOURCIN A. & DELLWO V. (2013). Rhythmic classification of languages based on voice timing. *Tranel Review*, p. 87–107.

GIBBON D. & GUT U. (2001). Measuring speech rhythm. In *EUROSPEECH 2001*, p. 95–98.

GRABE E. & LOW E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, **Vol. 7**, 515–546.

PELLEGRINO E. (2012). The perception of foreign accented speech. Segmental and suprasegmental features affecting degree of foreign accent in italian l2. *Mello H. et al. (Eds.) Proceeding of the 8 GSCP Conference*, p. 261–267.

PETTORINO M., MAFFIA M., PELLEGRINO E., VITALE M. & DE MEO A. (2013). *VtoV : a perceptual cue for rhythm identification*. University of Leuven (KU Leuven).

PISKE T., MACKAY I. & FLEGE J. (2001). Factors affecting degree of foreign accent in an l2 : a review. *Journal of Phonetics*, **29**(2), 191 – 215.

RAMUS F., NESPOR M. & MEHLER J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265–292.

ROSSATO S., ZHANG D., AJILI M. & BONASTRE J.-F. (2018). Suivre le rythme de tes paroles. In *Proc. XXXIIe Journées d'Études sur la Parole*, p. 37–45.

WHITE L. & MATTYS S. (2007). Calibrating rhythm : First language and second language studies. *J. Phonetics*, **35**, 501–522.