



HAL
open science

Caractérisation des accents étrangers en utilisant des techniques de reconnaissance automatique de la parole

Mohamed Ben Jannet

► **To cite this version:**

Mohamed Ben Jannet. Caractérisation des accents étrangers en utilisant des techniques de reconnaissance automatique de la parole. Linguistique. 2011. dumas-00609869

HAL Id: dumas-00609869

<https://dumas.ccsd.cnrs.fr/dumas-00609869>

Submitted on 20 Jul 2011

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.

Caractérisation des accents étrangers en utilisant des techniques de reconnaissance automatique de la parole

Projet réalisé par :
Mohamed Ameer Ben jannet

Mémoire de **Master 2**
Mention **Science du Langage**,
Spécialité **Modélisation et Traitements Automatique en**
Industries De la Langue : Parole, Ecrit, Apprentissage
Orientation Recherche
Année universitaire 2010-2011

Responsables :
Laurent Besacier
Solange Rossato

Membres du jury :
Georges Antoniadis
Isabelle Rousset

Remerciements

Je voudrais tous d'abord remercier M. Laurent *Besacier* et Mme. Solange *Rossato*, pour m'avoir accueilli au sein du LIG et pour avoir accepté d'être mes encadrants tout au long de ce stage. Un grand merci à M. Laurent *Besacier* qui a toujours su trouver les bons mots pour me remonter le morale et me pousser vers l'avant, pour tous ces critiques et ses conseils qui m'ont permis de bien mener mon stage, pour les disciplines de recherches et le travail en groupe qu'il m'a appris à respecter et aimer et pour avoir lu et commenté très soigneusement ce manuscrit. Un grand merci également à Mme. Solange *Rossato*, pour tous ce qu'elle m'a appris sur la phonétique, le traitement de signal et la statistique, pour ces conseils précieux et pour avoir lu et commenté soigneusement mon manuscrit.

J'adresse mes remerciements à M. Georges *Antoniadis* qui a été toujours présent tout le long du Master pour répondre à nos questions, et un grand merci pour tous les membres du jury qui m'ont honoré de leur présence à la présentation de mes travaux de stage.

Je tiens à remercier également tous les membres de l'équipe GTALP (Michel, Juliette, Benjamin, Pedro, David...) pour leur accueil et leur sympathie.

Un grand merci à mes amis (Mohamed Amine, Ayoub, Aymen, Horia, Jassem, Remus....) avec qui j'ai partagé des grands moments au cours de mon stage.

Enfin, je voudrais exprimer mes plus profonds remerciements à ma famille pour leurs encouragements durant tous le temps où j'ai passé mon Master.

Résumé :

L'apprentissage des langues étrangères est une des sources de la diversité d'accents observés chez les locuteurs non-natifs. Le signal de parole nous fournit des informations précieuses sur le locuteur. Ces informations peuvent renseigner sur sa maîtrise de la langue et les problèmes de prononciation qu'il peut avoir. Une analyse des caractéristiques de la parole des apprenants peut contribuer à perfectionner le processus d'apprentissage des langues étrangères. De plus, cette analyse peut aider à l'amélioration des applications technologiques telles que la reconnaissance automatique de la parole et du locuteur.

Notre mémoire propose une étude phonétique acoustique des accents étrangers en arabe, en utilisant des outils automatiques de reconnaissance de la parole : alignement forcé et décodage acoustico-phonétique, afin de caractériser la variation d'accent liée au niveau de langue du locuteur. Les analyses des résultats obtenus permettent de détecter des différences liées au niveau de langue du locuteur et les phonèmes susceptibles de causer des problèmes pour les apprenants de l'arabe.

Mots clés : apprentissage des langues, accents, prononciation, alignement forcé, étude phonétique acoustique

Sommaire

Résumé :	4
Sommaire.....	5
Liste des Tableaux :	7
Liste des figures :	7
Introduction:.....	8
Chapitre 1 : Contexte et état de l'art :	9
1. Les facteurs affectant l'acquisition de la langue :	9
1.1. Effet de l'âge sur l'acquisition d'une seconde langue :	9
1.2. Le degré d'exposition à la L1 par rapport à la L2 :	10
1.3. Interaction entre L1 et L2 :	10
2. Evaluation des niveaux de langues :	11
3. Outils automatiques de détection d'accents :	13
3.1. Caractérisation d'accent en utilisant des paramètres prosodiques :	13
3.2. Une étude récente de détection de la parole non native :	14
4. Conclusion.....	14
Chapitre 2 : Méthodologie.....	15
1. Introduction :	15
2. Corpus :	15
2.1. Présentation du corpus :	15
2.2. Caractéristique du corpus.....	16
2.3. Prétraitement et préparation des données :	16
a. Découpage du signal :	16
b. Adaptation des transcriptions :	17
c. Préparation du dictionnaire de prononciation :	17
2.4. Description du corpus :	18
3. Outils automatiques :	19
3.1. Boite à outils Sphinx :	19
3.2. Modélisation acoustique :	19
3.3. Alignement forcé :	20
3.4. Le décodage acoustico-phonétique :	21

4. Conclusion.....	22
Chapitre 3 : Analyses des résultats	23
1. Introduction :.....	23
2. Construction du modèle acoustique :.....	23
3. Analyse des scores globaux obtenus suite à l’alignement forcé :	26
3.1. Analyse du comportement des scores selon le niveau de langue :.....	27
3.2. Analyse de distribution des scores selon le niveau de langue :.....	28
a. corpusA1 :.....	29
3.3. Evaluation de la prononciation des apprenants :.....	32
4. Analyse en fonction des types de phonèmes :.....	33
5. Analyse des résultats pour un modèle acoustique « à gros grain » :	36
5.1. Construction du modèle acoustique :.....	36
5.2. Comparaison avec les résultats du modèle classique :.....	38
a. Comparaison des comportements des scores :	38
b. Comparaison de la distribution des scores :.....	39
6. Analyse des résultats du DAP :.....	41
6.1. Analyse du taux d’erreur pour les catégories phonétique :	42
6.2. Confusions :.....	43
Conclusion :.....	49
.....	50
Bibliographie.....	51

Liste des Tableaux :

Tableau 1: Critère d'évaluation du CECR.....	12
Tableau 2: translitération de l'Arabe vers le format Buckwalter	17
Tableau 3 : Distribution des locuteurs par corpus de mots.....	18
Tableau 4 : Distribution des mots dans le corpus.....	19
Tableau 5 : Table de correspondance entre les symboles utilisés et les symboles API	24
Tableau 6:les classes des phonèmes pour le modèle acoustique moyen	37
Tableau 7 : taux d'erreur pour chaque catégorie de locuteurs.....	42

Liste des figures :

Figure 1 : HMM à trois état.....	20
Figure 2 : Processus d'alignement forcé.....	21
Figure 3 : Décodage acoustico-phonétique	22
Figure 4 : Paramétrisation du signal.....	24
Figure 5 : Les fichiers d'entrer et de sortie du module SphinxTrain	25
Figure 6 : modélisation du phonème N dans le mot "E IH N AE"	26
Figure 7: Comportement des scores d'alignement forcé selon les niveaux de langue.....	27
Figure 8:distribution des scores d'alignement forcé dans le corpusA1	29
Figure 9:distribution des scores d'alignement forcé dans le corpusA2B1	30
Figure 10:la distribution des scores d'alignement forcé sur tous le corpus.....	31
Figure 11: Evaluation de la prononciation selon les scores acoustique	33
Figure 12: les scores des apprenants par catégorie phonétique.....	34
Figure 13 : Modélisation du N dans le modèle acoustique moyen	37
Figure 14 : scores d'alignement en utilisant le modèle acoustique « à gros grain ».....	38
Figure 15 : Distribution des scores obtenus en utilisant le modèle acoustique moyen	39
Figure 16 : évaluation des niveaux de prononciation	41
Figure 17 : taux d'erreur par phonème	42
Figure 18 : Extraction des confusions avec SCLITE.....	43
Figure 19 : confusions voyelles courtes voyelles longues	45
Figure 20 : le [h] un son qui attire tous les phonèmes.....	46
Figure 21 : confusion des sons proches.....	47
Figure 22 : Les approximantes.....	48

Introduction:

De nos jours, l'apprentissage des langues intéresse un grand nombre de personnes à travers le monde pour des raisons différentes : voyage, études à l'étranger, intérêts économique, immigration, intérêt scientifique, religieux et artistique etc....

Souvent, les locuteurs natifs parviennent à détecter un accent étranger dans la prononciation des personnes étrangères. Même pour des apprenants ayant atteint un niveau avancé et une maîtrise du vocabulaire et de la grammaire de la langue cible (L2) certains traits dans leur prononciation diffèrent de la prononciation des locuteurs natifs.

Aujourd'hui, l'identification automatique d'accent est un nouveau challenge lié à des domaines tel que la reconnaissance automatique de la parole, l'identification du dialecte ou l'identification du locuteur. L'accent contient des informations précieuses qui renseignent sur le locuteur, ces informations peuvent être utilisées pour construire des modèles adaptés aux locuteurs et qui peuvent aider pour différentes tâches liées à la reconnaissance vocale.

L'étude menée dans le cadre de ce mémoire porte sur l'utilisation de la reconnaissance automatique de la parole (RAP), dans un objectif d'aide à l'apprentissage des langues.

Plus précisément, notre travail consiste à trouver les outils automatiques qui permettent la détection et la caractérisation des accents suivant le niveau de langue du locuteur en essayant de localiser les problèmes de prononciation pour les apprenants. Ce qui permettra de trouver des solutions pour améliorer les systèmes de reconnaissance de la parole, pour aider les apprenants à localiser leurs problèmes de prononciation et guider les professeurs des langues à définir les exercices les mieux adaptés pour chaque niveau d'apprenants.

Dans le premier chapitre, nous présentons brièvement les différents facteurs influant la prononciation d'une langue seconde (L2) rendant ses caractéristiques différentes de celles de la langue maternelle (L1) et quelques outils permettant de détecter ses caractéristiques.

Dans le deuxième chapitre l'approche méthodologique envisagée est décrite : Nous présenterons, tout d'abord notre corpus de travail, ensuite les outils automatiques que nous avons utilisés dans notre travail: alignement forcé et décodage acoustico-phonétique (DAP).

Dans le troisième chapitre nous présenterons les analyses détaillées effectuées ainsi que l'ensemble des résultats obtenus avant de proposer les perspectives envisagées pour poursuivre ce travail.

Chapitre 1 : Contexte et état de l'art :

La communication orale est très importante durant la vie d'une personne, elle permet de prendre la parole, donner son point de vue et l'argumenter, mais il faut d'abord maîtriser la langue parlée par le groupe avec lequel on veut communiquer. Par ailleurs, l'apprentissage des langues étrangères devient de plus en plus important pour pouvoir communiquer d'une manière intelligible. Il est donc important que la communication se déroule dans le respect des règles de prononciation de la langue en question.

Le signal de parole renferme une quantité d'information précieuse mais complexe à analyser. Grâce aux recherches qui ont été menées pendant les dernières décennies les outils automatiques de reconnaissance de la parole nous permettent d'extraire ces informations qui peuvent renseigner les apprenants et les professeurs de langues sur les difficultés et les problèmes de prononciation pour mieux les surmonter. Ces informations sont aussi importantes pour l'amélioration des systèmes de reconnaissance de la parole. En effet les systèmes existant aujourd'hui ont atteint des performances assez élevées pour la parole native mais dès qu'il s'agit de parole non-native ce taux se dégrade rapidement. Ceci est dû aux caractéristiques de la parole non native qui diffèrent de celles de la parole native. Donc il faut adapter les systèmes à ces variations. Mais devant la grande diversité d'accent une représentation de toutes les prononciations et l'acquisition de corpus de parole non-native nécessite énormément de temps et de ressources. Donc il faut trouver des solutions plus flexibles qui tiennent compte de la différence d'accent.

Dans ce chapitre nous allons expliquer brièvement pourquoi la parole native est différente de la parole non-native. Nous allons ensuite mettre l'accent sur l'effet de la différence de niveau de langue sur la prononciation, en nous basant sur des critères définis par l'union européenne. Nous verrons ensuite quels sont les outils automatiques et les méthodes de détection et caractérisation d'accents existantes dans le domaine.

1. Les facteurs affectant l'acquisition de la langue :

1.1. Effet de l'âge sur l'acquisition d'une seconde langue :

L'âge d'acquisition d'une langue étrangère, est un critère important qui influence l'apprentissage de la langue. Des nombreuses recherches montrent que les meilleurs résultats sont pour les jeunes par opposition aux apprenants plus âgés, il y aurait donc une période critique dans l'acquisition d'une L2, période pendant laquelle l'apprentissage d'une langue est plus facile qu'à d'autres périodes de l'existence. Ce point de vue implique qu'après la période critique, il est plus difficile de réussir l'acquisition d'une seconde langue et il devient impossible de parvenir à une compétence semblable à celle d'un natif (Patkowski, 1990) et (Flege, 2001)

1.2. Le degré d'exposition à la L1 par rapport à la L2 :

La durée d'exposition et la motivation à apprendre la L2 sont des facteurs qui influencent l'apprentissage de la langue. Des recherches sur ce sujet ont montré que les personnes qui pratiquent la L2 avec des locuteurs natifs plus régulièrement que leur L1 atteignent des niveaux de langue supérieur à ceux qui continuent d'utiliser leur L1 couramment (Flege, 2001); le degré d'immersion dans un environnement de locuteurs natifs de la L2 est un facteur important qui affecte l'accent de l'apprenant.

1.3. Interaction entre L1 et L2 :

Les études portant sur l'interaction entre L1 et L2 de (Flege, 1995); ont montré que le mécanisme utilisé par l'enfant pour construire l'espace phonémique de la L1 reste intacte et sera accessible pour apprendre le L2 et que ses deux systèmes phonémiques (le système phonétique de la L1 et celui de la L2) existent dans le même espace donc ils s'influencent mutuellement (Flege, 1995).

Ainsi, lorsque l'apprentissage de la L1 se fait lentement pendant l'enfance et l'adolescence, et il devient plus probable que, en apprenant une seconde langue (L2), les apprenants jugent des voyelles et des consonnes de la L2 comme des instances des voyelles et des consonnes de la L1 qui sont préétablies dans leurs espace phonémique (Flege, 1995). Selon Flege dans le modèle SLM (Speech Learning Model) les apprenants d'une langue seconde ont plus de difficulté à acquérir les sons analogues à ceux de L1 par rapport aux nouveaux sons car les nouveaux sons permettent la création d'une nouvelle catégorie phonémique indépendante alors que les sons proches de ceux de langue native sont influencés par la représentation déjà établie au cours de l'apprentissage de la L1 (Flege, 1995). Ce modèle montre que l'apprentissage d'une langue seconde sera conditionné par la première.

D'autres recherches menées par (Kuhl et al, 2008) dans le modèle PME (Perceptual Magnet Effect) évoquent, l'existence d'une attraction magnétique induite par les prototypes phonémiques créés au cours de l'apprentissage de la L1 qui influence la perception des sons de la L2 qui se trouvent autour d'un prototype de la langue native. L'espace perceptif est donc dépendant de ce qui a été appris pendant l'apprentissage de la langue native.

Par conséquent, le modèle PME prédit que la difficulté d'apprendre un son d'une langue donnée dépendra de la distance qui le sépare d'un prototype de la langue native (kuhl, 2008).

2. Evaluation des niveaux de langues :

Comme on vient de voir l'apprentissage des langues est une tâche influencée par différents facteurs et les apprenants ont des niveaux de langue différents suivant leurs conditions d'apprentissage. Pour évaluer le niveau de langue des apprenants plusieurs tests et critères existent. Nous allons nous intéresser aux critères adoptés par l'Union Européenne puisque notre travail porte sur des Français apprenant l'arabe.

Le cadre européen commun de référence pour les langues (CECRL) essaye de définir une échelle d'évaluation de la maîtrise d'une langue qui est indépendante de l'organisme évaluateur, et transposable à n'importe quelle langue. Ce document définit une échelle qui se décompose en trois sections :

- Niveau élémentaire de compétences : Niveau A
- Niveau intermédiaire : Niveau B
- Niveau avancé : niveau C

Chaque section se décompose elle-même en deux niveaux le niveau 1 et le niveau 2.

Pour l'action pédagogique et l'évaluation, il est fondamental de comprendre que la distinction en six niveaux proposé dans l'échelle du CECR n'est pas suffisante. Un niveau global (A2 ou B1 par exemple) n'a que peu de sens.

Le niveau de compétences d'un élève ne peut s'observer qu'à travers des activités langagières, d'où l'importance de décrire ces six niveaux de compétences de façon distincte pour chaque type d'activité langagière (CERC, 2009).

Nous avons travaillé sur un corpus d'apprenant construit par Mohmed Belgacem du laboratoire LIDILEM qui contient des enregistrements de locuteurs ayant les niveaux de langue : A1, A2, B1. Voilà la grille d'évaluation du CECR pour ces niveaux.

		A1	A2	B1
Comprendre	Ecouter	Je peux comprendre des mots familiers et des expressions très courtes ou sujets de moi-même de mot familiers et de l'environnement concret et immédiat si les gens parlent immédiatement si les gens parlent lentement et distinctement.	Je peux comprendre des expressions et un vocabulaire très fréquent relatifs à ce qui me concerne de très près par exemple moi-même, ma famille les achats, l'environnement proche, le travail. Je peux saisir l'essentiel d'annonces et de messages simples et clairs.	Je peux comprendre les points essentiels quand un langage clair et standard est utilisé et s'il s'agit de sujets familiers Concernant le travail, l'école, les loisirs, ect. Je peux comprendre l'essentiel de nombreuses émissions de radio ou de télévision sur l'actualité ou sur des sujets qui m'intéressent à titre personnel ou professionnel si l'on parle d'une façon relativement lente et distincte.
		Je peux comprendre des noms familiers, des mots	Je peux lire des textes courts très simples. Je	Je peux comprendre des textes rédigés essentiellement dans

	Lire	ainsi que des Phrases très simples, par exemple dans des affiches ou des catalogues.	peux trouver une information particulière prévisible dans des documents courants comme les publicités, les prospectus, les menus et les horaires et je peux comprendre des lettres personnelles courtes et simples.	une langue courante ou une relative à mon travail. Je peux comprendre la description d'évènement, l'expression de sentiments et de souhaits dans des lettres personnelles.
Parler	Prendre part d'une conversation	Je peux communiquer, de façon simple, à condition que l'interlocuteur soit disposé à répéter ou à reformuler ses phrases plus lentement et à m'aider à formuler ce que j'essaie de dire. Je peux poser des questions simples sur des sujets familiers ou sur ce dont j'ai immédiatement besoin, ainsi que répondre à de telles questions.	Je peux communiquer lors de tâches simples et habituelles ne demandant qu'un échange d'information simple et direct sur des sujets et des activités familiers. Je peux avoir des échanges très brefs même si, en règle générale, je ne comprends pas assez pour poursuivre une conversation.	Je peux faire face à la majorité des situations que l'on peut rencontrer au cours d'un voyage dans une région où la langue est parlée. Je peux prendre part sans préparation à une conversation sur des sujets familiers ou d'intérêt personnel ou qui concernent la vie quotidienne.
	S'exprimer oralement en continu	Je peux utiliser des expressions et des phrases simples pour décrire mon lieu d'habitation et les gens que je connais.	Je peux utiliser une série de phrases ou d'expressions pour décrire en termes simples ma famille et d'autres gens, mes conditions de vie, ma formation, mon activité professionnelle actuelle ou récente.	Je peux m'exprimer de manière simple afin de raconter des expériences et des événements, mes rêves, mes espoirs ou mes buts. Je peux brièvement donner les raisons et les explications de mes opinions ou projets, je peux raconter une histoire ou l'intrigue d'un livre ou d'un film et exprimer mes réactions.
Ecrire	Ecrire	Je peux écrire une courte carte postale simple, par exemple de vacances, je peux parler des détails personnels dans un questionnaire, inscrire par exemple mon nom, ma nationalité et mon adresse sur une fiche d'hôtel.	Je peux écrire des notes et messages simples et courts. Je peux écrire une lettre personnelle très simple, par exemple de remerciements.	Je peux écrire un texte simple et cohérent sur des sujets familiers ou qui m'intéressent personnellement pour décrire des expériences et impressions.

Tableau 1: Critère d'évaluation du CECR

L'évaluation du niveau de langue et du niveau de prononciation pour les apprenants des langues semble être une tâche difficile et dépendante de plusieurs facteurs. Notre but durant ce stage est de savoir si à l'aide d'outils automatiques, il serait possible de détecter un écart à une norme de prononciation pour les apprenants, qui serait lié à des niveaux différents dans l'apprentissage de la L2.

3. Outils automatiques de détection d'accents :

Les outils de traitement automatique, et notamment la reconnaissance automatique de la parole se sont trouvés en face du problème de la variété d'accent y compris l'accent non natif. Les performances des systèmes de reconnaissance sont vite dégradées pour la parole non-native, s'est alors posé la question de détecter et caractériser l'accent du locuteur non natif afin d'appliquer les adaptations seulement si nécessaire (T.Ping, 2008).

3.1. Caractérisation d'accent en utilisant des paramètres prosodiques :

La prosodie est l'étude des phonèmes d'accentuation et d'intonation. Elle permet l'assertion, l'interrogation, l'injonction, l'exclamation... Les systèmes d'identification de l'origine du locuteur non natif basé sur les paramètres prosodiques utilisent la durée, l'énergie et la fréquence fondamentale principalement.

Durée : A l'aide des paramètres prosodiques les systèmes automatiques arrivent à détecter l'origine du locuteur en calculant le rapport entre la durée de la syllabe et la durée totale du mot. Le patron rythmique diffère en effet en fonction des langues : sur des mots dissyllabiques, un français à tendance à allonger la 2^{ème} syllabe, tandis que un japonais accentue la 1^{ère} syllabe et ce patron rythmique est acquis des l'acquisition des premiers mots.

Energie : selon la prononciation du locuteur l'énergie peut varier d'une syllabe à une autre. La prononciation d'un locuteur non natif est influencée par les habitudes acquises pendant son apprentissage de sa langue maternelle. Or, la distribution de l'énergie est également variable d'un locuteur à l'autre selon l'intensité de la voix. Il est alors important de normaliser lorsqu'on s'appuie sur l'énergie.

La fréquence fondamentale : La fréquence fondamentale marque l'intonation et permet souvent la distinction entre des mots ayant la même orthographe ou une prononciation similaire. Le locuteur non natif peut commettre des erreurs dans la prononciation de ces mots dûes à l'influence de sa langue maternelle ce qui permettra d'identifier son origine.

Différentes recherches se sont basées sur les paramètres prosodiques pour la caractérisation des accents étrangers, dont (Vieru-Dimulescu, 2006) qui a essayé d'identifier six accents étrangers en français et l'étude de (C.Horgues, 2006) portant sur la caractérisation de l'accent français pour la langue anglaise.

3.2. Une étude récente de détection de la parole non native :

L'influence de l'accent dans le signal de parole peut être faible par rapport à d'autres facteurs comme le genre du locuteur et la distorsion du canal de transmission.

Des études menées par (Xiao, 2011) visent à déterminer quelles sont les caractéristiques du signal les plus résistantes aux variations du canal et par suite trouver sur quel type de paramètres, il faut s'appuyer pour la détection du locuteur non natif.

Pour identifier les paramètres les plus adaptés à la tâche, Xiao a comparé les résultats obtenus à partir des analyses basées sur des mixtures de gaussiennes (GMM) appliqués sur différents type de données : la modulation du spectre du signal, l'analyse des formants, la fréquence fondamentale et les paramètres MFCC. Il a aussi effectué des analyses en combinant les différents paramètres cités et il a montré qu'un modèle basé sur une fusion entre la fréquence fondamentale (f_0) et la modulation du spectre est la plus résistante à la variation du canal de transmission. Cette combinaison donne les meilleurs résultats pour la détection de la parole non native à l'aide des GMM.

4. Conclusion

Dans ce chapitre nous avons présenté les problèmes liées à l'apprentissage des langues et les raisons qui font que les conditions d'apprentissage soient importantes dans l'analyse de la différence d'accent. Nous avons présenté quelques outils qui permettent de caractériser les différences d'accent : l'analyse des paramètres prosodiques et caractéristiques (MFCC, la fréquence fondamentale, modulation du spectre). Notre étude se situe dans le même contexte de la caractérisation des accents étrangers, et nous allons suivre une approche basée sur les scores d'alignements automatiques de signaux de parole d'apprenants dans une tâche de lecture de mots isolées. Le travail consiste à comparer les résultats des apprenants à ceux des natifs et se focalise sur les erreurs de prononciation et non sur les erreurs de syntaxique et lexique. Dans le prochain chapitre nous allons présenter notre corpus de travail ainsi que outils automatiques utilisés : alignement forcé et décodage phonémique.

Chapitre 2 : Méthodologie

1. Introduction :

Notre étude sur la différence d'accent des apprenants de langue nécessite des ressources linguistiques (corpus d'apprenants), et des outils automatiques permettant le traitement des données afin de détecter les différences existantes dans la parole en fonction du niveau de langue.

Dans ce chapitre, nous allons présenter les ressources exploitées et les différents outils automatiques utilisés.

2. Corpus :

Pour effectuer notre étude, il nous faut un corpus d'apprenants, dans ce but nous avons choisi de travailler sur un corpus d'apprenants de l'arabe (français apprenant l'arabe). Le français et l'arabe sont deux langues assez différentes, les apprenants sont confrontés à des nouveaux sons n'existant pas dans leur langue maternelle et aussi à des sons qui existent ou sont proches de ceux de leur langue.

Nous voulons observer le comportement des apprenants dans ces situations et voir si les outils automatiques peuvent détecter des différences entre des apprenants de différents niveaux.

2.1. Présentation du corpus :

Notre corpus de travail a été construit par M.Belgacem du laboratoire LIDILEM dans le but d'étudier les différences de prononciation et d'articulation des apprenants de la langue arabe afin de détecter leurs difficultés et de les aider à les surmonter.

Le corpus a été enregistré à l'université Stendhal de Grenoble, il est constitué de prononciations de mots isolés. C'est le professeur d'arabe de l'université qui a choisi les mots constituant le corpus de telle façon qu'on peut obtenir les différents phonèmes de l'arabe dans plusieurs contextes. Le corpus est constitué de trois groupes suivant : les mots appris aux cours du niveau A1, les mots appris aux cours de niveau A2 et les mots appris aux cours de niveau B1.

On demande ensuite aux apprenants de lire les mots un par un. Les élèves de niveau A1 prononcent seulement les mots qui sont appris dans le cours des A1, les autres élèves (A2 et B1) prononcent tous les mots des cours A1, A2 et B1.

2.2. Caractéristique du corpus

Le corpus est constitué d'enregistrements des apprenants avec les transcriptions correspondantes et il propose aussi un étiquetage précis des informations concernant les locuteurs ce qui facilitera par la suite notre analyse des données et des résultats.

Pour chaque locuteur on a les informations suivantes :

- Le nom du locuteur
- Le sexe
- Le niveau de langue
- La langue maternelle, ici le français uniquement.

Pour l'évaluation du niveau de prononciation, il n'existe pas des règles précises pour cette tâche délicate. Nous avons alors décidé d'évaluer nos apprenants en comparant leurs prononciations à celle des natifs mais comme le corpus ne contient pas de locuteurs natifs nous avons décidé d'acquérir ces données importantes pour notre étude. Pour cela nous avons demandé à des locuteurs de langue maternelle arabe de lire les mêmes mots qui ont été prononcés par les apprenants. Nous avons obtenu des enregistrements correspondant à deux locuteurs masculins âgés de 24 ans de langue maternelle arabe tunisien.

2.3. Prétraitement et préparation des données :

Pour qu'on puisse exploiter les données du corpus avec les outils automatiques du LIG une phase de prétraitement et d'adaptation importante a été réalisée pour mener à bien notre travail.

a. Découpage du signal :

La première tâche a consisté dans le découpage et l'organisation des enregistrements. A l'aide du logiciel Audacity nous avons écouté et découpé tous les signaux enregistrés et nous avons enlevé les parties endommagées ou qui contiennent beaucoup de bruit. A la fin, nous avons obtenu une base de données constituée d'enregistrements au format WAV. Chaque enregistrement contient le signal correspondant à la prononciation d'un seul mot, le nom du fichier est une abréviation du nom du locuteur et un numéro correspondant au mot

prononcé.

Exemple :

L'enregistrement : « AU11.wav » correspond au locuteur Aurore qui a prononcé le mot numéro 11.

b. Adaptation des transcriptions :

Les outils automatiques que nous avons choisi d'utiliser n'acceptent pas les caractères graphiques de l'arabe. Nous devons donc traduire les transcriptions vers un codage accepté par ces outils. Nous avons choisi de translittérer les caractères arabe vers le format nommé « Buckwalter¹ ». Des outils automatiques présents au LIG nous permettent d'effectuer cette tâche facilement.

Transcriptions en arabe	Transcriptions en Buckwalter
<s> سكن </s>	<s> skn </s> (au11)
<s> طالب </s>	<s> TALb </s> (au12)
<s> جزائري </s>	<s> jzA}ry </s> (au13)
<s> ماذا </s>	<s> mA*A </s> (au14)
<s> شارع </s>	<s> mA*A\$ArE </s> (au15)
<s> نكلم </s>	<s> tkl~m </s> (au16)
<s> اسبانيا </s>	<s> AsbAnyA </s> (au21)
<s> صف </s>	<s> Sf </s> (au22)
<s> لقاء </s>	<s> lqA' </s> (au23)
<s> هل </s>	<s> hl </s> (au24)
<s> جامعة </s>	<s> jAmEp </s> (au25)
<s> تفاح </s>	<s> tfAH </s> (au26)

Tableau 2: translittération de l'Arabe vers le format Buckwalter

c. Préparation du dictionnaire de prononciation :

Une ressource indispensable est le dictionnaire de prononciation. Il contient la transcription phonétique de tous les mots présents dans le corpus. A l'aide d'un phonétiseur automatique de l'arabe développé au LIG nous avons obtenu un fichier contenant une transcription phonétique possible pour chaque mot que nous avons vérifié manuellement pour corriger les erreurs et ajouter d'autres transcriptions si le mot a plusieurs prononciations possibles.

1Représentation ASCII des caractères arabe

Voici un extrait du dictionnaire de prononciation obtenu :

skn S AE K AE N AE

TAIb TT AE: L IH B

jzA}ry J AE Z AE: E IH R IH Y

mA*A MAE: DH AE

\$ArE SH AE: R IH AI

tkl~m T AE K AE LL AE M

tkl~m(2) T AE K AE LL AE MAE

AsbAnyA E IH S B AE: N Y AE:

2.4. Description du corpus :

Une fois les prétraitements et l'organisation du corpus terminés, le corpus obtenu compte 1919 mots prononcés par neuf locuteurs. Mais comme nous l'avons déjà expliqué les locuteurs de niveau de langue A1 ne prononcent que les mots issus du cours A1. La distribution des locuteurs suivant les mots prononcés est présentée dans le tableau 3.

Niveau de langue	mots du niveau A1	mots du niveau A2	mots du niveau B1	Nombre de mots prononcé
A1	3 locuteurs	/	/	360 mots
A2	2 locuteurs	2 locuteurs	2 locuteurs	508 mots
B1	1 locuteur	2 locuteurs	2 locuteurs	407 mots
Natifs	2 locuteurs	2 locuteurs	2 locuteurs	644 mots

Tableau 3 : Distribution des locuteurs par corpus

Comme nous pouvons remarquer dans le tableau ci-dessous notre corpus peut se diviser en deux sous corpus : le corpus des mots A1 qui sont prononcés par tout le monde (noté par suite corpusA1) ; le deuxième corpus prononcé par les apprenants de niveau A2 ,B1 ainsi que natifs et que nous allons appeler corpusA2B1, Cette décomposition nous permet de faire une analyse en comparant les différents locuteurs sur les mêmes difficultés.

Après l'organisation et le traitement du corpus nous avons enlevé les enregistrements inexploitable et voilà la distribution des mots dans le corpus suivant le niveau de langue.

Niveau de langue des locuteurs	Nombre de mots A1	Nombre de mots A2	Nombre de mots B1	Total
A1	360 mots	0 mots	0 mots	360 mots
A2	105 mots	169 mots	234 mots	508 mots
B1	120 mots	170 mots	117 mots	407 mots
Natifs	240 mots	170 mots	232 mots	644 mots
Tous les niveaux	825 mots	509 mots	583 mots	1919 mots

Tableau 4 : Distribution des mots dans le corpus

3. Outils automatiques :

Les outils automatiques pour l'analyse sont un aspect important dans notre étude car nous nous basons sur ces outils pour mener notre étude et essayer de caractériser automatiquement la différence de niveau de langue entre les locuteurs. Avant de décrire les méthodes utilisées, nous allons commencer par décrire la boîte à outils expérimentale qui nous a permis d'obtenir nos résultats.

3.1. Boîte à outils Sphinx :

La boîte à outils CMU sphinx est l'une des plus connues et les plus robustes dans le domaine de la reconnaissance vocale, elle contient des outils permettant l'entraînement d'un modèle acoustique, la compilation d'un modèle de langage et différents traitements permettant l'analyse et l'exploitation des signaux de parole.

Nous allons utiliser sphinx3, c'est une version libre de sphinx compatible avec le système d'exploitation Unix, l'utilisation de cette boîte amène modularité, flexibilité et facilité de développement dans notre projet.

3.2. Modélisation acoustique :

Les approches statistiques et les modèles probabilistes sont largement utilisés, de nos jours, dans les systèmes de reconnaissance automatique de la parole. Ces approches, notamment celles basées sur les modèles de Markov cachés(HMM), ont atteint des performances remarquables avec des vocabulaires de plus en plus importants et une robustesse au bruit et à la variabilité des locuteurs de plus en plus grandes (Rabiner, 1993).

Un modèle acoustique est un modèle statistique qui peut produire une ou plusieurs

hypothèses phonétiques pour chaque segment de parole (vecteur acoustique). Il est similaire à un générateur d'hypothèses et il se base sur des modèles statistiques de phonèmes, qui sont entraînés sur des données de parole contenant les différentes unités de la langue dans différents contextes. Les hypothèses sont générées grâce au HMM (Hidden Markov Modèle).

Les modèles de Markov cachés sont très utilisés dans les domaines de reconnaissance des formes, ces modèles ont subi plusieurs améliorations au fil du temps, notamment en introduisant des notions comme les corrélations entre trames et de modélisation de trajectoires.

Un modèle de Markov caché est un automate stochastique à nombre d'état fini N (Rabiner, 1993). Un processus aléatoire se déplace d'état en état. On peut noter E_t l'état où se trouve le processus à l'instant t .

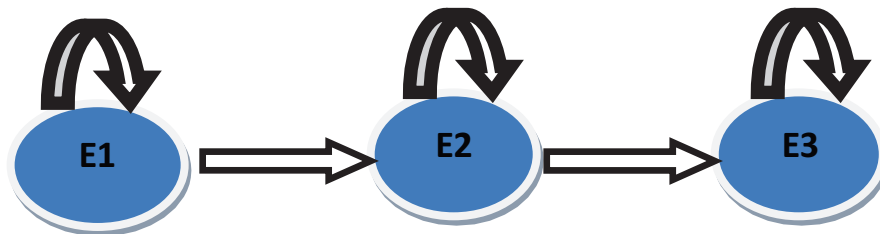


Figure 1 : HMM à trois état

Chaque état représente la probabilité de produire l'évènement X à l'état E à un instant donné.

- Etat = $\text{prob}(X_k/E_i)$: probabilité de produire l'évènement X_k sur l'état E_i .
- Transition : Probabilité de passer d'un état à un autre $\text{prob}(E_{i+1}/E_i)$.

Comme nous avons expliqué dans le paragraphe précédent la tâche d'alignement forcé et le DAP nécessite un modèle acoustique.

Pour notre travail nous avons entraîné un modèle acoustique sur 5h de parole, ces enregistrements correspondent à 6 locuteurs différents et portent sur divers sujets : politique, sport et news. Ce corpus a été construit par M.Belgacem et moi-même durant mon projet de fin d'étude.

3.3. Alignement forcé :

Ce traitement consiste à aligner les signaux de parole avec les transcriptions

correspondantes, repérer la description phonétique du signal dans le dictionnaire phonétique, repérer les frontières phonémiques dans l'enregistrement et affecter à chaque phonème un score suivant la distance qui le sépare de la norme qui est le modèle acoustique.

Nous avons réalisé l'alignement grâce à l'outil sphinx-align (dans la boîte sphinx3) qui nous a permis d'aligner le signal de parole avec la transcription. Pour cela, il suffit de lui donner les chemins vers les données nécessaires qui sont les signaux sous format MFCC, les transcriptions, le dictionnaire phonétique et le modèle acoustique.

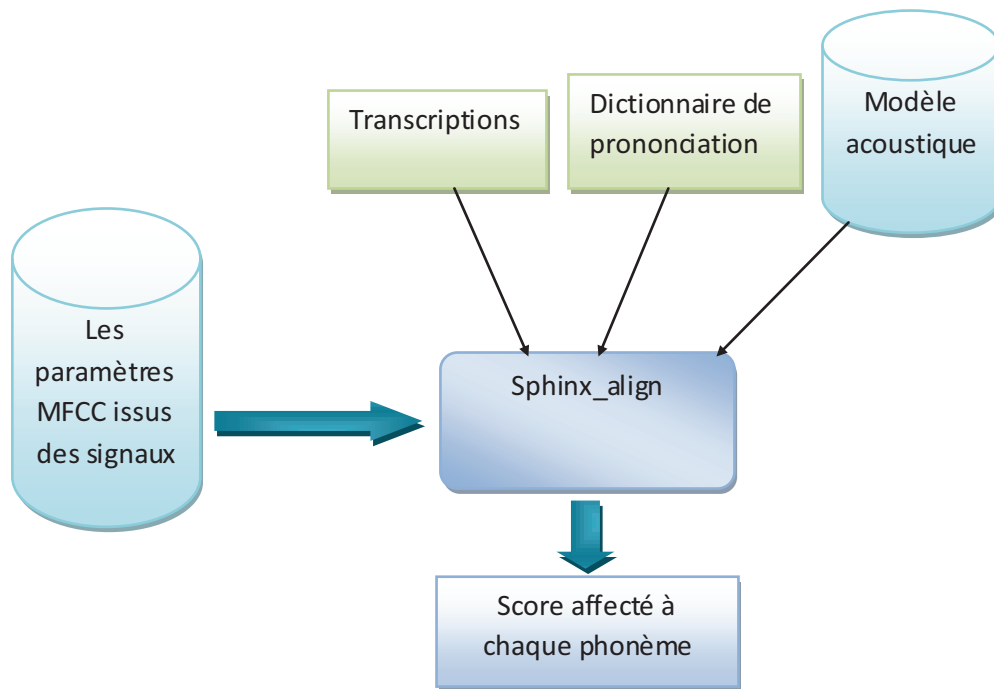


Figure 2 : Processus d'alignement forcé

Une fois l'alignement forcé terminé nous obtenons un corpus segmenté en phonèmes avec les fichiers résultats qui contiennent le temps de début et fin pour chacun en nombre de trames et le score qui lui a été affecté, on trouve à la fin du fichier la somme de tous les scores pour chaque signal.

3.4. Le décodage acoustico-phonétique :

Le décodage acoustico-phonétique (DAP) est le processus qui fournit à partir d'un signal de parole la séquence de phonèmes la plus probable.

La sortie du décodeur, comme l'alignement forcé, se base sur les vecteurs de paramètres extraits à partir du signal et les hypothèses émises par le modèle acoustique et le modèle de langage (lorsque il y en a un).

Dans notre travail, nous nous intéressons à détecter les erreurs dues à des problèmes de

prononciation, pour cette raison nous avons utilisé un modèle de langage plat qui émet des hypothèses équiprobables pour les phonèmes, ce qui signifie que les sorties du décodage acoustico-phonétique ne dépendent que de la modélisation des phonèmes dans le modèle acoustique.

Le décodeur ne traite pas les signaux sous format WAV il faut d'abord extraire les paramètres MFCC. De cette façon les éléments nécessaires pour le décodage sont :

- Les paramètres MFCC
- Le Modèle acoustique
- Le modèle de langage plat

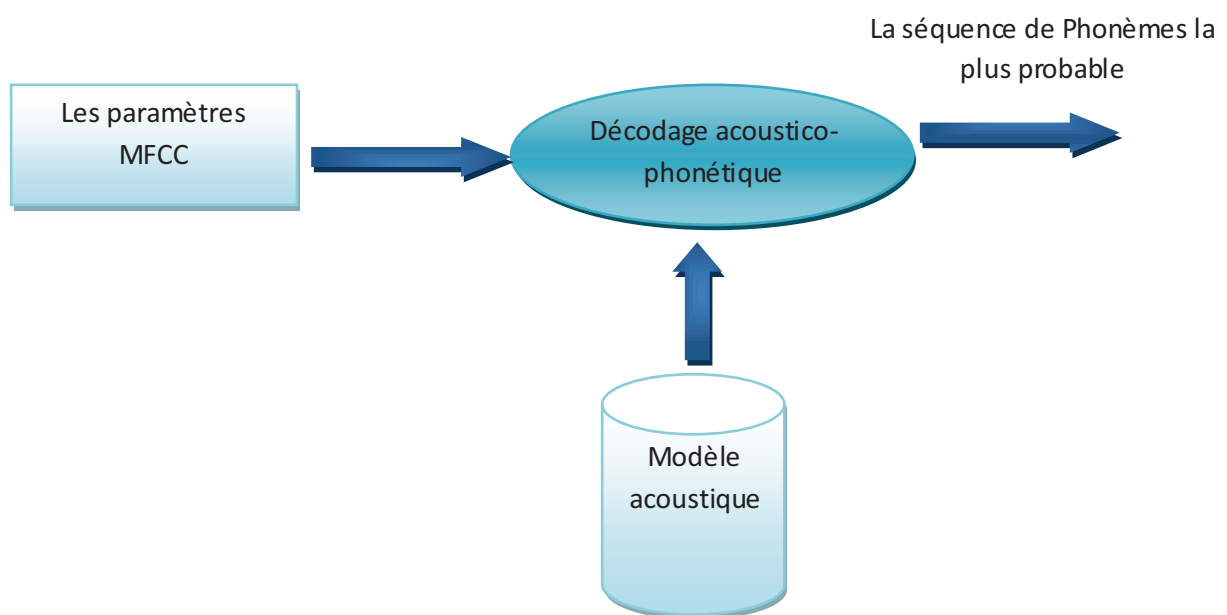


Figure 3 : Décodage acoustico-phonétique

4. Conclusion

Dans ce chapitre, nous avons présenté notre corpus de travail, les différents niveaux de langue que nous avons traités ainsi que les outils automatiques utilisés. Ces données sont importantes pour notre étude. Elles nous permettront d'analyser le comportement de chaque groupe de locuteurs et leurs caractéristiques.

Chapitre 3 : Analyses des résultats

1. Introduction :

Dans le chapitre précédent nous avons décrit notre corpus de travail et les outils que nous avons utilisés. Dans ce chapitre nous allons présenter et analyser les résultats de l'alignement forcé et du DAP. Tout d'abord, nous allons voir si le comportement des scores peut donner des indications significatives sur le niveau de langue, ensuite nous allons regarder en détail les scores pour chaque phonème afin de voir quelles sont les phonèmes qui posent plus particulièrement des problèmes aux apprenants de l'arabe et les confusions qui sont le plus souvent observées chez les apprenants.

2. Construction du modèle acoustique :

Pour pouvoir entraîner notre modèle acoustique nous avons 5h d'enregistrement de parole native arabe et les transcriptions correspondantes, nous avons aussi le dictionnaire phonétique qui contient la transcription en phonèmes de tous les mots du corpus ; le tableau ci dessous présente les notations utilisées dans le dictionnaire et les symboles API correspondant.

Notations utilisées	Symboles API
AE	[a]
AE :	[a :]
IH	[i]
IY	[i :]
UH	[u]
UW	[u :]
F	[f]
DH	[ð]
TH	[θ]
S	[s]
AI	[ʕ]
H	[h]
HH	[ħ]
GH	[ɣ]
B	[B]
T	[T]
D	[D]

K	[K]
SH	[ʃ]
KH	[x]
Z	[z]
Q	[q]
E	[ʔ]
J	[ɕ]
R	[r]
Y	[j]
W	[w]
L	[l]
EL	[ʔl]
M	[m]
N	[n]
TT	[tʰ]
SS	[sʰ]
DD	[dʰ]
DH2	[ðʰ]

Tableau 5 : Table de correspondance entre les symboles utilisés et les symboles API

Avant de lancer l'entraînement du modèle acoustique, une phase de paramétrisation est indispensable pour faire les analyses acoustiques et extraire les caractéristiques de la parole. Pendant la phase de paramétrisation, des transformations de Fourier rapides et des bans de filtres sont appliqués au signal pour obtenir une suite de vecteurs acoustiques appelée MFCC (Mel-Frequency Cepstral Coefficients). Nous avons utilisé le module wave2feat de SphinxTrain pour extraire les paramètres MFCC.

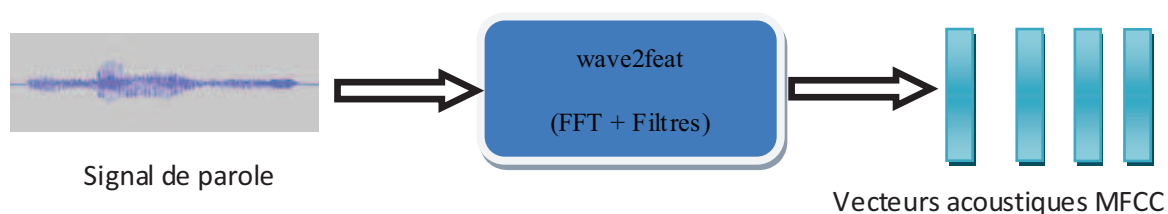


Figure 4 : Paramétrisation du signal

Une fois la phase de paramétrisation terminée nous pouvons alors construire notre modèle acoustique et lancer l'entraînement à l'aide de l'outil SphinxTrain. Pour cela il nous faut les données suivantes :

- Les paramètres acoustiques MFCC obtenus de la phase de paramétrisation.
- Les transcriptions.
- La liste des phonèmes.
- Le dictionnaire de prononciation.
- La liste des caractères qui ne représentent pas de la parole et qui existent dans les transcriptions.

La figure 5 présente les différents fichiers d'entrée et de sortie en utilisant SphinxTrain pour la construction du modèle acoustique.

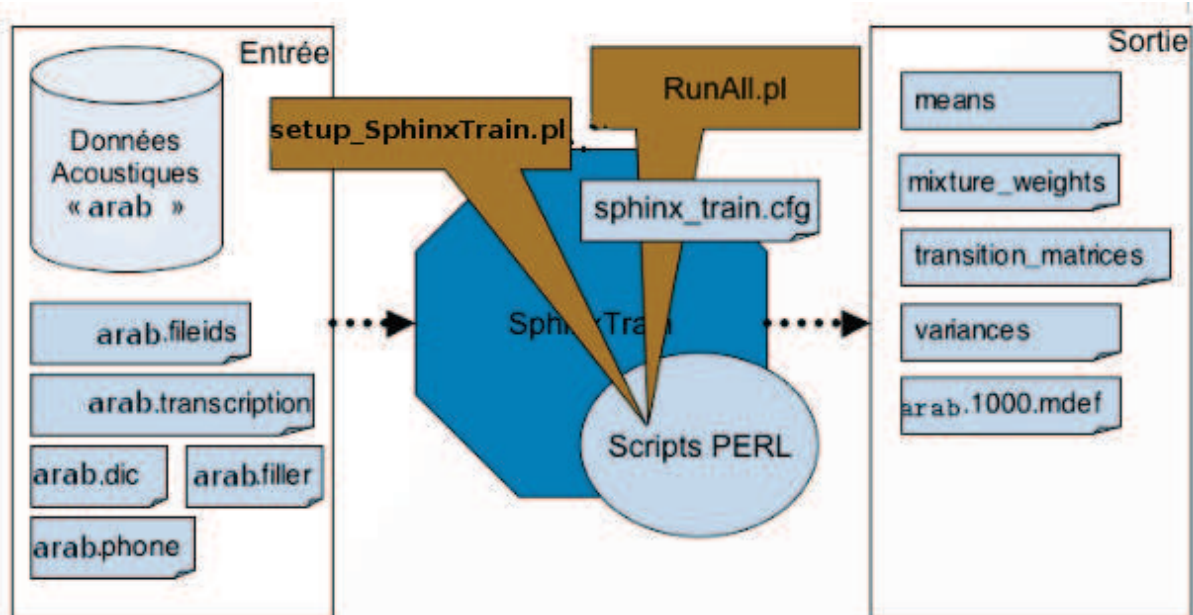


Figure 5 : Les fichiers d'entrer et de sortie du module SphinxTrain

Les fichiers : (means, mixture_weight, transition_matrices, variances, arab.mdef) obtenus constituent notre modèle acoustique.

Dans le cas du modèle acoustique classique que nous avons utilisé pour dans les premières mesures, un phonème est modélisé par un HMM à trois états représentant ses début, milieu et fin (Voir exemple de modélisation du phonème N dans le mot « E IH N AE »).

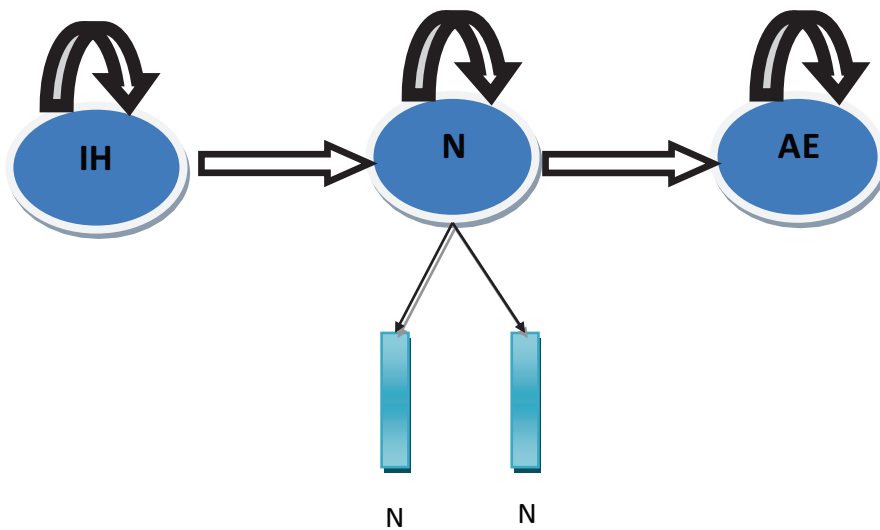


Figure 6 : modélisation du phonème N dans le mot "E IH N AE"

Les phonèmes encadrant le phonème à modéliser, dans notre cas le N, ne sont pas modélisés entièrement, seules les parties de transitions (passage d'un phonème à un autre) sont représentées.

Le modèle acoustique construit constitue une donnée importante pour notre travail sur laquelle se base l'alignement forcé et le décodage des phonèmes.

3. Analyse des scores globaux obtenus suite à l'alignement forcé :

Pour commencer nous allons regarder les résultats de l'alignement forcé et essayer de les analyser.

A l'issue de l'alignement forcé les fichiers générés affichent le score affecté à chaque phonème ainsi que la somme des scores de tous les phonèmes. Cependant, analyser la somme des scores n'est pas une tâche intéressante car le score global dépend de la longueur du mot. En effet plus le nombre des phonèmes dans le mot est élevé plus la somme sera importante. Il est donc plus intéressant de regarder la moyenne des scores à la place de la somme. A l'aide d'un script Perl nous avons donc extrait le score moyen pour chaque enregistrement.

3.1. Analyse du comportement des scores selon le niveau de langue :

Afin de pouvoir comparer le comportement des scores selon le niveau de langue nous avons extrait les scores pour chaque catégorie de locuteur (niveau de langue) séparément, ensuite nous avons trié les scores et enfin, nous avons tracé sur la Figure7 qui présente la répartition de tous les scores obtenus à l'issue de l'alignement forcé suivant le niveau de langue des locuteurs.

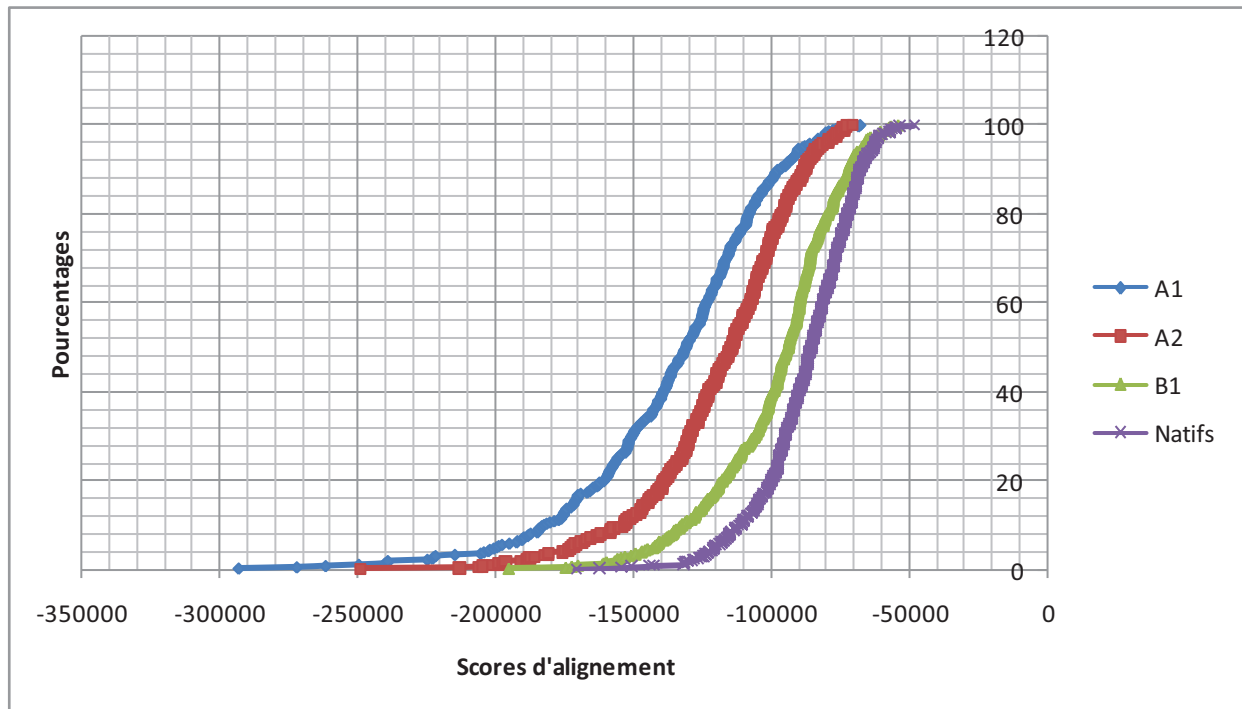


Figure 7: Comportement des scores d'alignement forcé selon les niveaux de langue

Chaque point de la courbe représente un score moyen pour un locuteur donné obtenu par l'alignement forcé pour un enregistrement (prononciation d'un mot). L'abscisse renseigne sur le score obtenu en ce point et l'ordonnée sur le pourcentage des scores inférieurs ou égaux au score obtenu en ce point.

Nous pouvons facilement voir qu'il y a une différence entre le comportement des scores pour des locuteurs de niveau de langue différent. Nous remarquons aussi que plus le niveau de langue des apprenants est avancé plus les scores se rapprochent de ceux des natifs.

Nous pouvons aussi remarquer à partir de la forme des courbes obtenues que les courbes des A1 et des A2 ont un comportement similaire alors que le comportement de la courbe des B1 qui ont un niveau de langue plus avancé se rapproche plus du comportement des Natifs.

Pour mieux analyser la relation entre le niveau de langue et le score acoustique nous avons effectué des tests statistiques : ANOVA et correction de Bonferroni en utilisant le logiciel « R ».

- ANOVA : « ANlysis Of VAriaence » ou analyse de la variance : c'est un test statistique utilisé pour mesurer le degré de dépendance d'une variable par rapport à une ou plusieurs autres variables. (wiki, 2011).
- Correction de Bonferroni : c'est un test statistique utilisé pour faire des comparaisons multiples dans le but de vérifier la dépendance ou l'indépendance de n hypothèse par rapport à un ensemble de données.
- P : renseigne sur la probabilité de tomber sur l'hypothèse nulle, elle doit être inférieure à un seuil de 0.01 pour rejeter l'hypothèse nulle et dire que nous avons une différence significative entre des populations.

Nous avons appliqué l'ANOVA sur l'ensemble de nos scores d'alignement groupés selon le niveau de langue du locuteur et nous avons obtenu une valeur du « P » $\ll 0.01$ ce qui signifie que les scores n'appartiennent pas à la même population, dans notre cas, les scores n'appartiennent pas à des locuteurs ayant le même niveau de langue. Toutefois on ne peut pas dire que la différence sera significative si on compare les scores d'un groupe par rapport à un autre.

Les résultats du test statistique confirment donc notre hypothèse : les scores d'alignement sont différents selon le niveau de langue du locuteur.

3.2. Analyse de distribution des scores selon le niveau de langue :

A travers cette analyse nous voulons trouver un moyen d'évaluer la prononciation des locuteurs et voir où se concentrent les scores pour chaque catégorie d'apprenants. Afin de faire cette analyse nous avons pris le plus mauvais et le meilleur score obtenu pour nos locuteurs (tous niveaux inclus) et nous avons découpé l'intervalle en tranches de scores égales (10000) et pour chaque tranche nous avons calculé le pourcentage de scores qui se trouve dans cette partie de la courbe.

Comme nous avons expliqué dans la méthodologie, le corpus se décompose en deux sous corpus : corpusA1 et corpusA2B1 nous allons voir le comportement des scores en analysant tout d'abord les scores obtenus par les locuteurs en prononçant les mots se trouvant dans chaque sous corpus séparément ensuite nous allons voir la distribution des scores sur l'intégralité du corpus.

Sur la figure 8 ci-dessous nous pouvons observer les résultats pour le premier sous corpus : le corpusA1. Nous rappelons que ce sous-corpus comprend des locuteurs de tous les niveaux de langue.

Pour plus d'information concernant le nombre de locuteur et le nombre de mots par corpus consulter les tableaux 3 et 4.

a. corpusA1 :

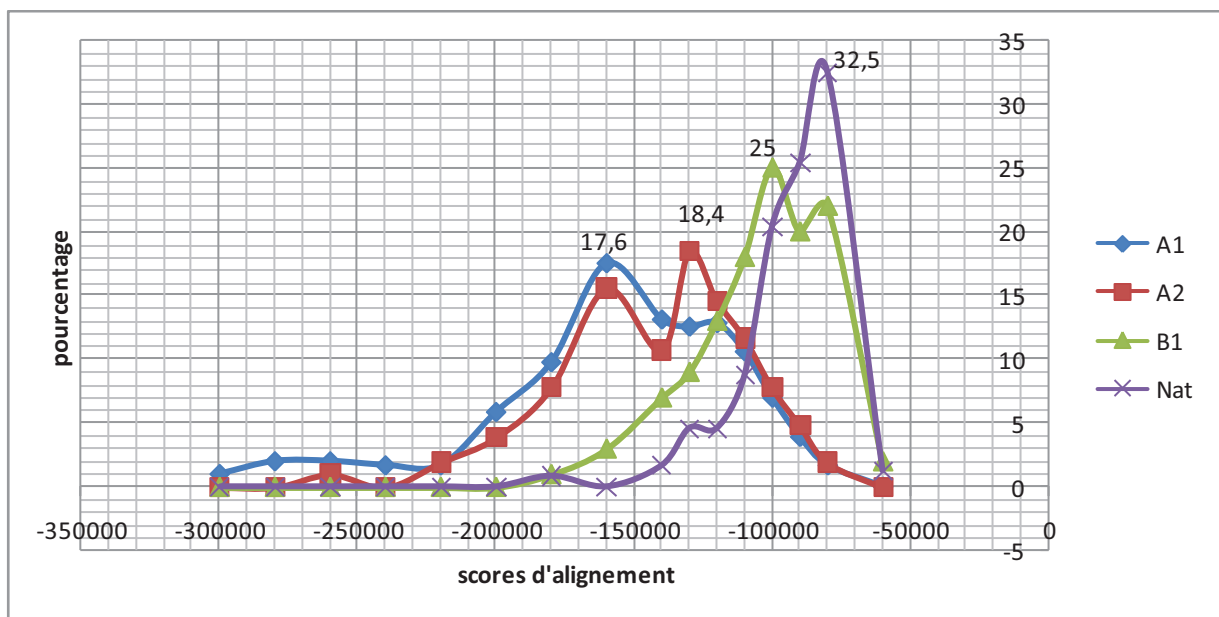


Figure 8:distribution des scores d'alignement forcé dans le corpusA1

Nous pouvons clairement voir sur cette figure que plus le niveau de langue est avancé, plus la distribution des scores se rapproche de celle des natifs. La courbe des locuteurs de niveau B1 est la plus proche de celle des natifs, car les mots prononcés dans ce premier sous corpus sont appris dans le cours de niveau A1. Ils sont donc plus faciles à prononcer que les autres mots. Si la distribution des scores des apprenants de niveau B1 est très proche de celle des natifs alors que les courbes obtenues pour les locuteurs de niveau A1 et A2 sont très différents de celle des natifs. Ces résultats montrent qu'il y a encore un écart important entre leur prononciation et la prononciation native des mots.

Là aussi nous pouvons voir une similarité dans la distribution des scores pour les A1 et les A2 et une distribution plus proche à celle des Natifs pour les B1.

Pour vérifier que les scores sont différents en passant d'un niveau de langue à un autre. Nous avons appliqué une correction de Bonferroni sur les scores d'alignement.

Les résultats du test sont représentés dans une matrice où chaque valeur représente le « P » obtenu en comparant les scores des locuteurs ayant le niveau de langue de la ligne avec les scores des locuteurs ayant le niveau de langue de la colonne.

$$\begin{pmatrix} & A1 & A2 & B1 \\ A2 & 0,282 & & \\ B1 & < 0,01 & < 0,01 & \\ Natifs & < 0,01 & < 0,01 & < 0,01 \end{pmatrix}$$

Le résultat de la correction de Bonferroni confirme nos hypothèses sur les scores d'alignement qui sont distincts en passant d'un niveau de langue à un autre. Mais on remarque qu'il n'y a pas de différence significative entre les scores obtenus pour les A1 et les scores des A2 pour ce sous-corpus, c'est peut être par ce que on n'a pas suffisamment de données pour les locuteurs de niveau A2 dans cette partie du corpus.

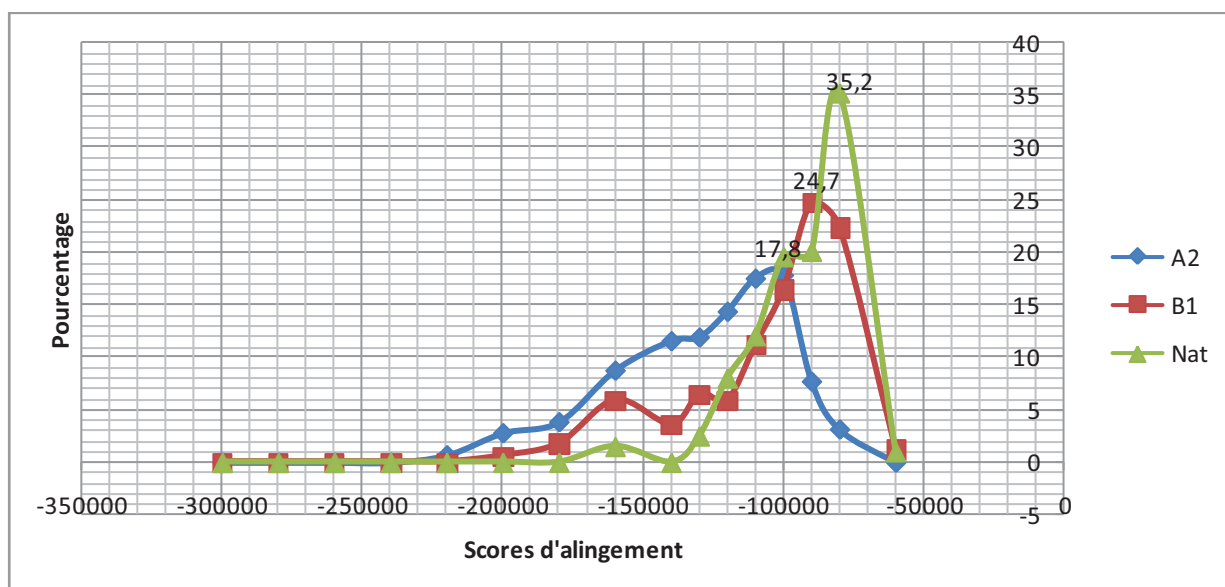


Figure 9: distribution des scores d'alignement forcé dans le corpus A2B1

Les apprenants de niveau A1 n'ont pas prononcé les mots de ce sous corpus il y a seulement les locuteurs de niveau A2, B1 et les natifs. Les mots constituant ce corpus sont plus difficiles que ceux dans le corpus A1. Si nous comparons les résultats affichés sur la figure 9 et ceux de la figure 8 nous pouvons remarquer que la courbe des natifs a gardés presque la même allure avec un pic vers -80000, nous pouvons remarquer aussi que la distribution des scores pour les niveaux A2 et B1 :

- comme pour le corpus précédent la courbe des B1 est proche de celle des natifs.
- pour les A2 on remarque un pic à -100,000 meilleur que sur le corpusA1, peut être, parce que les mots sont appris récemment ou parce qu'on a plus de données pour les locuteurs de niveau A2 dans le corpusA2B1.

La matrice obtenue en appliquant la correction de Bonféroni :

$$\begin{pmatrix} & A2 & B1 \\ B1 & < 0.01 & \\ Natifs & < 0.01 & < 0.01 \end{pmatrix}$$

Les résultats du test statistique affichent des valeurs très inférieures à 0.01, ça signifie que les scores comparés n'appartiennent pas à la même population d'apprenants, ce qui renforce nos hypothèses : les scores d'alignements diffèrent selon le niveau de langue du locuteur.

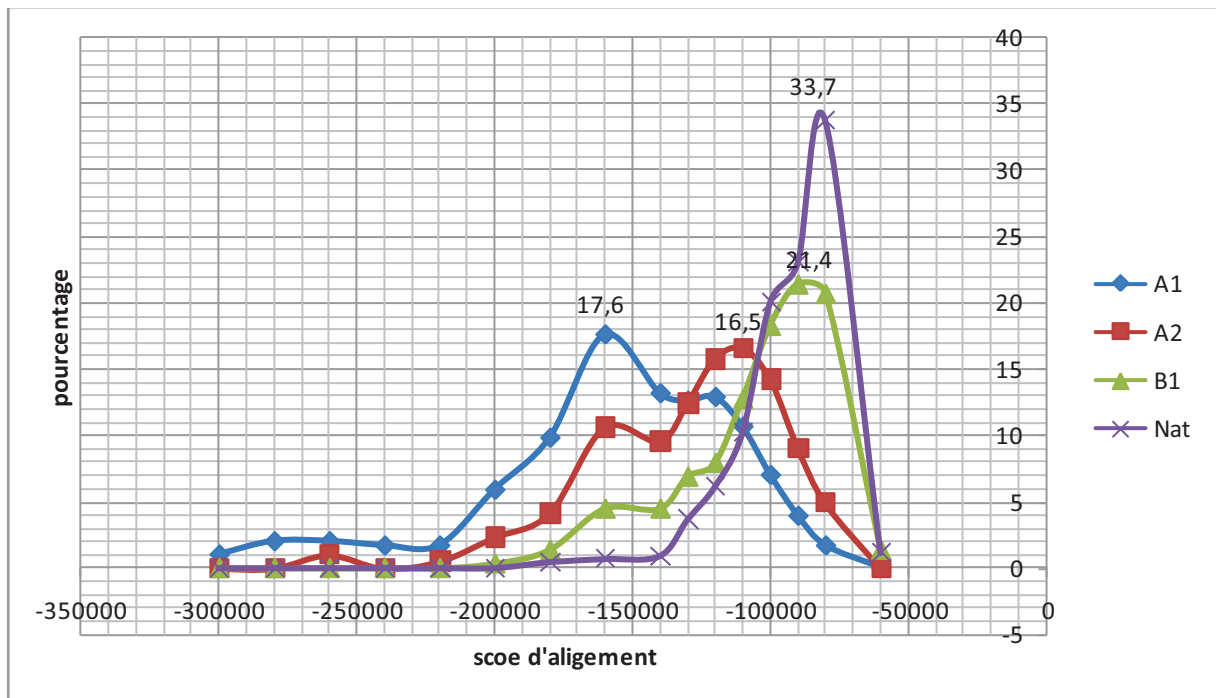


Figure 10: la distribution des scores d'alignement forcé sur tout le corpus

Sur la Figure 10, nous pouvons remarquer facilement la différence de distribution des scores pour des apprenants ayant des niveaux de langue différents, les courbes des apprenants se rapprochent de celle des natifs lorsque le niveau de langue augmente. L'analyse des scores relevés sur tout le corpus figure 10, nous permet de dire que le niveau de langue a une

influence sur les scores d'alignement forcé donc une influence sur les outils de Reconnaissance de la parole et par suite les scores de l'alignement forcé peuvent donc être un indice qui renseigne sur le niveau de langue du locuteur.

$$\begin{pmatrix} & A1 & A2 & B1 \\ A2 & < 0.01 & & \\ B1 & < 0.01 & < 0.01 & \\ Natifs & < 0.01 & < 0.01 & < 0.01 \end{pmatrix}$$

Les résultats de la correction de Bonferroni appliquée sur tout le corpus renforcent nos hypothèses et confirment que si nous avons suffisamment de données pour les locuteurs, les scores d'alignement seront différents en passant d'un niveau de langue à un autre et peuvent donner des informations renseignant sur le niveau de langue du locuteur.

3.3. Evaluation de la prononciation des apprenants :

Afin d'estimer la qualité de prononciation des locuteurs nous allons faire une évaluation des apprenants par rapport aux locuteurs natifs, pour cela nous avons suivi une approche qui permet l'estimation du niveau de langue d'une population d'apprenants en la comparant à une référence, dans notre cas les natifs. Cette approche consiste donc à calculer la moyenne des scores des locuteurs natifs et l'écart type « σ », pour ensuite diviser notre plage de scores en trois catégories :

- Bonne prononciation : **scores > moyenne - « σ ».**
- Prononciation moyenne : **moyenne - 2« σ » < scores < moyenne - « σ ».**
- Mauvaise prononciation : **scores < moyenne - 2 « σ ».**

D'après nos données, le score moyen pour les mots prononcés par les locuteurs natifs est de -87704 avec un écart type de -17701.

Pour chaque groupe de locuteur (niveau de langue) on calcule le pourcentage des scores cumulés dans les intervalles définis. La figure11 affiche les histogrammes obtenus.

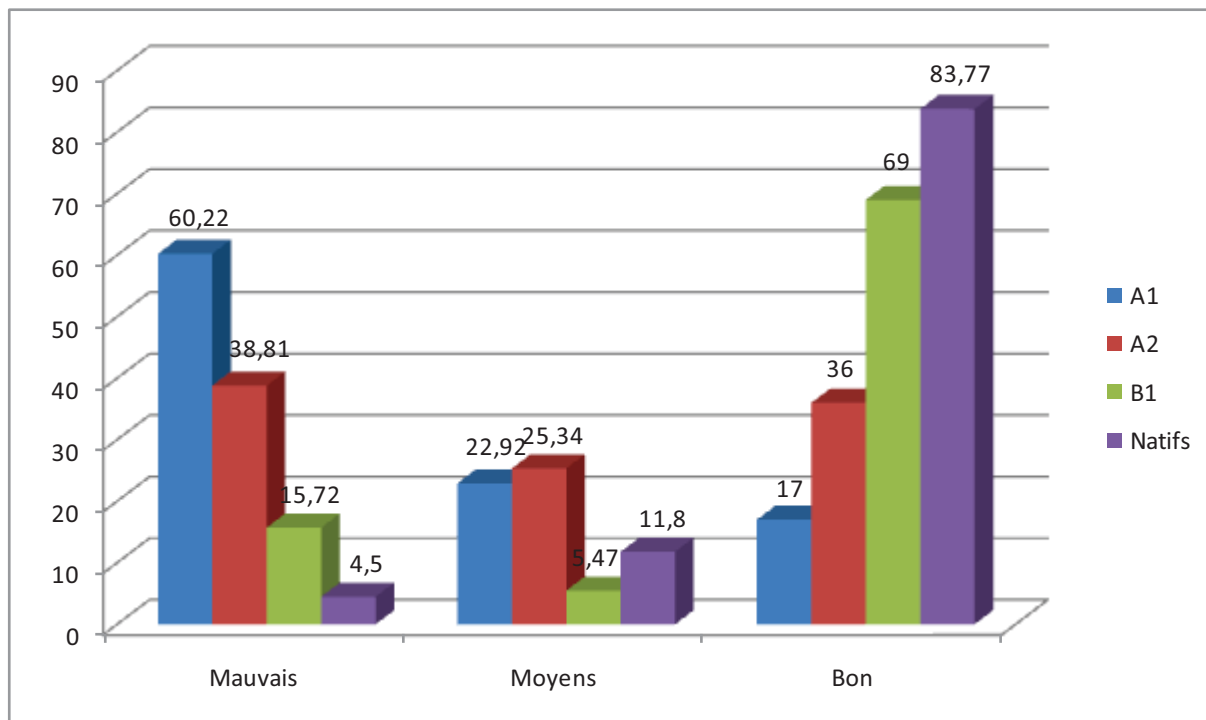


Figure 11: Evaluation de la prononciation selon les scores acoustique

Nous pouvons voir clairement que la répartition des scores des apprenants suit bien leurs niveaux de langue respectifs, et que l'écart entre les différents niveaux est facilement observable. Les résultats du test d'évaluation du niveau de langue viennent confirmer les résultats présentés précédemment et nous pouvons dire que le niveau de langue a une influence sur les scores de l'alignement forcé et ceux-ci peuvent donc être un indice pour l'estimation du niveau de langue des apprenants.

4. Analyse en fonction des types de phonèmes :

Nous avons vu que les scores d'alignement sont différents d'un niveau de langue à un autre et maintenant nous voulons analyser les scores pour les phonèmes pour voir si le score d'alignement peut donner un indice sur les phonèmes qui posent problème pour les apprenants.

Dans cette analyse, nous avons reparti les phonèmes de l'arabe en sept catégories selon les difficultés recensées dans le corpus comme suit :

- Voyelles courtes : ([i] ; [u] ; [a]) sont au nombre de trois et font parties des voyelles de la langue française.
- Voyelles longues : ([i:] ; [u:] ; [a:]) en français les voyelles longues ne sont pas très fréquentes, elles sont souvent utilisées en fin de mot.

- Pharyngales : ([ʕ] ; [ħ]) elles sont articulées avec la racine de la langue et le pharynx et sont susceptibles de causer des difficultés de prononciation pour les apprenants.
- Le Hamza [ʔ] : est une consonne glottale son lieu d'articulation est la glotte. Elle est reconnue en phonétique internationale comme une occlusive glottale et existe en allemand et en maltais.
- Des sons se trouvant dans d'autres langues européennes : ([ð] ; [ʁ] ; [x] ; [θ] ; [h] ; [r]) la difficulté dans l'apprentissage, peut dépendre également des autres (L2) langues acquises par le locuteur.
- Emphatiques : ([sˤ] ; [dˤ] ; [tˤ] ; [ðˤ]) ils sont marqués par plusieurs caractéristiques : recul de la racine de la langue, pharyngalisation et vélarisation. Ils sont proches de certains sons de la langue Française ([s] ; [d] ; [t] ; [ð]) et causent généralement des problèmes en perception et en production pour les apprenants de l'arabe.
- Sons non problématiques : existent dans les deux langues et ne causent normalement pas de problèmes pour les apprenants en général.

Nous avons défini ces catégories à l'aide de l'enseignant de l'arabe de l'université Stendhal.

La Figure 12 résume les scores obtenus pour les apprenants pour chaque une des catégories de phonèmes ci-dessus.

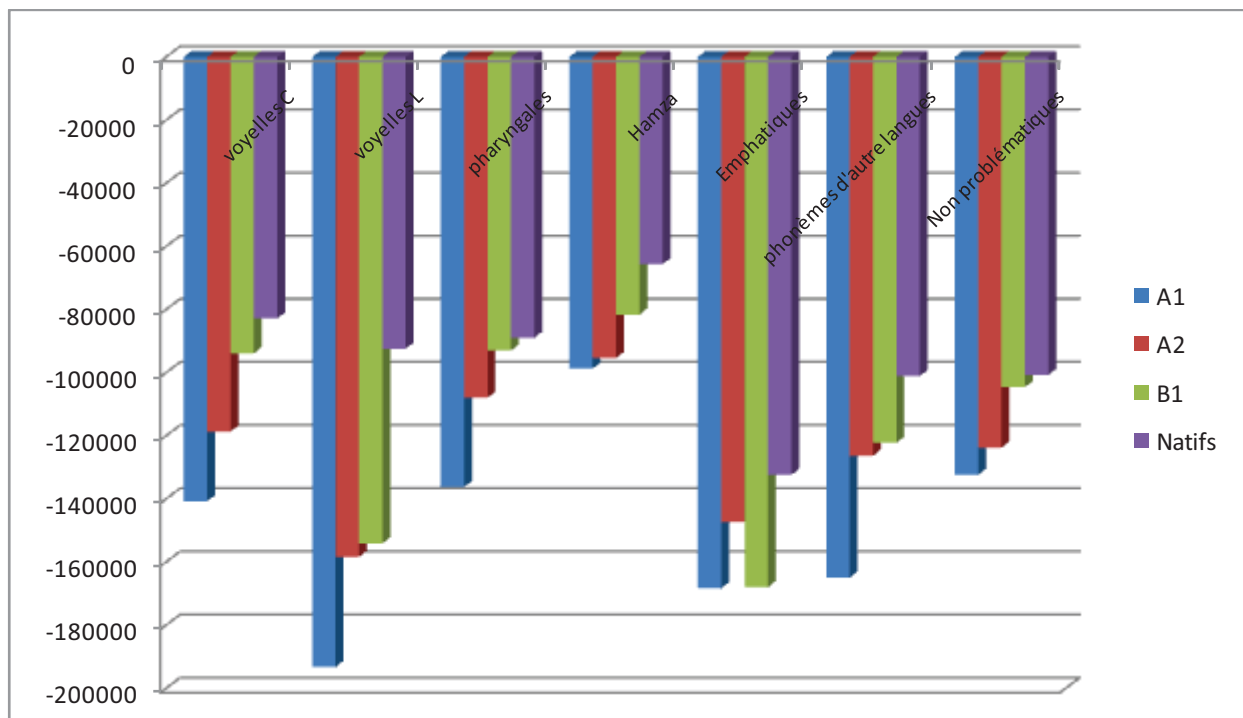


Figure 12: les scores des apprenants par catégorie phonétique

Sur la figure 12 on peut remarquer que les scores suivent le niveau de langue respectif des locuteurs pour toutes les catégories.

Pour évaluer la prononciation des apprenants et déterminer les sons qui leurs causent problèmes nous allons comparer leurs scores acoustiques pour chaque catégorie de phonèmes à celle des natifs, c'est l'écart qui déterminera le degré de difficulté pour les apprenants.

Nous avons divisé les voyelles en deux groupes les voyelles courtes et les voyelles longues. Les voyelles courtes de l'arabe existent dans la langue française, mais les apprenants peuvent être gênés par les nouveaux contextes des sons. En regardant les scores on peut voir que le score des B1 est proche de celui des natifs, ce qui veut dire que les voyelles courtes ne causent pas de problèmes pour cette catégorie d'apprenants. Pour les A1 et les A2 on remarque un écart assez important entre leurs scores et celui des natifs, ils sont possiblement gênés par les nouveaux contextes de ces phonèmes.

Les voyelles longues ne sont pas fréquentes dans la langue française et elles sont la plupart du temps prononcées en fin des mots, par contre dans l'arabe les voyelles longues sont très fréquentes et peuvent se trouver dans différents positions dans les mots. En regardant les scores, on voit un écart très important entre les scores des natifs et les scores apprenants. Si on compare cet écart avec celui des autres catégories phonétiques, on peut déduire que ce sont les voyelles longues qui causent le plus grand problème pour les apprenants. Ce qui veut dire que la durée des phonèmes est une caractéristique difficile à maîtriser pour les apprenants.

Les pharyngales [ħ] [ʕ] : elles sont essentiellement utilisées en nord-Afrique, au proche orient et en Colombie Britannique, elles ne sont pas présentes dans les langues européennes. En regardant les scores, les B1 et les Natifs ont presque les mêmes scores et que l'écart entre eux et les A2 n'est pas très important, on peut donc dire que les B1 et les A2 commencent à bien maîtriser ses nouveaux sons alors que les A1 ont plus de difficultés de prononciation. Ce résultat confirme l'hypothèse qui dit que les sons n'ayant pas des similarités avec les phonèmes de la langue maternelle des apprenants sont plus faciles à apprendre car cela engendre la création d'une nouvelle catégorie dans l'espace phonétique du locuteur.

La hamza [ʔ] : En regardant les scores, on peut voir que les scores des apprenants A1, A2 et B1 sont proches et l'écart avec les natifs n'est pas très important. Comme pour les pharyngales c'est un nouveau son qui n'a pas de similitude avec les sons de la langue native des apprenants ce qui facilite son apprentissage.

Les emphatiques [ʃˤ] ; [dˤ] ; [tˤ] ; [ðˤ] : En regardant les scores, on peut remarquer que les natifs ont obtenu des scores faibles pour cette catégorie, cela laisse dire qu'il y a un problème de modélisation pour ces phonèmes au niveau du modèle acoustique. Par contre si on regarde la différence entre les scores obtenus pour les natifs et les scores obtenus pour les apprenants on remarque qu'il y a un écart important qui permet de dire que les emphatiques posent problème pour les apprenants de l'arabe. Cela confirme l'hypothèse qui dit que les sons de la L2 ayant des similitudes avec des sons de la L1 sont les plus difficiles à maîtriser.

Les phonèmes [h] [x] [θ] [ð] [r] sont des sons qui n'existent pas en français mais qui peuvent exister dans d'autres langues européennes comme l'anglais et l'espagnol, les locuteurs peuvent maîtriser un ou plusieurs sons parmi ces phonèmes. En regardant l'écart entre les scores des apprenants et des natifs, on remarque que ces sons causent un grand problème pour les A1. Les A2 et les B1 ont presque le même score, leur écart avec les natifs est encore important, ils n'ont pas encore maîtrisé ces phonèmes mais ils ont des meilleures performances par rapport au A1.

Les sons non problématiques sont les consonnes communes entre les deux langues. Leur prononciation ne pose pas problème pour les apprenants en général. Sur le graphe on peut voir que le B1 ont obtenu un score très proche de celui des natifs, ces sons ne causent pas de problème pour eux. Par contre l'écart entre les scores obtenu par les A1, A2 et ceux obtenu par les Natifs est encore important, ils sont encore gênés par les nouveaux contextes de ces phonèmes.

5. Analyse des résultats pour un modèle acoustique « à gros grain » :

Le but de cette expérience est tout d'abord de montrer que les résultats obtenus ne sont pas liés aux spécificités des phonèmes et que l'écart observé entre les natifs et les apprenants n'est pas dû à des confusions entre les sons de la L2, mais à des erreurs de prononciations qui dépendent de leur maîtrise de la langue. Les erreurs peuvent être causés par une non maîtrise de la durée des phonèmes, des problèmes liés au mode ou au lieu d'articulation.

5.1. Construction du modèle acoustique :

Le choix des classes est un critère très important pour nos expériences qui nous aidera par la suite à analyser les problèmes des apprenants. Nous avons choisi de classer nos phonèmes selon le mode d'articulation. En phonétique articulatoire le mode d'articulation désigne un ensemble de propriétés :

- le type d'obstruction au passage de l'air.
- le canal vocal ou la cavité de résonance utilisée.
- l'organe qui pointe vers un lieu d'articulation.
- le type d'écoulement.
- le mécanisme à l'origine de l'écoulement.

A partir des 37 phonèmes au départ nous avons obtenu 9 classes. Le tableau 4 montre la répartition choisie.

Classes	Notation utilisée	Les Phonèmes de la classe
Voyelles	VO	[i] ; [u] ; [a] ; [i :] ; [u :] ; [a :]
Occlusives	OC	[b] ; [t] ; [d] ; [k] ; [q] ; [ʔ]
Fricatives	FR	[f] ; [s] ; [z] ; [ʃ] ; [ʒ] ; [θ] ; [x] ; [ʁ] ; [ʁ̥] ; [h] ; [ɦ]
Emphatiques	EM	[t ^ʕ] ; [s ^ʕ] ; [d ^ʕ] ; [ð ^ʕ]
Nasals	NA	[n] ; [m]
Affriquées	AF	[dʒ]
Latérales	LA	[l]
Spirantes	SP	[j] ; [w]

Tableau 6: les classes des phonèmes pour le modèle acoustique « gros grain »

Après avoir réparti nos phonèmes en classes nous avons changé les transcriptions phonétiques dans le dictionnaire en utilisant les nouvelles notations, ensuite nous avons relancé l'entraînement du nouveau modèle acoustique.

Le modèle acoustique « à gros grain » obtenu diffère du modèle classique dans les observations modélisées par les HMM. Pour mieux expliquer la différence reprenant l'exemple du phonème N dans le mot « E IH N AE » la phonétisation dans le nouveau dictionnaire est la suivante « OC VO NA VO ».

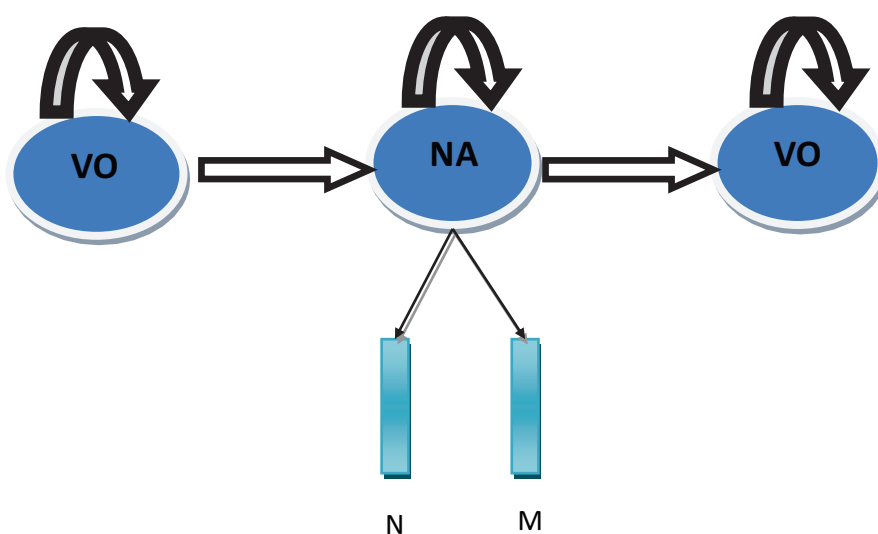


Figure 13 : Modélisation du N dans le modèle acoustique moyen

Comme vous pouvez le constater dans le cas du modèle acoustique « à gros grain », celui-ci ne fait pas de différence entre le M et le N. Ce n'est pas important car le but de cette expérience est de montrer que la différence entre la parole des apprenants et celles des natifs est causée plutôt par des problèmes de prononciation et pas par une différence d'accent ou de débit.

5.2. Comparaison avec les résultats du modèle classique :

Nous avons appliqué l'alignement forcé sur notre corpus en utilisant le modèle acoustique moyen, et à partir des scores obtenus nous avons appliqué les mêmes analyses que celles effectuées sur les scores obtenus en utilisant le modèle acoustique classique pour pouvoir faire la comparaison des résultats.

a. Comparaison des comportements des scores :

A partir des résultats de l'alignement forcé nous avons extrait les scores et nous avons appliqué les normalisations nécessaires (les scores moyens par mot et par phonème) en suite nous avons représenté les scores obtenus sur la figure14 ou chaque point est un score obtenu à partir de l'alignement forcé pour un enregistrement. Nous allons faire une comparaison de ces résultats avec ceux obtenus a partir de l'alignement forcé réalisé avec le modèle acoustique classique (ces résultats sont déjà représentés sur la Figure7).

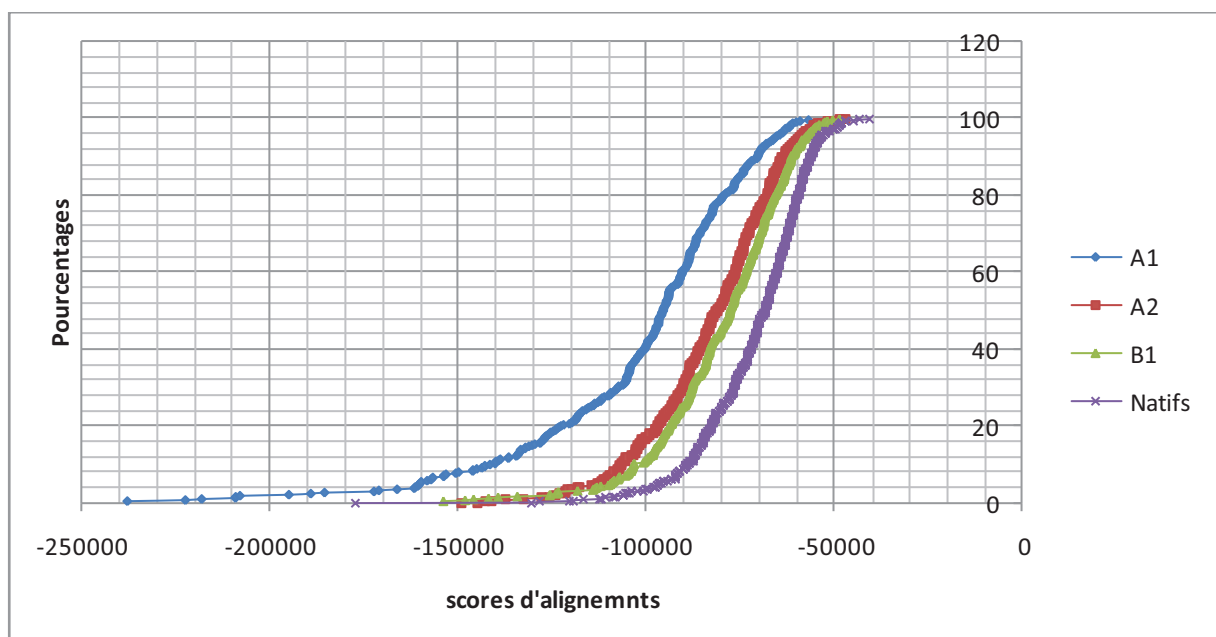


Figure 14 : scores d'alignement en utilisant le modèle acoustique « à gros grain »

En comparant les courbes obtenues avec celles de la Figure7 (obtenus pour le modèle acoustique classique) on remarque une amélioration des scores pour toutes les catégories

de locuteurs, cette amélioration est due à la largeur des classes des phonèmes dans le modèle moyen ce qui donne plus de tolérance et par la suite des scores meilleurs pour l'alignement forcé. Malgré cette tolérance on remarque que le comportement des scores continue à respecter le niveau de langue respectif des locuteurs.

Maintenant si on regarde le comportement des scores par rapport aux natifs qui sont notre référence, on remarque que pour les A2 l'écart avec les natifs s'est réduit et leur courbe devient très proche de celle des B1 ce qui montre qu'ils ont moins de problèmes liés au mode d'articulation que les A1, leur difficulté est plutôt liée à la durée des phonèmes et peut être au lieu d'articulation. Par contre pour les A1 on voit bien que leur écart avec les natifs est encore important ce qui nous permet de dire que les apprenants de niveau A1 ont des problèmes liés au mode d'articulation.

Nous avons aussi appliqué une ANOVA sur les scores obtenus à partir du modèle acoustique « à gros grain » et nous avons obtenu une valeur de « P » $\ll 0.01$ ce qui veut dire les scores d'alignement n'appartiennent pas à la même population de locuteurs. Et par suite on peut dire que même en utilisant un modèle acoustique qui manque de précision on arrive à détecter la différence de niveau de langue entre les locuteurs.

b. Comparaison de la distribution des scores :

Nous avons procédé de la même manière qui nous a permis de tracer la Figure6 : calcul du pourcentage des scores repartis dans des intervalles égaux (10000) se trouvant entre le meilleur et le plus mauvais score obtenu.

La Figure 15 représente les distributions classées par niveau de langue.

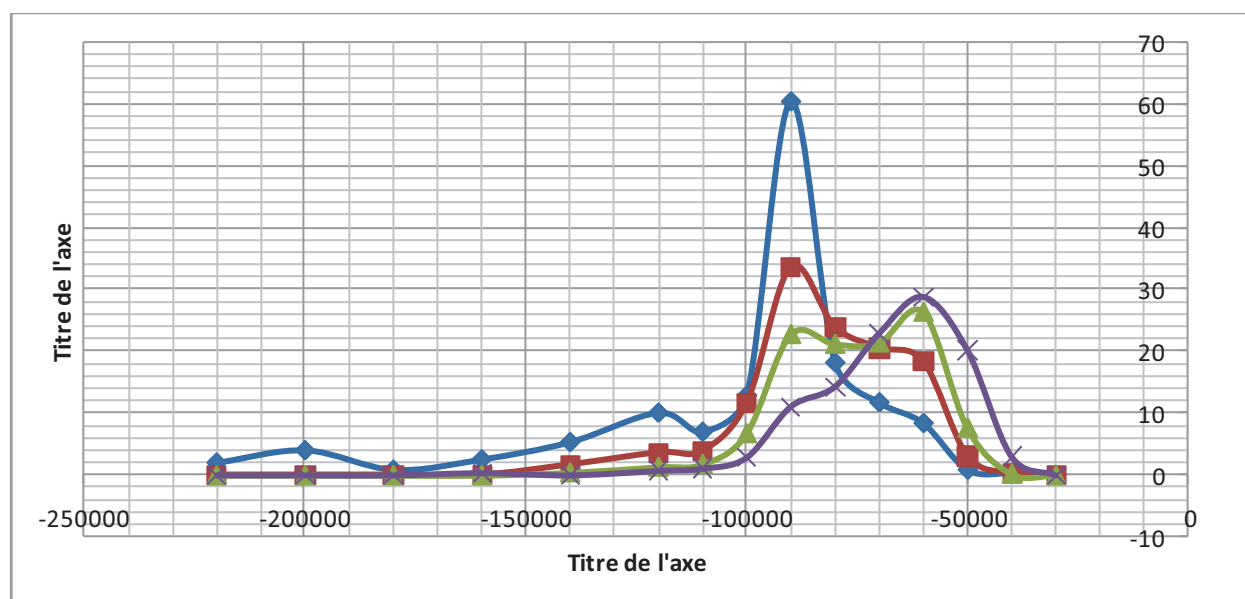


Figure 15 : Distribution des scores obtenus en utilisant le modèle acoustique moyen

Si on compare les résultats affichés sur la figure 15 avec ceux de la Figure10, on peut observer une amélioration globale des scores pour tous les locuteurs et d'une manière plus importante pour les apprenants, cette amélioration est expliquée par la tolérance qu'offre le modèle moyen (largeur des classes). Malgré le rapprochement des courbes de distribution sur ce graphe, on peut encore voir que la distribution continue à respecter le niveau de langue des locuteurs.

Nous avons appliqué une correction de Bonferroni sur les scores d'alignement obtenus en utilisant le modèle acoustique « à gros grain » les résultats sont affichés dans la matrice ci-dessous.

$$\begin{pmatrix} & A1 & A2 & B1 \\ A2 & < 0.01 & & \\ B1 & < 0.01 & 0.065 & \\ Natifs & < 0.01 & < 0.01 & < 0.01 \end{pmatrix}$$

Les résultats du test statistique indiquent que les scores ne sont pas homogènes et qu'il y a une différence significative entre les scores obtenus pour des locuteurs de niveau de langue différents même si nous utilisons un modèle acoustique « à gros grain » qui manque de précision. Seuls les scores des A2 et des B1 ne semblent pas séparables. Ce résultat confirme les hypothèses que nous avons tiré en analysant la figure 14, les scores des locuteurs de niveau A2 se sont améliorés remarquablement en utilisant le modèle acoustique « à gros grain » qui offre plus de tolérance pour les fautes liées au lieu d'articulation, ce qui laisse suggérer que les apprenants de niveau A2 ont des problèmes liés au lieu d'articulation.

Pour vérifier si on arrive encore à observer la différence de niveau de langue en utilisant le modèle acoustique « à gros grain », nous avons utilisé la même approche d'évaluation de niveau de langue utilisée dans la première expérience. La Figure 16 représente le diagramme obtenu.

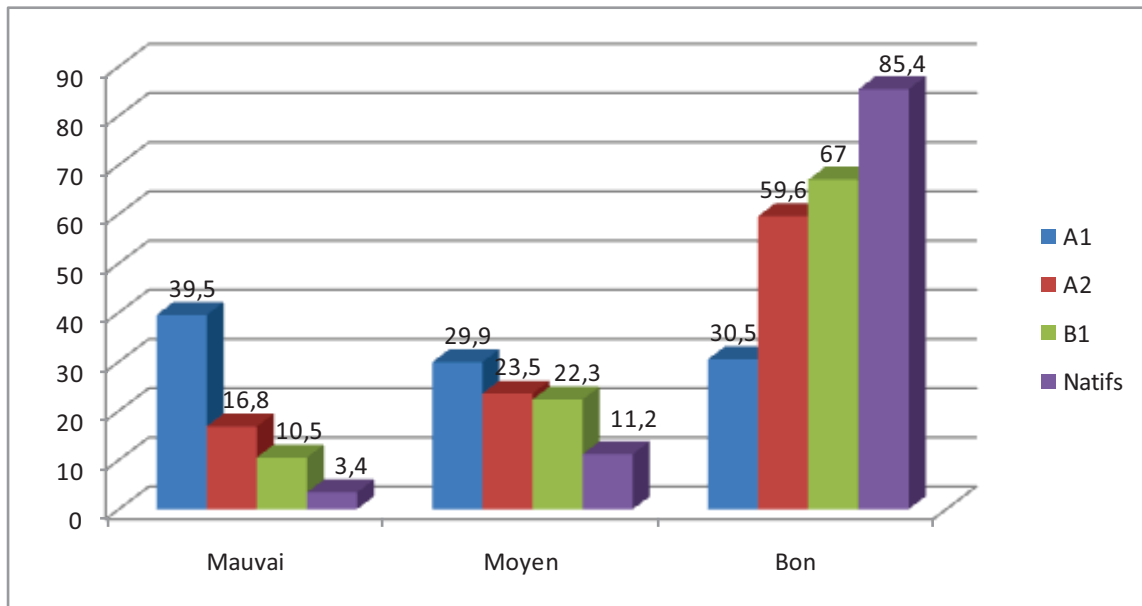


Figure 16 : évaluation des niveaux de prononciation

Si on compare ces résultats avec ceux de la figure11 on peut voir qu'il n'y a pas de gros changement pour les natifs et les B1, alors que pour les A1 et les A2 on peut observer des améliorations significatives. Ces améliorations sont du à la tolérance due modèle acoustique aux problèmes liés au lieu d'articulation. On pourra donc conclure que les A1 et A2 ont des difficultés liées au lieu d'articulation des phonèmes, alors que les B1 n'ont plus ce problème.

Sur la figure 16 : on peut observer aussi que le résultat de l'évaluation suit le niveau de langue respective des locuteurs. On peut donc conclure que les modèles acoustiques moyens peuvent être un « à gros grain » pour repérer les difficultés de prononciation des apprenants et peuvent aussi renseigner sur le niveau de prononciation mais avec moins de précision que les modèles acoustiques classiques.

6. Analyse des résultats du DAP :

A l'issu du décodage acoustico-phonétique, on obtient la transcription phonétique que le décodeur a reconnue à partir des enregistrements de parole. En comparant les sorties du DAP et celle de l'alignement forcé on peut déterminer le taux d'erreur de phonème (phone error rate PER) : celui-ci rassemble : le pourcentage d'insertions, le pourcentage de suppressions et le pourcentage de substitutions, qui sont des données très importantes pour notre analyse.

Dans la littérature on peut voir que le taux d'erreur pour le décodage acoustico-phonétique est souvent au tour de 50% ; voir par exemple les travaux de (Xiao, 2011) qui a trouvé un taux d'erreur de 52% pour le DAP. Comme nous l'avons expliqué dans la partie méthodologie, nous n'avons pas utilisé de modèle de langage durant le décodage, et c'est pour cette raison que nous avons obtenu des taux d'erreurs élevés. Le tableau montre le taux d'erreur obtenu pour les différentes catégories de locuteurs.

	A1	A2	B1	Natifs
Taux d'erreur	85	80	73	67

Tableau 7 : taux d'erreur de phonèmes pour chaque catégorie de locuteurs

6.1. Analyse du taux d'erreur pour les catégories phonétique :

A partir des sorties du DAP nous avons extrait le taux d'erreur pour chaque phonème ; ensuite nous avons groupé les phonèmes par catégorie. La Figure 17 montre le taux d'erreur calculé pour chaque catégorie.

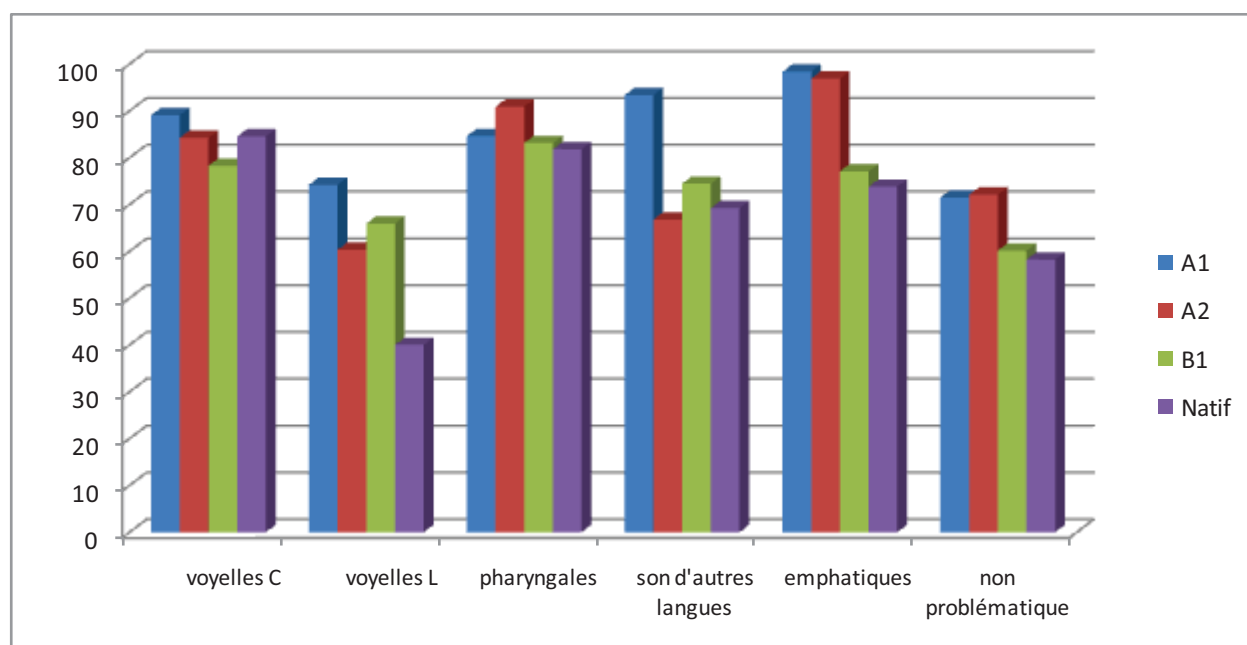


Figure 17 : taux d'erreur par phonème

A partir de la Figure 17, on peut remarquer que ce sont les voyelles longues qui ont le taux d'erreur le plus faible cela est dû au fait qu'elles sont plus stables grâce à leur allongement, ce qui les rendent plus faciles à reconnaître.

Pour pouvoir analyser nos données nous allons comparer les résultats des apprenants à ceux des natifs. On peut remarquer que pour les voyelles courtes et les pharyngales il n'y a pas de différence significative entre le taux d'erreur obtenu par les natifs et celui obtenu par les apprenants donc on ne peut rien conclure.

Pour les sons d'autres langues on remarque que l'écart entre les natifs et les A1 permet de dire qu'ils ont plus de problèmes que les autres locuteurs pour cette catégorie de phonèmes.

On remarque aussi que le taux d'erreur le plus élevé est obtenu pour les emphatiques et que l'écart entre les natifs et apprenants de niveau A1 et A2 est important. On peut alors dire que la prononciation des emphatiques leur cause un gros problème.

La classe des sons notée « non problématique » contient une grande diversité de phonèmes avec caractéristiques différentes et des taux d'erreur qui peut varier énormément d'un phonème à un autre.

6.2. Confusions :

Les confusions sont les substitutions observées pendant la phase de décodage, lorsque le décodeur reconnaît un phonème différent de celui qui se trouve à la même position dans la sortie de l'alignement forcé. Cette erreur peut provenir d'une erreur de prononciation due à des confusions entre les phonèmes dans l'espace phonémique du locuteur.

Pour extraire les confusions nous avons utilisé le logiciel SCLITE : c'est un logiciel libre et gratuit qui permet de faire des analyses et des statistiques sur les sorties du décodeur. L'extraction des confusions consiste à aligner des fichiers référence qui sont les sorties d'alignement forcé et qui correspondent à la suite des phonèmes décrivant le signal de parole aux fichiers hypothèse obtenus à la sortie du décodage. On déduit ainsi, les substitutions commises par les locuteurs. Ces substitutions peuvent être un indice sur les confusions des phonèmes chez les apprenants.

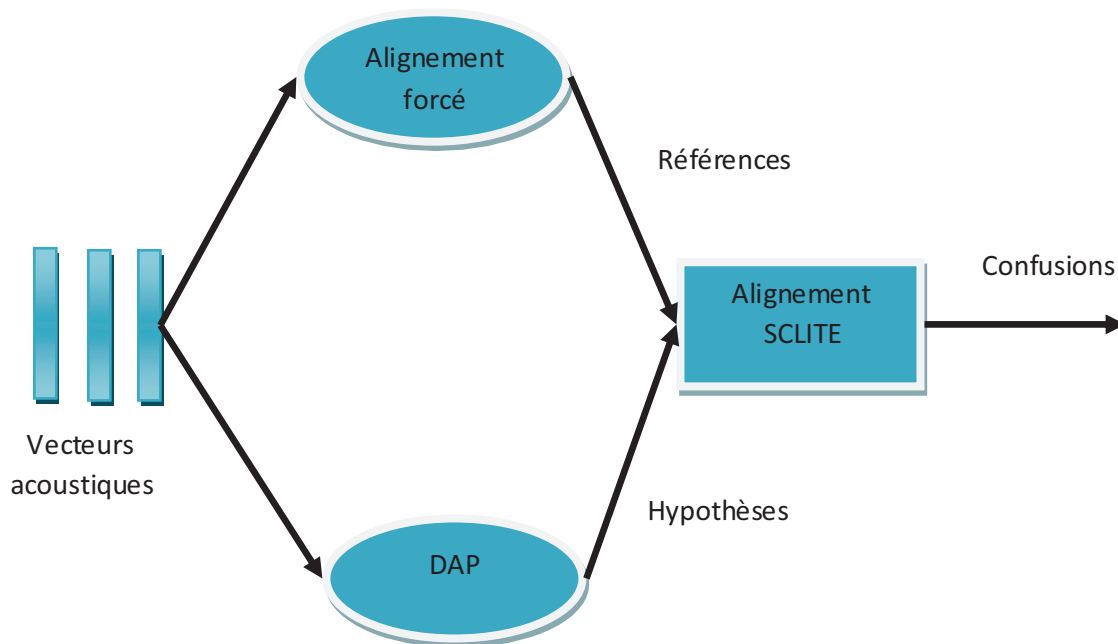


Figure 18 : Extraction des confusions avec SCLITE

Les sorties de SCLITE nous fournissent beaucoup de résultats statistiques sur la sortie du DAP concernant les suppressions, les insertions et les substitutions. Nous nous intéressons ici seulement aux résultats concernant les substitutions.

Exemples de sortie de SCLITE :

CONFUSION PAIRS Total (93)
With >= 1 occurrences (93)

- 1: 12 -> n ==> h
- 2: 7 -> r ==> h
- 3: 6 -> ai ==> h
- 4: 5 -> ae ==> h
- 5: 5 -> b ==> h
- 6: 5 -> ih ==> iy
- 7: 5 -> uh ==> uw
- 8: 4 -> ih ==> ae:
- 9: 4 -> l ==> h
- 10: 4 -> m ==> n
- 11: 3 -> ae ==> f

Les phonèmes sont classés selon le nombre de fois où ils ont été substitué par un certain phonème, la première colonne à gauche indique le rang, la deuxième indique le nombre de fois ou le phonème se trouvant avant la flèche (==>) a été remplacé par le phonème se trouvant après la flèche.

A partir des sorties de SCLITE nous avons calculé les statistiques pour chaque groupe de locuteur selon leur niveau de langue respective et pour normaliser nos résultats, nous avons transformé les résultats obtenus en pourcentage. Le pourcentage est calculé par rapport au nombre d'occurrences du phonème substitué dans les enregistrements prononcés par un groupe de locuteur ayant le même niveau de langue.

Nous allons afficher seulement les résultats des confusions qui sont supérieurs à un seuil de 1.5% ; pour plus de détail on consultera les matrices de confusion dans l'annexe.

Nous avons séparé les phonèmes confondus suivant la cause de la confusion. Sur la figure 15 on affiche les confusions causées par la différence de durée du phonème, pour l'arabe ce problème est surtout lié à l'opposition voyelles courtes voyelles / longues.

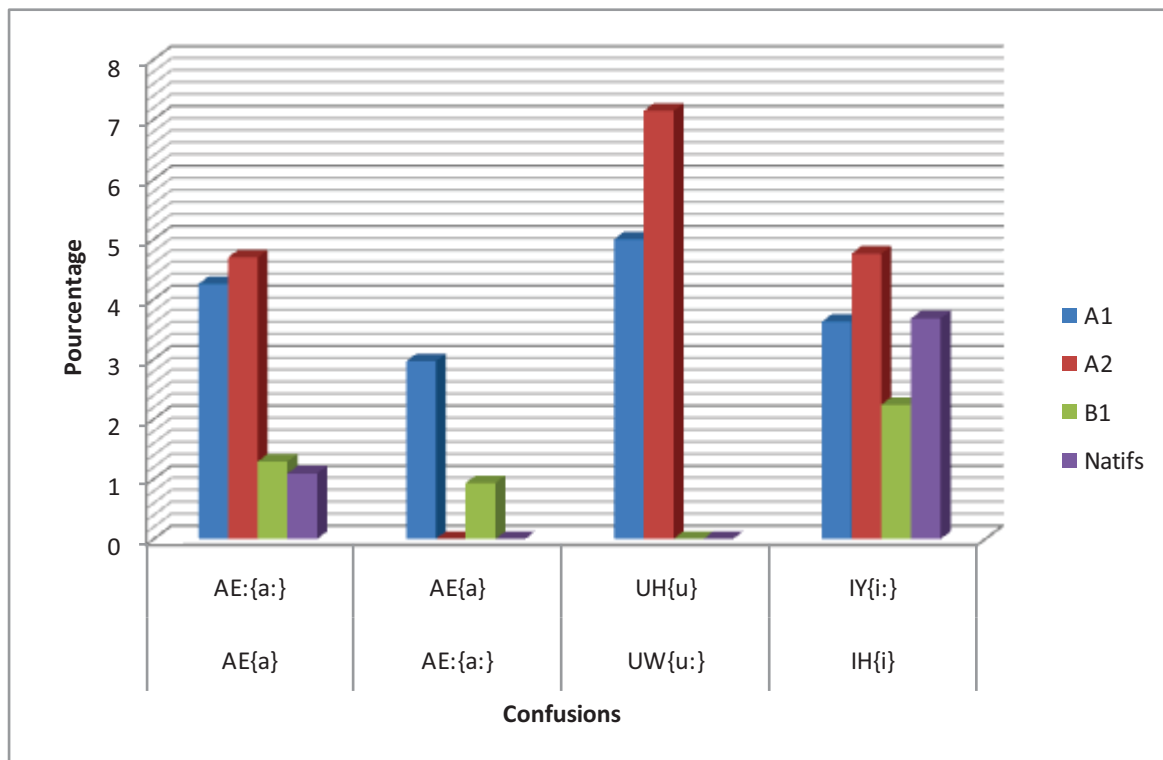


Figure 19 : confusions voyelles courtes voyelles longues

La modélisation de la différence entre voyelles courtes et voyelles longues est un problème souvent rencontré dans la reconnaissance automatique de la parole car les HMM ne modélisent pas bien la durée du phonème.

Sur la figure, les phonèmes substitués sont affichés en bas, alors qu'en haut on trouve les phonèmes insérés par confusion.

On peut voir que ce sont les A1 et les A2 qui font les plus de confusions entre les voyelles courtes et longues. Nous avons aussi vérifié les résultats sur les signaux et nous avons remarqué que les A1 et les A2 ne maîtrisent pas la différence de longueur entre les voyelles. Par contre l'analyse des confusions et la vérification manuelle des enregistrements montrent que les B1 ont beaucoup moins de problèmes liés à la différence « voyelles longues » et « voyelles courtes ».

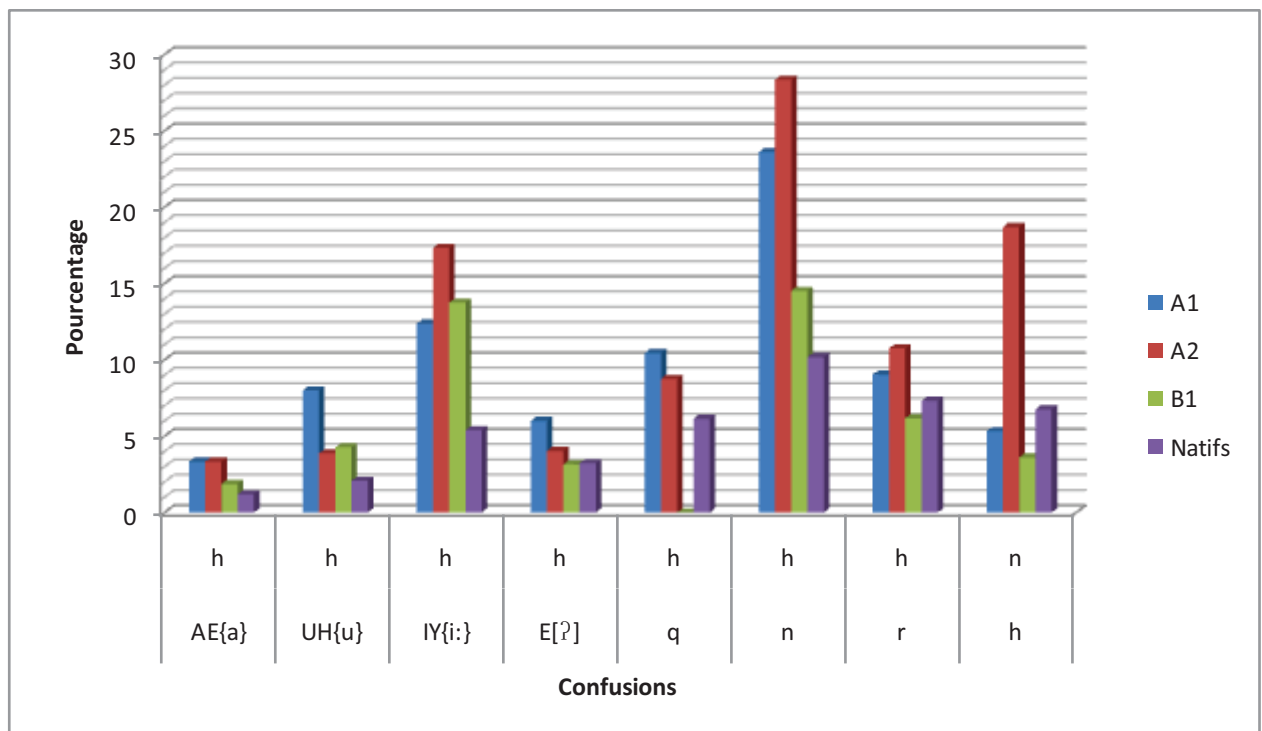


Figure 20 : le [h] un son qui attire tous les phonèmes

La lettre h existe dans l'alphabet du français mais elle n'est prononcée que rarement, toute fois elle a été introduite dans l'API du français. Dans la phonologie le [h] est une consonne glottale qui n'a pas de point d'articulation particulier, et il est prononcé dans les différents emplacements dans les mots comme une consonne ordinaire.

Nous avons remarqué que les différents phonèmes sont confondus avec [h] pour tous les locuteurs mais la confusion est plus importante dans le cas de la parole des apprenants. Dans la littérature, on trouve que les français apprenants l'anglais ont un problème avec la prononciation [h] et les enseignants disent que les élèves insèrent ce phonème dans les phrases même si il n'existe pas. Cependant, nous n'avons pas analysé ce phonème plus en détail.

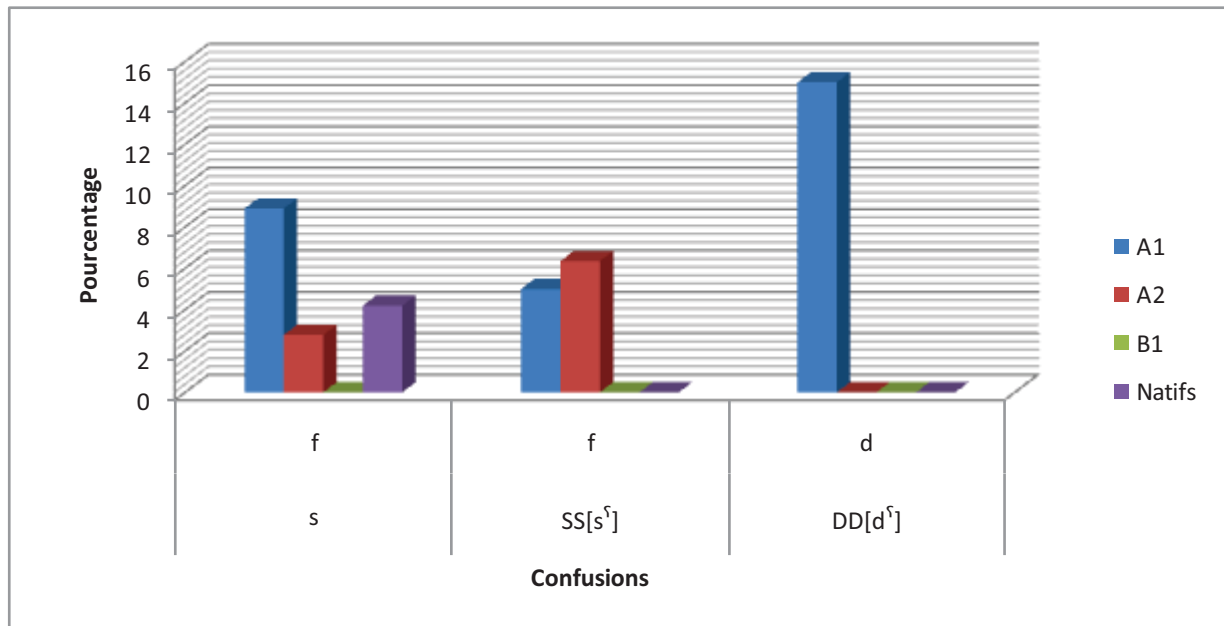


Figure 21 : confusion des sons proches

La figure 21 reporte les confusions entre les sons dite proche. Le point commun entre ces sons est le mode d'articulation. La confusion est remarquée particulièrement chez les A1 et les A2.

La confusion /s/~/f/ : correspond à deux fricatives : le /s/ est une fricative sifflante alvéolaire produite par le rapprochement de la pointe de la langue vers la région alvéolaire, alors que /f/ est une fricative labiodentale produite par le rapprochement de la lèvre inférieure aux dents du haut. La différence se situe donc au lieu d'articulation. La confusion est remarquée chez les A1.

La confusion /sʰ/~/f/ : a lieu entre la fricative le /f/ déjà décrite précédemment et l'emphatique /sʰ/ qui est une sifflante sourde proche du /s/ caractérisé par une résonance emphatiques et une position de la langue courbée vers le bas avec la pointe de la langue qui pointe vers le derrière des dents. [Wikipédia] Sa similarité avec /s/ cause des problèmes de perception pour les non natifs. On observe sur la figure 21 des confusions entre le /f/ et /sʰ/ chez les A1 et les A2. Nous pensons que ces apprenants ne maîtrisent pas bien la différence entre le /s/ et /sʰ/ ce qui nous ramène au cas précédent (confusion du /s/ avec le /f/).

La confusion /dʰ/~/d/ : a lieu entre : le /d/ qui est une occlusive dentale voisée articulée avec la langue qui prend contact avec le bourrelet formé par les alvéoles avec une vibration des cordes vocales et le /dʰ/ qui diffère par une résonance emphatique avec la langue qui est retournée et sa partie antérieure qui prend contact avec l'arrière du palais. Les modes d'articulation des deux phonèmes sont alors différents mais ils sont proches du point de vue acoustique. Les confusions sont observées pour les A1 seulement.

L'analyse des ces résultats confirme que les emphatiques sont difficiles a maîtriser pour les apprenants et ils sont confondus avec les sons proches. Les confusions du /s/ avec le /f/ et

du /s^ʃ/ avec /f/ confirment l'existence des difficultés liées au lieu d'articulation chez les A1 et les A2.

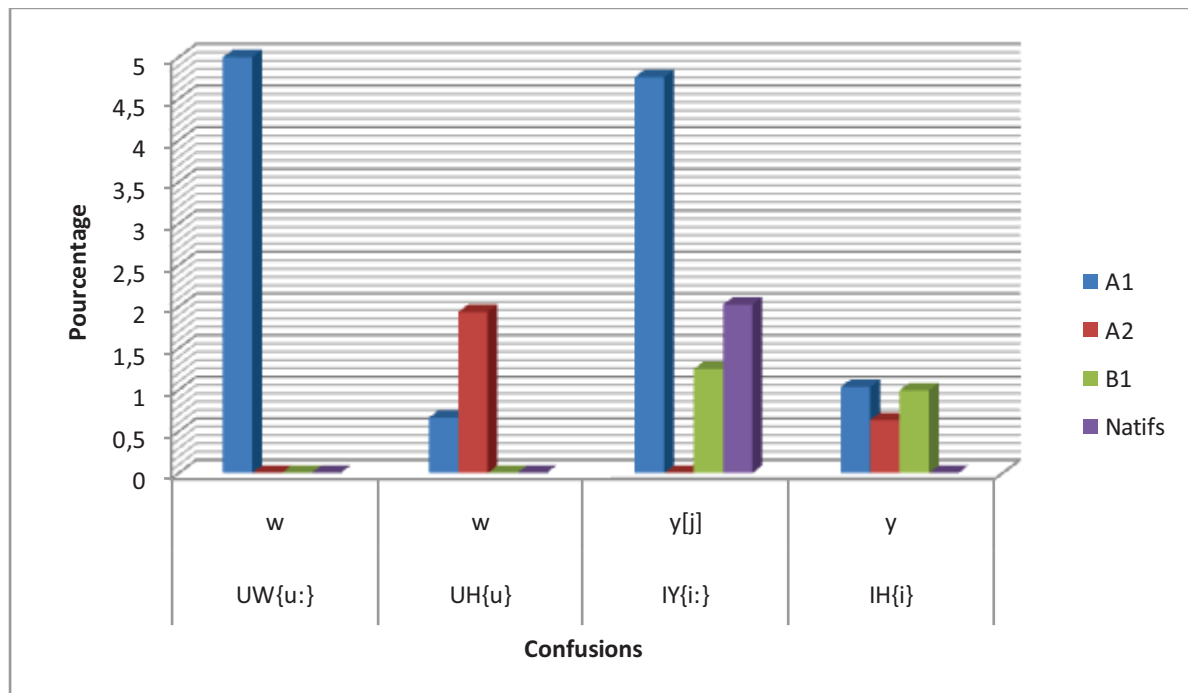


Figure 22 : Les approximantes

La figure 22 rassemble d'autres confusion observés : « Les glissantes sont des éléments phonétiques intermédiaires entre voyelles et consonnes. Elles font parties de la classe des approximantes. En français elles sont au nombre de trois : /j/, /ɥ/ et le /w/. Leur articulation se fait en référence à l'articulation des trois voyelles fermées : /i/ pour /j/ (palatale), /y/ pour /ɥ/ (labio-palatale) et /u/ pour /w/ (labio-vélaire). Le lieu d'articulation des approximantes se réfère à celui des trois voyelles fermées » (C.Meunier, 2007).

L'analyse des résultats du décodage des phonèmes affichés sur la figure 22 montre que les apprenants font des confusions pour les deux couples /i/~j/ et /u/~w/.

Pour le couple /i/~j/ on remarque que se sont les A1 et se de façon moins importante les A2 qui font des confusions pour ces deux sons.

Pour le couple /u/~w/ on remarque que même les natifs font beaucoup des confusions. Pour ces sons on n'a pas suffisamment de données pour conclure.

L'analyse des confusions obtenues suite au décodage acoustico-phonétique confirme les hypothèses tirées de la comparaison des résultats de l'alignement forcé faite avec le modèle acoustique « à gros grain » et le modèle acoustique classique qui montrent que les apprenants de niveau A1 ont des gros problèmes liés au lieu d'articulation. On a aussi remarqué que les A2 continuent à faire des confusions liées au lieu d'articulation mais beaucoup moins que les A1.

Les confusions voyelles courtes, voyelles longues observés confirment que les apprenants ne maîtrisent pas bien la longueur des phonèmes. Les confusions observées pour les natifs concernant ces sons montrent, quant à elles, qu'on ne modélise pas très bien la longueur des phonèmes avec les HMM.

Nous avons aussi observé que le /h/ est un son qui attire les autres sons de la langue ; ce phénomène est plus accentué dans le cas de la parole des apprenants.

Conclusion :

Au cours de ce stage nous avons pu caractériser des différences qui se trouvent au niveau de la parole pour des locuteurs de différents niveaux de langue en utilisant des outils automatiques de reconnaissance de la parole : alignement forcé et DAP.

Nous avons montré que le niveau de langue du locuteur a une influence sur les scores de l'alignement forcé et que la variation des scores n'est pas aléatoire et peut être un indice pour reconnaître le niveau de langue du locuteur.

L'analyse des scores d'alignements nous a permis aussi d'étudier les phonèmes qui causent des problèmes aux élèves. Nous avons montré que les sons ayant des caractéristiques proches de celles de la langue maternelle des apprenants sont les plus difficiles à maîtriser. Dans notre cas (L1 : français, L2 : arabe) se sont les phonèmes emphatiques qui sont les plus durs à assimiler.

Une autre caractéristique difficile à maîtriser pour les apprenants est la durée des phonèmes. Nous avons remarqué cette difficulté au cours des analyses faites sur les voyelles courtes et les voyelles longues de l'arabe.

Nous avons montré aussi que, en utilisant un modèle acoustique « à gros grain », les scores d'alignement forcé continuent à nous renseigner sur le niveau de langue du locuteur. Ce qui nous a permis de dire que la différence observée à travers l'analyse des scores d'alignement est la distance qui sépare la prononciation des apprenants et celles des locuteurs natifs et pas la différence d'accents.

Les analyses des résultats obtenus en utilisant le modèle acoustique « à gros grain » nous a permis aussi de dire que les apprenants de niveau A1 ont plus de difficultés liées aux lieux d'articulations. Ce qui a été confirmé par l'analyse des confusions, à l'issue du décodage acoustico-phonétique (DAP).

Malgré tout, on peut envisager de nombreuses améliorations et perspectives.

Nous allons tout d'abord commencer par les améliorations qu'il faut envisager pour nos outils et données.

- Nous avons remarqué que la durée du phonème n'est pas bien modélisée par les HMM, il faut donc trouver des solutions pour prendre en compte ce paramètre dans le modèle acoustique.
- Améliorer la modélisation des emphatiques dans le modèle acoustique en augmentant leur nombre d'occurrence dans le corpus d'apprentissage, est aussi nécessaire.

Les perspectives qui peuvent être envisagées pour mieux caractériser le problème sont :

- Construire des modèles acoustiques en se basant sur des groupements de phonèmes pour localiser les causes qui engendrent les fautes de prononciation.
- Effectuer des analyses sur des apprenants ayant des langues maternelles différentes et faire des comparaisons.
- Analyser les résultats dans le cas de la parole continue.
- Recueillir suffisamment de données d'apprenants pour essayer d'analyser les paramètres sur les quels peut se baser un système automatique de détection de niveau de langue.

Bibliographie

[Flege 2001] [Article] / aut. Flege // Interactions between the Native and Second-language Phonetic Systems. - [s.l.] : SPEECH COMMUNICATION, 2001. - page 2-6.

[Flege,1995] [Article] / aut. Flege // Second-language speech learning: theory, findings, and problems. - [s.l.] : Speech Perception and Linguistic Experience., 1995. - Timonium, MD; York Press, 233-277, page 6-9.

[kuhl et al, 2008] [Article] / aut. kuhl al // Phonetic learning as pathway to language: new data and native language magnet theory expanded. - 2008. - Philosophical Transaction of royal society.

C. Horgues 2006 [Thèse] / aut. C.Horgues // Prosodie et perception de l'accent français en anglais. - 2006. - Université Paris 7, Denis Diderot.

C.Meunier, 2007 [Article] / aut. C.Meunier // Phonétique acoustique. - Marseille : [s.n.], 2007. - Neurophysiologie de la parole 13, Les dysarthries.

CERC [Livre] / aut. CERC // Les outils du conseil de l'Europe en classe de langue. - [s.l.] : Didier, 2009. - Cadre européen commun, chapitre 3, Page 37,38.

Patkowski, 1990 [Article] / aut. Patkowski // Age and accent in a second language. - 1990.

Rabiner, 1993 [Livre] / aut. Rabiner // Fundamentals of Speech Recognition. - 1993. - Prentice-Hall.

T.Ping, 2008 [Thèse] / aut. T.Ping // Automatic Speech Recognition for Non-native speakers. - 2008, Université J. Fourier

Vieru-Dimulescu 2006 [Article] / aut. Vieru-Dimulescu // Identification perceptive d'accents étrangers en français. - 2006.

Wikipedia [En ligne] / aut. wiki. - 5 6 2011. - 5 6 2011.

Lien/ http://fr.wikipedia.org/wiki/Analyse_de_la_variance.

Xiao,2011 [Article] / aut. Xiao, Laurent Besacier, Eric Castelli // Speech Modulation features for Robust Nonnative Speech Accent Detection. - Florence-Italie : [s.n.], Interspeech2011.

Annexe 1 :

Les transcriptions des mots prononcés dans le corpus classer suivant la classe de langue.

1) Echantillon de mots connus au terme du niveau A1 (CECR) :

Voyelle longue / voyelle courte gémination	Mots contenant des sons inconnus ع ح	Sons existant dans d'autres langues étrangères ه ه ذ خ ر	Hamza médiane et finale	Lettres emphatiques ق ص ض ط ظ	Sons non problématiques	
تَكْم	شارع	ماذا	جزائري	طالب	سكن	1.
تَقَّاح	جامعة	هل	لقاء	صفت	اسبانيا	2.
سنة	عربي	ذهب	سؤال	طلب	لبنان	3.
أمي	مرحباً	أهلاً	سأل	برنقال	في	4.
جدتي	على	تشرّفنا	مئة	أيضاً	أين	5.
شاي	عليكم	إثنان	كأس	شقة	السلام	6.
سمير	عربية	ثلاثة	سوداء	طابق	إلى	7.
سامر	عصير	خمسة	بيضاء	قهوة	تونسي	8.
جمال	عمرى	شرب	زرقاء	فاصوليا	أكل	9.
جميلة	عشرون	أخ	حمراء	حمص	تين	10.
تتورة	وحدى	أخت	صفراء	بنطلون	بناية	11.
شمام	واحد	خال	الأحد	أزرق	بيت	12.
يسير	أربعة	برنقال	الإثنين	أبيض	بلد	13.
سير	سبعة	بأذنان	الثلاثاء	أصفر	مدينة	14.
بريد	تسعة	كريم	الأربعاء	أخضر	ألف	15.
بارد	عشرة	درس	أبناء	طويل	أبي	16.
عالم	بعيد	ثمن	دقائق	قصير	ابن	17.
علم	ساعة	المغرب	نائم	قميص	فنجان	18.
يدور	حي	هذا	جائع	قلم	ليمون	19.
دُر	عمي	هذه	ماء	قبل	موز	20.

2) Echantillon de mots connus au terme du niveau A2 (CECR) :

Voyelle longue / voyelle courte gémination	Mots contenant des sons inconnus ع ح	Sons existant dans d'autres langues étrangères ن ه ذ ح ر	Hamza médiane et finale	Lettres emphatiques ق ص ض ط ظ	Sons non problématiques	
سافر	حساب	اتّخار	زبائن	فندق	زبون	1.
سفر	اعتماد	رسالة	طائرة	مصرف	وكالة	2.
مسجل	فتح	خدمة	غداء	أغلق	سياسي	3.
مال	عُطلة	خُطوط	عشاء	بطاقة	باب	4.
مجلة	أسبوع	جريدة	أصدقاء	طابع	تسليّة	5.
شباك	عملة	ثقافي	اتّمان	صندوق	أسس	6.
حارّ	تحويل	نافذة	إلغاء	صيف	دوليّة	7.
حرّ	حجز	شهرية	رأي	ضيف	كلمة	8.
مجانّة	اجتماعي	تاريخ	آراء	موظف	يومية	9.
صرف	مُحقّق	زيارة	قضاء	مضمون	سنويّة	10.
صراف	مُتوّعات	زار	أسئلة	مطار		11.
شهد	بحر	غرفة	سائح	ألصق		12.
شاهد	ممنع	غادر	رائد	شخص		13.
معالم	حرارة	أخبار	تأشيرة	رياضة		14.
معلم	عاصفة	ذهاب	مائدة	صحيفة		15.

3) Echantillon de mots connus au terme du niveau B1 (CECR)

Voyelle longue / voyelle courte gémination	Mots contenant des sons inconnus ع ح	Sons existant dans d'autres langues étrangères ن ه ذ ح ر	Hamza médiane et finale	Lettres emphatiques ق ص ض ط ظ	Sons non problématiques	
مرشّح	محكمة	وزير	رئاسة	اتّفاق	فاز	1.
أُكّد	حكومة	وزارة	مسؤول	مفاوضات	فوز	2.
دمار	حُكم	برلمان	اعتداء	مقاطعة	سياسات	3.
دمر	ترشيح	انتخابات	لقاء	مقاومة	زلزال	4.
صوت	علاقة	خاسر	فائز	مظاهرة	ولاية	5.

6.	دولة	نظير	نائب	داخليّة	اقتراع	صوت
7.	مجلس	قتيل	رأسمالية	خارجيّة	مشروع	حلّ
8.	لجنة	قانون	استأنف	هجوم	حزب	حالة
9.	سلام	وطنيّ	فقراء	انفجار	نزاع	استقلّ
10.	بيان	إقليميّ	أغنياء	قوّات	حرب	استقال
11.	جيش	مصدر	الأحزاب	إرهاب	حوار	صرّح
12.	وفد	وظيفة	الأراضي	هجرة	ضحية	فجرّ
13.	مندوب	وضع	الإدارة	مهمة	احتلال	اجتمع
14.	بلديّة	خطة	الاستقلال	ثورة	مقعد	اجتماع
15.	أفاد	سلطة	فائدة	اختيار	أعلن	ذكر
16.	سجين	نظام	وزراء	سفير	حقوق	نكّر
17.	سجن	منظمة	سفراء	سفارة	عسكريّ	اتّهم
18.		خطاب	رؤساء	مُدكّرة	عدالة	اتّهام
19.		تطبيق	سجناء	اهتمّ	حرية	وقّع
20.		إطلاق	لائحة	دور	شيوعيّ	واقع

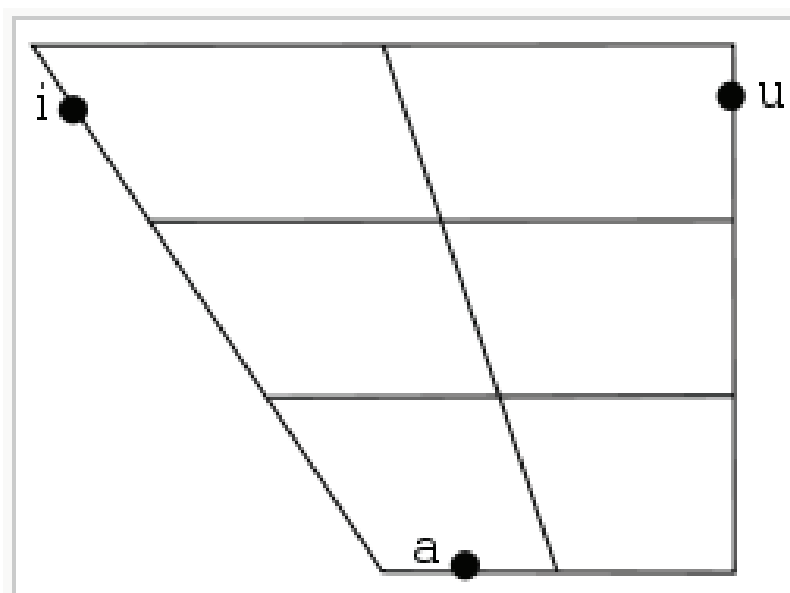
Annexe 2

L'alphabet phonétique arabe : tableau des consonnes

	Labial	Plain		Emphatic		Palatal	Velar	Uvular	Pharyngeal	Epiglottal	Glottal
		Dental	Alveolar	Alveolar	Dental						
Nasal	m		n								
Stop	voiceless	(p)	t	tʰ			k	q			ʔ
	voiced	b	d	dʰ			g				
Fricative	voiceless	f	θ	s ʃ	sʰ			x~χ		ħ~h	h
	voiced	(v)	ð	z ʒ~dʒ	zʰ~ðʰ			ɣ~ʁ		ʕ~ʕ	
Approximant	w			ɹ~ɹʰ			j				
Trill				ɾ~ɾʰ							

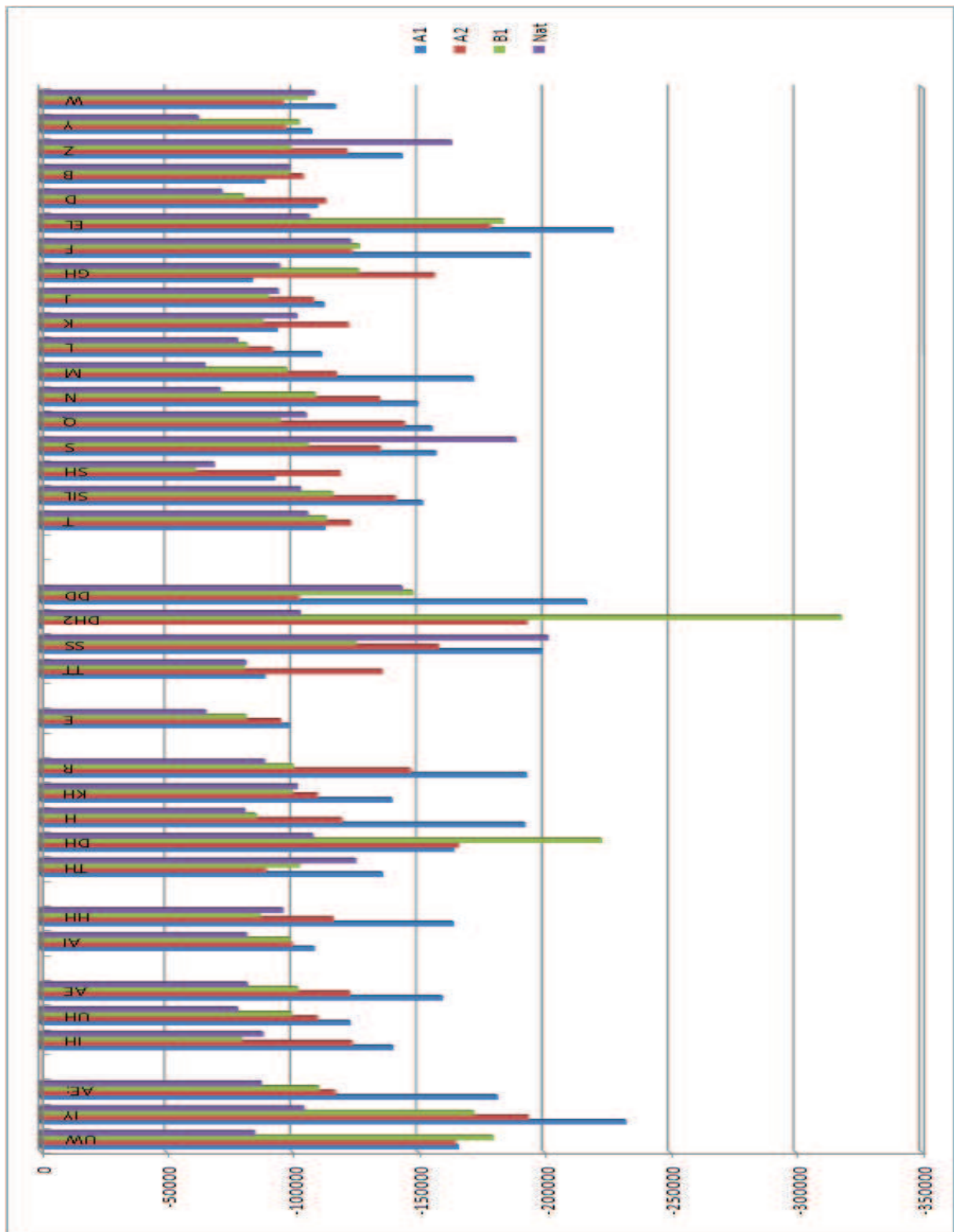
http://en.wikipedia.org/wiki/Arabic_phonology

Tableau des voyelles arabe:



Annexe 3 :

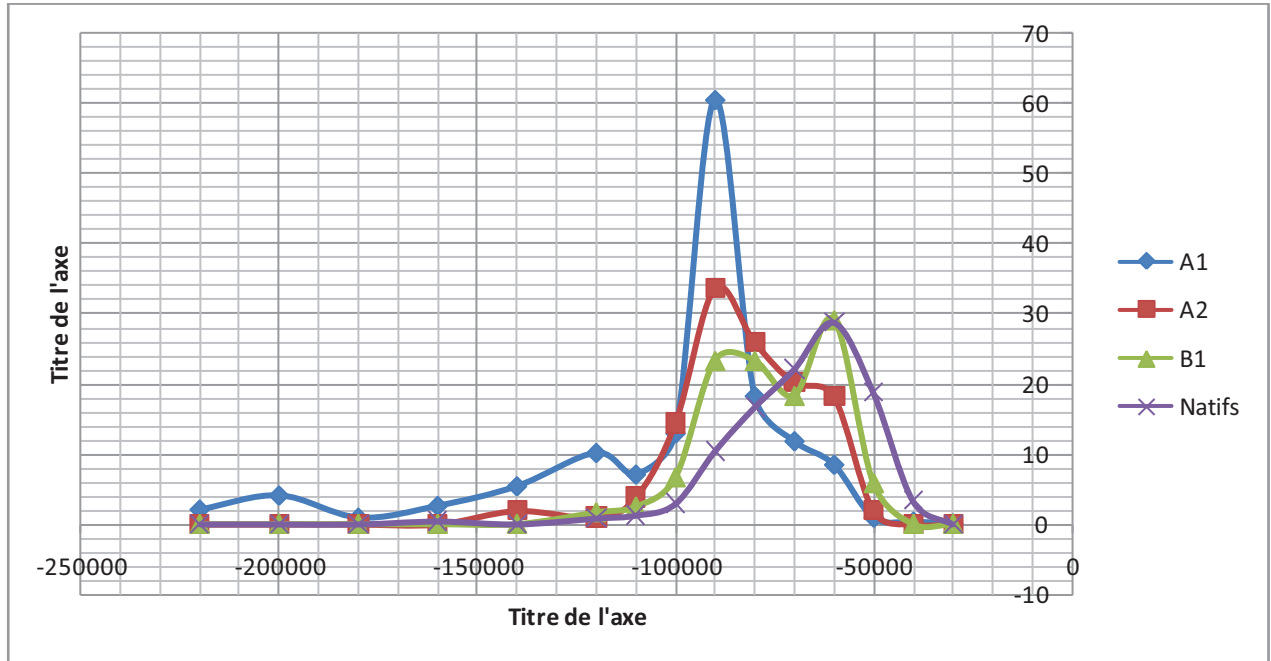
Les scores d'alignement forcé par phonème.



Annexe 4 :

La distribution des scores sur les corpusA1 et le corpusA2 en utilisant le modèle acoustique « à gros grain ».

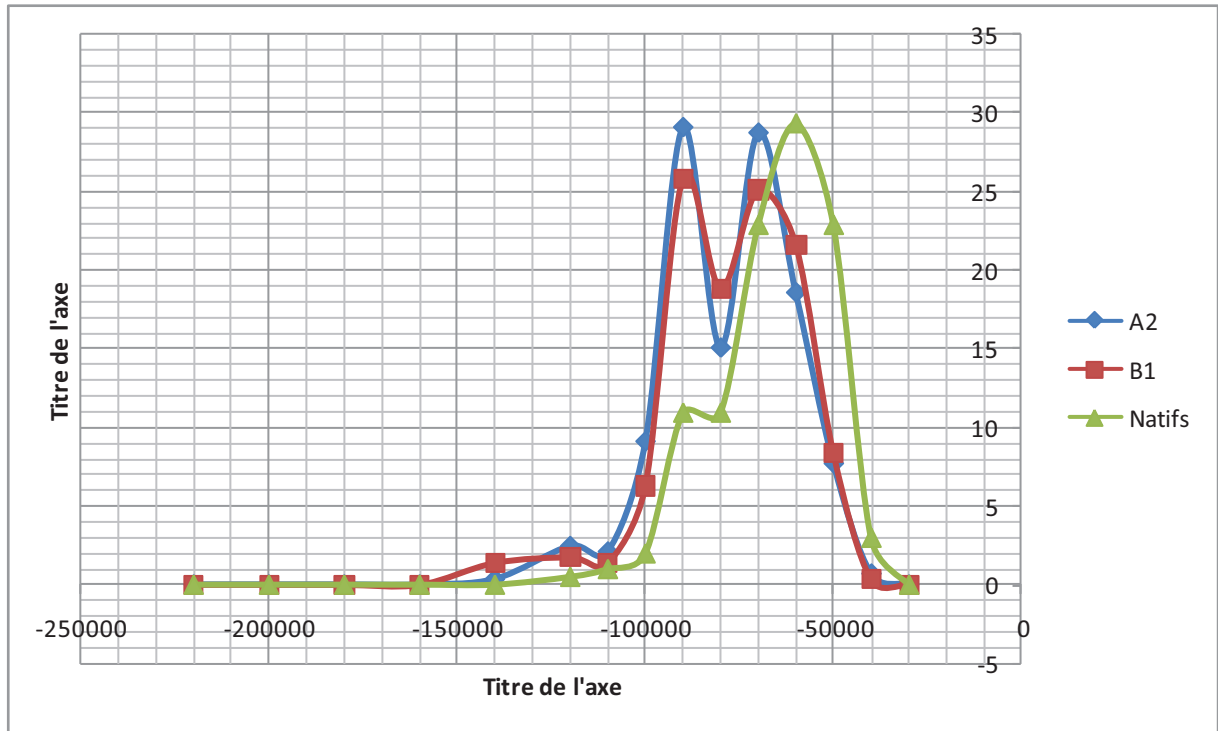
Courbe pour le corpusA1 :



Les résultats de la correction de Bonferroni :

$$\begin{pmatrix} & A1 & A2 & B1 \\ A2 & < 0.01 & & \\ B1 & < 0.01 & 0.245 & \\ Natifs & < 0.01 & < 0.01 & < 0.01 \end{pmatrix}$$

Courbe pour le corpusA2B1 :



Les résultats de la correction de Bonferroni :

$$\begin{pmatrix} & A2 & B1 \\ B1 & < 0.19 & \\ Natifs & < 0.01 & < 0.01 \end{pmatrix}$$

Annexe 5 :

Les confusions :

Alignement	décodage	A1	A2	B1	natif
AE{a}	AE:{a:}	28	38	9	11
AE{a}	h	22	27	13	12
AE{a}	AI[ɿ]	13	4	0	0
AE{a}	f	9	0	7	16
AE{a}	n	7	6	2	0
AE{a}	b	2	0	0	0
AE{a}	d	1	0	0	0
AE{a}	IY{i:}	2	0	0	0
AE{a}	sil	2	0	3	0
AE{a}	t	1	0	0	2
AE{a}	UW{u:}	1	0	0	4
AE{a}	y	1	0	0	0
AE{a}	E[ʔ]	2	0	0	0
UH{u}	h	12	10	10	5
UH{u}	AE:{a:}	12	0	0	0
UH{u}	AE{a}	5	0	0	0
UH{u}	w	1	5	0	0
UH{u}	AI[ɿ]	2	2	0	0
UH{u}	IH{i}	2	0	1	1
UH{u}	n	2	3	0	0
UH{u}	sil	2	0	1	0
IH{i}	AE{a}	10	0	0	5
IH{i}	IY{i:}	7	15	9	12
IH{i}	n	3	0	0	0
IH{i}	AE:{a:}	2	12	0	0
IH{i}	AI[ɿ]	4	2	0	0
IH{i}	d	2	0	2	0
IH{i}	el	1	0	1	0
IH{i}	h	7	5	0	0
IH{i}	k	2	0	0	2
IH{i}	SH[ʃ]	3	0	0	1
IH{i}	y	2	2	4	0
AE:{a:}	AE{a}	7	0	4	0
AE:{a:}	h	5	0	0	2
AE:{a:}	r	2	7	5	0
UW{u:}	w	2	0	0	0
UW{u:}	h	1	0	1	0
UW{u:}	UH{u}	2	3	0	0
IY{i:}	y	5	0	1	3
IY{i:}	h	13	13	11	8

IY{i:}	r	3	0	3	0
IY{i:}	AE:{a:}	2	0	0	0
DD[d ^ʰ]	d	3	0	0	0
DD[d ^ʰ]	h	3	0	0	0
SS[s ^ʰ]	h	6	1	0	3
SS[s ^ʰ]	f	2	4	0	0
KH[x]	h	4	1	0	0
KH[x]	HH[ħ]	1	0	0	0
KH[x]	UW{u:}	1	0	0	0
E[ʔ]	AE{a}	2	0	0	0
E[ʔ]	AE:{a:}	2	0	0	0
E[ʔ]	AI[ɿ]	2	0	0	3
E[ʔ]	IY{i:}	2	0	1	0
E[ʔ]	n	3	0	0	0
E[ʔ]	sil	2	0	4	0
E[ʔ]	TH[θ]	1	0	0	0
E[ʔ]	h	12	9	8	10
AI[ɿ]	h	3	12	4	8
TH[θ]	f	1	0	0	2
n	h	47	55	36	29
s	f	8	3	0	6
l	ae:	5	0	0	0
l	f	5	0	0	0
l	h	7	20	4	6
l	n	9	0	0	7
m	r	6	0	1	0
n	sil	5	0	8	0
q	f	3	3	0	2
r	n	3	0	2	0
sil	h	4	8	4	0
y	iy	4	0	1	3
b	ae	3	0	0	0
b	ae:	2	0	0	0
d	b	2	0	0	0
h	n	4	26	7	10
h	sil	3	2	14	0
m	n	6	0	0	10
m	sil	3	0	0	0
m	y	2	0	2	0
n	ae:	3	0	0	1
q	h	7	7	0	7
q	sil	2	0	0	0
r	h	15	23	15	20
r	ae:	3	0	0	0
r	f	2	3	0	4
r	sil	2	0	6	0

s	sh	2	0	1	0
t	h	5	13	5	0
b	n	1	0	0	0
b	y	1	0	0	0
d	ae:	1	0	0	0
d	h	8	0	4	4
d	r	2	0	0	0
d	sil	1	0	0	0
d	t	1	2	0	0
d	uw	3	0	0	0
dh	t	1	0	0	0
el	ae:	1	0	0	0
f	h	3	0	11	0
f	y	1	0	1	0
gh	e	1	0	0	0
h	d	1	0	0	1
hh	h	3	7	0	3
j	s	1	0	0	0
j	sh	5	0	1	0
k	iy	1	0	0	0
l	y	1	0	0	0
m	el	1	0	0	0
m	f	1	0	0	0
m	h	12	0	7	5
m	l	1	0	6	0
n	ae	3	0	0	0
n	d	1	0	0	0
n	el	2	0	0	0
n	iy	2	0	0	0
q	t	1	0	0	0
q	uw	1	0	0	0
r	ae	1	0	0	0
r	el	1	0	0	0
r	kh	1	0	0	0
r	t	1	0	0	0
r	y	1	0	0	0
s	ai	1	0	0	0
s	r	1	0	0	0
t	ae:	1	0	0	0
t	ai	1	0	0	0
t	j	1	0	0	0
t	y	1	3	1	0
y	ae	2	0	0	0
y	ae:	1	0	0	0
y	ai	2	0	0	0
y	r	2	0	0	0



Déclaration anti-plagiat

Document à scanner après signature
et à joindre au mémoire électronique

DECLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : *Ben Jannet* PRENOM : *Mohamed A. meus*.....

DATE : *08/07/2011*....

