Reconnaissance automatique de la parole : génération des prononciations non natives pour l'enrichissement du lexique

Ismael Bada, Dominique Fohr, Irina Illina

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

Reconnaissance automatique de la parole : génération des prononciations non natives pour l'enrichissement du lexique

Ismaël Bada, Dominique Fohr, Irina Illina

Université de Lorraine, CNRS, Inria, LORIA
F-54000 Nancy, France
{ismael.bada,dominique.fohr,irina.illina}@loria.fr

RÉSUMÉ

Dans cet article nous proposons une méthode d'adaptation du lexique, destinée à améliorer les systèmes de la reconnaissance automatique de la parole (SRAP) des locuteurs non natifs. En effet, la reconnaissance automatique souffre d'une chute significative de ses performances quand elle est utilisée pour reconnaître la parole des locuteurs non natifs, car les phonèmes de la langue étrangère sont fréquemment mal prononcés par ces locuteurs. Pour prendre en compte ce problème de prononciations erronées, notre approche propose d'intégrer les prononciations non natives dans le lexique et par la suite d'utiliser ce lexique enrichi pour la reconnaissance. Pour réaliser notre approche nous avons besoin d'un petit corpus de parole non native et de sa transcription. Pour générer les prononciations non natives, nous proposons de tenir compte des correspondances graphèmes-phonèmes en vue de générer de manière automatique des règles de création de nouvelles prononciations. Ces nouvelles prononciations seront ajoutées au lexique. Nous présentons une évaluation de notre méthode sur un corpus de locuteurs non natifs français s'exprimant en anglais.

ABSTRACT

In this study we propose a method for lexicon adaptation in order to improve the automatic speech recognition (ASR) of non-native speakers. ASR suffers from a significant drop in performance when used to recognize the speech of non-native speakers, since the phonemes of the foreign language are frequently poorly pronounced by these speakers. To take into account this problem of erroneous pronunciations, we integrate non-native pronunciations in the lexicon and subsequently we use this augmented lexicon for speech recognition of non-natives speakers. To realize our approach we need a small corpus of non-native speech and its transcription. To generate non-native pronunciations, we take into account relationships graphemes-phonemes in the analyzed pronunciations, with a view to automatically generating rules for creating new pronunciations, which will be added to the lexicon. We present an evaluation of our method on a corpus of non-native French speakers, pronouncing sentences in english.

MOTS-CLÉS : reconnaissance automatique de la parole, locuteurs non natifs, lexique
KEYWORDS : automatic speech recognition, non-native speech, lexicon

1 Introduction

Les systèmes de reconnaissance automatique de la parole (SRAP) ont fait des progrès continus au fil des années, grâce à l'utilisation des réseaux de neurones artificiels dans l'élaboration des modèles acoustiques et à la puissance de calcul toujours plus élevée, entraînant une baisse notable du taux d'erreur de reconnaissance. Mais le traitement de la parole non native reste encore un défi à relever.

Volume 1 : Journées d'Études sur la Parole, pages 27–35.
HAL : hal-02798511.

Des études phonologiques (Park et Culnan, 2019) ont mis en évidence les problèmes posés par les locuteurs non natifs : les prononciations des phonèmes de la langue parlée, influencées par celles de la langue maternelle, des intonations différentes, des prononciations hachées, des hésitations et des auto-corrections. Plusieurs approches ont été tentées pour rendre les SRAP tolérants à la parole non native. Elles sont de deux types : l'une utilise l'**augmentation automatique du lexique** avec des prononciations alternatives, et l'autre consiste à **modifier les modèles acoustiques** pour les rendre compatibles avec les phonèmes de la langue parlée. Cela peut être effectué en utilisant un corpus de parole non native (Duan *et al.*, 2017 ; Li *et al.*, 2016) ou seulement la parole native (Lee et Glass, 2015 ; Das et Hasegawa-Johnson, 2015). Une combinaison de ces deux approches est également envisageable (Tan, 2008 ; Goronzy *et al.*, 2004). Une bonne reconnaissance des noms propres dans un contexte multilingue est obtenu en utilisant un modèle acoustique multilingue et de transcriptions nativisées émergeant de G2P (*grapheme-to-phoneme*) de ces langues (Reveil *et al.* 2010).

L'enrichissement du lexique par génération automatique de prononciations alternatives peut se faire *via* un jeu de règles pré-établies de substitution de phonèmes (Schaden, 2004). Ces règles sont issues d'études phonologiques entre les différences de prononciations de phonèmes entre la langue parlée et la langue native. Une autre variante automatisée consiste à utiliser une base de données de parole non native afin de générer une matrice de confusion entre les phonèmes de la langue parlée et les phonèmes de la langue native (Livescu *et al.*, 2000).

L'adaptation des modèles acoustiques en vue de rendre tolérant le SRAP à la parole non native peut consister à utiliser deux ensembles de modèles acoustiques, l'un adapté à la langue maternelle du locuteur, et l'autre à la langue non native. Cela permet de faire une reconnaissance phonétique et un alignement sur la parole non native, dans le but d'extraire les différences de prononciations de phonèmes et de les intégrer dans les modèles acoustiques (Morgan J., 2004). Enfin une variante consiste à intégrer la graphie des mots (Bouselmi G. et al., 2006) dans l'étude des différences de prononciations entre la parole native et celle non native, afin de générer des règles de création de prononciations, règles qui seront ensuite utilisées pour modifier les modèles acoustiques pour les rendre tolérant à une parole non native.

Dans notre article nous nous intéressons à l**'augmentation du lexique.** Notre méthode pour traiter le problème de la parole non native emprunte certaines idées de deux approches : l'enrichissement du lexique canonique par génération de règles, et la prise en compte des correspondances graphèmes-phonèmes dans l'élaboration de ces règles. Nous ne modifions pas le modèle acoustique, c'est donc une approche rapide à mettre en œuvre. Pour réaliser notre approche nous avons besoin d'un petit corpus de parole non native et sa transcription qui nous permet de générer ces nouvelles règles de prononciations. Par rapport aux méthodes de l'état de l'art, notre approche permet d'utiliser simultanément des phonèmes de la langue cible et de la langue maternelle du locuteur. De plus, seul le lexique de l'application est modifié.

2 Méthodologie proposée

Notre approche de prise en compte de la parole non native dans un système de reconnaissance de la parole s'appuie sur l'hypothèse que les locuteurs non natifs **peuvent être influencés par la graphie des mots étrangers à prononcer.** Dans ce cas, les locuteurs non natifs peuvent prononcer certains phonèmes d'un mot en utilisant des règles de phonétisation issues de leur langue maternelle. Par exemple, le graphème « e » du mot anglais *zero* doit se prononcer /ɪ/ (zero /z ɪ r oʊ/). Mais un

français risque de le prononcer /e/ car le graphème « e » ne se prononce pratiquement jamais */i/* en français.

Pour prendre en compte ce phénomène, nous proposons d'enrichir le lexique du système de reconnaissance en y ajoutant les prononciations non natives en tenant compte de la graphie des mots. Nous proposons effectuer seulement les **substitutions** de certains phonèmes natifs par des phonèmes non natifs en fonction de la graphie. Le lexique augmenté sera utilisé pour effectuer la reconnaissance.

Voici en quelques lignes le résumé de notre méthode. Dans une première étape, nous extrayons les correspondances phonème-graphème à partir d'un *lexique phonétique canonique* (lexique de départ du SRAP) : nous effectuons un lien entre chaque phonème d'un mot du lexique et les graphèmes composant ce mot. Puis, à l'aide d'un corpus audio de parole non native, nous identifions les associations entre les phonèmes attendus et les phonèmes réellement prononcés (en utilisant un alignement forcé et une reconnaissance phonétique). A partir de cela, nous créons des règles de substitution des phonèmes en tenant compte des couples graphème-phonème. Ces règles seront appliquées au lexique destiné à la reconnaissance de la parole des locuteurs non natifs, afin de l'enrichir en nouvelles prononciations. La Figure 1 illustre ces étapes. Dans les sections suivantes nous présenterons ces étapes plus en détails et avec des exemples.

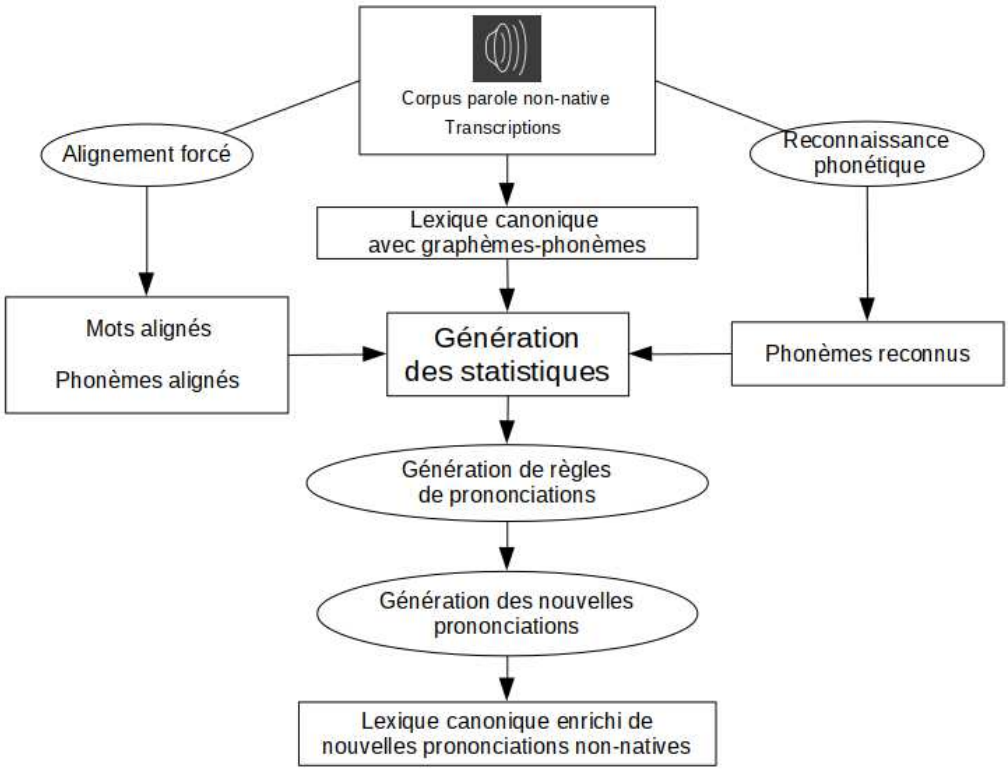

Figure 1*:* Les étapes de la génération des nouvelles prononciations non natives.

2.1 Création d'un lexique avec les associations graphème-phonème

Dans un système de reconnaissance automatique de la parole, **un lexique** phonétisé est une liste de mots avec leurs prononciations (suite de phonèmes). Dans le cadre de notre étude, nous avons également besoin d'un second type de lexique, qui contient ces mêmes mots, mais avec des associations graphème-phonème pour chaque mot. Ce lexique sera généré automatiquement de la manière suivante : en utilisant un algorithme de programmation dynamique, à chaque phonème (ou

groupe de phonèmes) d'un mot du lexique canonique (CMU dictionary), nous affectons le graphème (ou le groupe des graphèmes) qui lui correspond. Voici quelques exemples d'associations graphème-phonème générées :

nogo	n : n	o : əʊ	g : g	o : əʊ
weather	w : w	ea : e	th : ð	er : ɚ
taxi	t : t	a : æ	x : k=s	i : i
zero	z : z	e : ɪ	r : r	o : oʊ

On peut noter que ces associations sont parfois du type « un graphème associé à un phonème » (par exemple e:ɪ) ou parfois du type plusieurs graphèmes associés à un phonème (par exemple ea:e) ou du type « un graphème associé à plusieurs phonèmes » (x:k=s).

2.2 Alignement forcé et reconnaissance phonétique

Pour générer les variantes de prononciations non natives, nous utilisons un corpus audio de parole non native pour lequel nous avons la transcription. Tout d'abord, nous effectuons un alignement forcé de corpus non natif en utilisant le signal audio, les transcriptions et un lexique canonique. Le résultat est un alignement au niveau des phonèmes et au niveau des mots. Ensuite, nous effectuons une reconnaissance phonétique du même corpus, trame par trame (sans lexique ni modèle de langage). Par exemple, pour le mot *nogo* à l'issu de l'alignement forcé nous obtenons :

```
1.530 0.130 n
1.660 0.100 əʊ
1.760 0.070 g
1.830 0.680 əʊ
```

A l'issu de la reconnaissance phonétique trame par trame, pour le segment de parole correspondant à la prononciation du phonème əʊ du mot *nogo*, débutant à la trame 166 et se terminant à la trame 175, nous obtenons les probabilités suivantes :

```
166   əʊ 0.43   n 0.21   ɑ 0.13   u 0.11   ɔ 0.05
167   əʊ 0.88   u 0.04   ɑ 0.04   ɔ 0.02
168   əʊ 0.73   ɔ 0.13   u 0.10   ɑ 0.02
169   əʊ 0.65   ɔ 0.26   u 0.03   ɑ 0.02
170   əʊ 0.62   ɔ 0.27   u 0.03   ɑ 0.02   ʌ 0.01
171   əʊ 0.52   ɔ 0.36   u 0.03   ʌ 0.02   r 0.02   # 0.01
172   əʊ 0.67   ɔ 0.21   r 0.04   l 0.02   ʌ 0.02   ɑ 0.02
173   əʊ 0.78   ɔ 0.10   r 0.04   u 0.02   ʌ 0.01
174   r 0.31   əʊ 0.22   ɔ 0.11   u 0.11   m 0.07   ŋ 0.04
175   r 0.66   əʊ 0.11   g 0.04   m 0.04   ɔ 0.02   ɜː 0.02
```

On peut noter que la somme des probabilités pour chaque trame somme à 1 (les phonèmes qui ont une probabilité inférieure à 0.01 ne sont pas affichés).

2.3 Création de statistiques

Pour créer des statistiques, nous allons utiliser les résultats de l'alignement forcé et de la reconnaissance phonétique. Pour chaque association graphème-phonème, pour chaque phonème reconnu, et pour chaque trame correspondant à ce phonème, nous cumulons les probabilités obtenues. De plus, nous calculons le nombre de trames affecté à chaque association (graphème - phonème, phonème reconnu). Ainsi, nous obtenons des statistiques où pour chaque couple

graphème-phonème, nous avons la distribution en pourcentage des phonèmes reconnus, et leur nombre d'apparitions. Nous appelons ces paramètres α et β. Par exemple, pour le graphème-phonème *e:*ɪ nous avons obtenu :

Association	Phonèmes reconnus			
e: ɪ	ɪ *(40%) (3200)*	e *(35%) (2800)*	ʌ *(23%) (1840)*	i: *(2%) (160)*

Ce tableau montre que le phonème *ɪ* lorsqu'il est associé à la graphie « e », est dans 40 % des cas correctement reconnu comme ɪ, mais dans 35 % des cas il est reconnu comme e.

2.4 Création de règles de prononciations alternatives et génération de nouvelles prononciations

A l'aide de ces statistiques, nous créons des règles de prononciations pour les associations graphème-phonème selon deux critères : nous utilisons seulement les associations graphème-phonème qui ont un pourcentage suffisamment élevé et un nombre d'apparitions est suffisamment grand. On obtient alors un ensemble de règles qui serviront à générer de nouvelles prononciations. Par exemple, en utilisant un seuil de 20 % de taux de phonèmes reconnus et un nombre d'apparitions de 1500, pour notre corpus de parole non native nous obtenons les règles suivantes :

e : ɪ → e
e : ɪ → ʌ

Ces règles seront appliquées à l'ensemble des mots du lexique qui contient ces associations graphème-phonème. Cela permet de créer un nouveau lexique enrichi de prononciations alternatives non natives. Par exemple, pour le mot anglais *zero* du lexique avec les associations graphème-phonème suivantes :

zero z : z e : ɪ r : ɾ o : əʊ

et deux règles trouvées e:ɪ → e, e:ɪ → ʌ, après application de ces règles nous ajouterons dans notre lexique les prononciations suivantes :

zero z e ɾ əʊ
zero z ʌ ɾ əʊ

3 Expériences

Cette partie décrit les corpus utilisés et le protocole expérimental.

3.1 Les corpus utilisés

Pour valider l'approche proposée dans cet article, nous avons utilisé trois corpus de parole non natives. Ces trois corpus ont été prononcés par des français qui s'expriment en anglais :

- ***Nombres*** **:** Il est constitué d'enregistrements audio de locuteurs français non natifs lisant des nombres en anglais. Dans ce corpus, il y a 15 locuteurs, chacun a prononcé 66 commandes comprenant des nombres et de chiffres. La durée totale du corpus est de 78 minutes, le nombre des mots différents est 70. Le corpus contient de la parole lue.
- ***HIWIRE*** : Ce corpus est constitué d'enregistrements audios de locuteurs français non natifs lisant des commandes simples d'aéronautique en anglais. Ce corpus a été enregistré dans le cadre du projet *HIWIRE* (Bouselmi, 2008). Dans ce corpus, 31 locuteurs français ont chacun prononcé 100 commandes. Des données audios ont été enregistrées dans

différents bureaux à l'aide de micro-casque (*close-talking microphone*). Au total il y a 128 minutes de parole enregistrée. La taille du vocabulaire est de 130 mots différents. Le corpus correspond à la parole lue.

- ***Aéro* :** ce corpus contient des commandes aéronautiques en anglais prononcées par 2 locuteurs français. Les enregistrements correspondent à des commandes aéronautiques complexes. A la différence des deux corpus précédents qui étaient de la parole lue, ce corpus est un enregistrement de parole spontanée et de nombreuses hésitations et reprises sont présentes. La durée d'enregistrement est de 27 minutes. Le vocabulaire est de 900 mots.

La Table 1 résume ces corpus. Tous les corpus ont été transcrits manuellement. Nous utilisons le corpus *Aero* et une partie du corpus *HIWIRE* pour valider l'approche proposée et les **autres corpus** pour la **génération automatique des règles** de prononciations non natives.

Corpus	*Taille de vocabulaire (nbr. de mots)*	*Nbr. de locuteurs*	*Durée (minutes)*
Nombres	70	15	78
HIWIRE	130	31	128
Aero	900	2	27

Table 1 : Statistiques des corpus.

3.2 Système de reconnaissance

Notre système de reconnaissance est fondé sur la boîte à outils de reconnaissance vocale *Kaldi* (Povey et al., 2011). Nous utilisons les modèles acoustiques senones de type TDNN (*Time Delay Neural Network*).

Nous avons développé deux systèmes de reconnaissance.

- **SRAP_anglais**. Il est appris sur un corpus de conférences en **anglais** (TED-LIUM). L'apprentissage des modèles acoustiques anglais TDNN a été réalisé en utilisant les 452 heures du corpus d'apprentissage. 39 phonèmes anglais plus un modèle de silence et un modèle de bruit sont utilisés.
- **SRAP_anglais-français**. Il est appris en utilisant le même corpus que précédemment (les 452 heures de TED-LIUM) auquel on a ajouté un corpus radiophonique **français** (247 heures des corpus ESTER et ETAPE (Gravier *et al.*, 2012). Les phonèmes français et anglais sont appris simultanément : 30 phonèmes français, 39 phonèmes anglais, un modèle de respiration, un modèle de bruit et un modèle de silence. Ce système permet de reconnaître une personne s'exprimant en français ou en anglais et même de changer de langue au cours d'une même phrase.

Le lexique du système de reconnaissance **SRAP_anglais** utilise uniquement des phonème anglais. Le lexique du système **SRAP_anglais-français** contient les mêmes mots avec les mêmes prononciations (anglaises) mais certains mots ont également des prononciations qui utilisent les phonèmes français. Par exemple, pour les noms de villes françaises (comme, par exemple, *Marseille*), nous avons ajouté les prononciations canoniques françaises (avec des phonèmes français). C'est aussi le cas pour les mots transparents, c'est-à-dire les mots qui existent en français et en anglais, comme *ok* ou *possible.*

Les modèles de langage utilisés sont propres aux deux corpus de test : *HIWIRE* et *Aero*. Pour le corpus *HIWIRE*, le modèle de langage est une boucle de mots. Pour le corpus *Aéro*, nous avons appris un modèle tri-gramme sur un corpus textuel de commandes aéronautiques.
Les résultats de reconnaissance seront présentés en terme de taux d'erreur de mots (*Word Error Rate, WER*).

3.3 Protocole expérimental

Les nouvelles règles de prononciations non natives sont générées en utilisant soit le corpus *Nombres* seul, soit le corpus *Nombres* et le corpus *HIWIRE*. Les règles générées sont ensuite appliquées au lexique de départ pour générer des nouvelles prononciations non natives. Puis ce lexique augmenté est utilisé pour effectuer la reconnaissance des phrases du corpus de développement et de test. Pour chaque configuration, nous choisissons les valeurs qui maximisent le taux de reconnaissance sur le corpus de développement pour les paramètres suivants :

- le seuil de la distribution en pourcentage des phonèmes : paramètre α ;
- le seuil de nombre minimal de trames d'apparitions des phonèmes : paramètre β.

4. Résultats

Nous étudions l'impact du lexique de prononciation avec variantes non natives et l'impact des modèles acoustiques. Nous nous intéressons également à la génération des règles en utilisant le corpus de *Nombres* ou le corpus de *Nombres* et le corpus *HIWIRE*. α et β étant optimisés sur un corpus de développement.

Les résultats des expériences sont présentés dans les tables 2 et 3. Dans ces tables « Système de base » désigne le SRAP avec le lexique canonique avant l'enrichissement du lexique. La colonne *Phonèmes* désigne quel ensemble de phonèmes est utilisé pour représenter les phonèmes du lexique et les modèles acoustiques. Notons que le corpus *Aéro* contient la parole de seulement 2 locuteurs, donc il est délicat d'évaluer la variabilité des prononciations générées par notre approche. En revanche ce corpus correspond à de la parole spontanée et son vocabulaire est de 900 mots. De l'autre coté, le corpus *HIWIRE* présente une plus variabilité en termes de locuteurs (31). Mais le vocabulaire de ce corpus est plus petit (130 mots) et la parole est lue. Pour générer les règles de prononciations, pour la Table 2 nous utilisons le corpus *HIWIRE* entier, pour la Table 3 nous utilisons une partie de *HIWIRE* (cf. les explications ci-dessous).

	Phonèmes	*Corpus pour générer les règles*	*Nbr. de nouvelles prononciations*	*Corpus Aero WER (%)*
Système de base	anglais			13,84
	anglais+français			12,43
Approche proposée	anglais	*Nombres*	290	13,6
		Nombres et HIWIRE	290	13,6
	anglais+français	*Nombres*	219	12,38
		Nombres et HIWIRE	1148	**12,09**

Table 2. Résultats de reconnaissance en terme de WER (%). Corpus de test : *Aero*. α et β sont optimisés sur le corpus de développement.

La Table 2 montre que pour le corpus Aéro, utiliser juste les phonèmes anglais ne semble pas améliorer le taux d'erreurs (ligne 3 et 4 de Table 2). En revanche, en utilisant les phonèmes anglais et français notre approche de génération de nouvelles prononciations améliore les performances du SRAP. En générant les règles à partir des corpus *Nombres et HIWIRE,* nous obtenons 12,09 % WER par rapport à 12,43 % obtenu par le système de base.

La Table 3 montre les résultats de reconnaissance sur le corpus *HIWIRE* dans les mêmes configurations que les expériences présentées dans la Table 2. Pour *HIWIRE* nous avons utilisé la technique de validation croisée : nous utilisons la parole de 10 locuteurs pour générer les règles, puis 10 locuteurs comme corpus de développement et puis 11 locuteurs restant pour le test. Puis nous répétons l'opération en sélectionnant une autre répartition de locuteurs. Nous itérons 3 fois.

Dans la Table 3 nous observons que le nombre de règles générées est plus petit que celles données dans la Table 2, car le vocabulaire de *HIWIRE* contient seulement 130 mots par rapport au vocabulaire de 900 mots du corpus *Aero*. Nous observons une légère amélioration de taux d'erreurs de reconnaissance avec notre approche en utilisant les phonèmes anglais : de 9,7 % WER nous passons à 9,5 %. L'utilisation des phonèmes anglais+français n'améliore pas les résultats. Nous allons faire des investigations pour comprendre pourquoi le meilleur résultat n'est pas obtenu avec les phonèmes anglais+français et comment il est possible de mieux sélectionner les règles de prononciations non natives.

	Phonèmes	*Corpus pour générer les règles*	*Nbr. de nouvelles prononciations*	*HIWIRE WER (%)*
Système de base	anglais			9,7
	anglais+français			10,0
Approche proposée	anglais	*Nombres*	19	**9,5**
		Nombres et HIWIRE	35	**9,5**
	anglais+français	*Nombres*	20	9,9
		Nombres et HIWIRE	15	10,0

Table 3. Résultats de reconnaissance en terme de WER (%). Corpus de développement et de test : *HIWIRE (*validation croisée*)*. α et β sont optimisés sur le corpus de développement.

5. Conclusion

Dans cette étude nous avons présenté notre approche de génération des nouvelles prononciations pour l'enrichissement du lexique. Cette approche est conçue pour l'adaptation à la parole de locuteurs non natifs. Les nouvelles prononciations sont générées à partir des règles de prononciation déduites en utilisant un petit corpus représentatif de parole non native. Les règles prennent en compte les graphies des mots et permettent de substituer certains phonèmes par des phonèmes plus appropriés pour la parole non native. L'avantage de notre méthode est que seul le lexique du SRAP est modifié, les modèles acoustiques et le modèle de langage restent inchangés. Les expériences montrent que cette approche est pertinente pour enrichir le lexique de SRAP. Une validation sur un corpus de parole non native plus important serait intéressante. Nous travaillons sur la prise en compte des insertions et des suppressions de phonèmes effectuées par les locuteurs non natifs.

Remerciements

Les auteurs remercient la DGA (Direction Générale de l'Armement), Thales AVS et Dassault Aviation qui soutiennent le financement de cette étude et du programme scientifique «Man-Machine Teaming» dans lequel se déroule ce projet de recherche.

Références

BOUSELMI G. *(2008)*. Contributions à la Reconnaissance Automatique de la Parole Non Native. Thèse Université Lorraine.

BOUSELMI G., FOHR D., ILLINA I., AND HATON J.-P. (*2006)*. Fully Automated Non-Native Speech Recognition Using Confusion-Based Acoustic Model Integration and Graphemic Constraints. *In Proc. ICASSP*, France.

DAS A. AND HASEGAWA-JOHNSON M. *(2015)*. Cross-lingual Transfert Learning During Supervised Training in Low Resource Scenarions. *In Proc. of Interspeech.* pp. *3531–3535, 2015.*

DUAN R., KAWAHARA T., DANTSUJI M., AND ZHANG J. *(2017)*. Articulatory Modeling for Pronunciation Error Detection without Non-Native Training Data based on DNN Transfer Learning. *IEICE Transactions on Information and Systems, vol. E100D, no. 9, pp. 2174–2182.*

GRAVIER G., ADDA G., PAULSON N., CARRÉ M., GIRAUDEL A., GALIBERT O. (2012). The ETAPE Corpus for the Evaluation of Speech-based TV Content Processing in the French Language , *LREC.*

GORONZY S., RAPP S., KOMPE R. (2004). Generating non-native pronunciation variants for lexicon adaptation. *In Journal Speech Communication.*

LEE A. AND GLASS J. *(2015)*. Mispronunciation Detection Without Non-Native Training Data. *in Proc. of Interspeech, pp. 643–647.*

LI W., SINISCALCHI S. M., CHEN N. F., AND LEE C. *(2016)*. Improving Non-Native Mispronunciation Detection and Enriching Diagnostic Feedback with DNN-based Speech Attribute Modeling. *in Proc. of ICASSP, pp. 6135–6139.*

LIVESCU K. AND GLASS J. *(2000),* Lexical Modeling of Non-Native Speech for Automatic Speech Recognition, *In Proc. ICASSP.*

MATASSONI M., GRETTER R., FALAVIGNA D., GIULIANI D. (2018). Non-Native Children Speech Recognition Through Transfer Learning. *ArXiv:1809.09658.*

MORGAN J. (2004), Making a Speech Recognizer Tolerate Non-Native Speech Through Gaussian Mixture Merging, *In Proc. InSTIL/ICALL.*

PARK S. AND CULNAN JOHN (*2019)*. A comparison between native and non-native speech for automatic speech recognition.*The Journal of the Acoustical Society of America* 145, 1827.

POVEY D., GHOSHAL A., BOULIANNE G.I, BURGET L., GLEMBEK O., GOEL N., HANNEMANN M., MOTLICEK P., QIAN Y., SCHWARZ P., SILOVSKY J., STEMMER G., VESELY K. (*2011)*. The Kaldi Speech Recognition Toolkit, *IEEE 2011 ASRU Workshop.*

PRÉVEIL B., MARTENS J.P., VAN DEN HEUVEL H. (2010). Improving proper name recognition by adding automatically learned pronunciation variants to the lexicon. *In Proc. LREC.*

SCHADEN S. *(2004),* Generating Non-Native Pronunciation Lexicons by Phonological Rule, *InProc ICSLP2004.*

TAN T.-P. (2008), Automatic Speech Recognition for Non-Native Speakers, *thèse de l'université Joseph Fourrier,* 2008.