Analyse factorielle pour une représentation vectorielle des états des Modèles de Markov Cachés

Mohamed Bouallègue Driss MATROUF Georges Linares
Université d'Avignon, LIA, France
mohamed.bouallegue, driss.matrouf, georges.linares@univ-avignon.fr

RÉSUMÉ

Dans cet article nous proposons une représentation vectorielle des états des HMM. L'idée de base nous a été inspirée par l'approche SGMM (Subspace Gaussian Mix- ture Models paradigm). Modéliser les états des HMM par de simples vecteurs (au lieu de GMM) permet de faciliter un grand nombre de tâche dans le cadre du traitement automatique de la parole, comme la classification automatique des états ou des phonèmes. La visualisation graphique des états devient possible grâce à cette représentation vectorielle. Le représentation vectorielle des états permet de voir les états comme un nuage de points dans un espace multi-dimensionnel, ce qui permet d'étudier ses caractéristiques en utilisant des techniques d'analyse de données. Dans cet article nous expliquerons comme obtenir cette représentation et comme l'utiliser pour réaliser la procédure de partage d'états pour fabriquer des modèles contextuels avec des états partagés. Nous comparons notre approche avec celle fondée sur l'utilisation des arbres de décision.

ABSTRACT

Subspace Gaussian Mixture Models for Vectorial HMM-states Representation

In this paper we present a vectorial representation of the HMM states that is inspired by the Subspace Gaussian Mix- ture Models paradigm (SGMM). This vectorial representation of states will make possible a large number of applications, such as HMM-states clustering and graphical visualization. Thanks to this representation, the Hidden Markov Model (HMM) states can be seen as sets of points in multi-dimensional space and then can be studied using statistical data analysis techniques. In this paper, we show how this representation can be obtained and used for tying states of an HHM-based automatic speech recognition system without any use of linguistic or phonetic knowledge.

MOTS-CLÉS : Représentation vectorielle des états des HMM, partage d'états, classification des états d'un HMM.

KEYWORDS: HMM-state vector representation, Speech recognition, Acoustic Modelling, HMM states clustering..

1 Introduction

Depuis les première applications dans le domaine de traitement de la parole, la problématique du calcul de la similitude entre phonèmes (ou allophones) a été posée comme sujet de recherche par la communauté scientifique. En reconnaissance de la parole, Cette mesure est utilisée principalement dans la procédure de "parage d'états". Cette approche proposée par Young (Young, 1992)

consiste à associer la même fonction de densité de probabilité (probability density function : pdf) aux états des phonèmes contextuels acoustiquement proches. Le "partage d'états" est une procédure incontournable dans la processus de la modélisation acoustique afin d'avoir un équilibre entre la quantité de données d'apprentissage (pour phonème contextuel) et le nombre de paramètres à estimer dans les modèles acoustiques. Elle est basée sur la définition d'une distance entre les états des MMC (mesure de similitude). En supposant que chaque phonème est modélisé par une seule gaussienne, Young (Young et Woodland, 1993) a utilisé la similitude entre phonème en utilisant une distance entre deux gaussiennes, appelée la divergence de Kullback-Leibler [1]. Pour calculer la distance entre deux mélanges de gaussiennes, Mak a proposé une expression alternative de la similitude basée sur la distance de Bhattacharyya [2] (Mak et Barnard, 1996). La distance de Bhattacharyya permet de mesurer la distance théorique entre deux distributions gaussiennes. Mak a proposé dans son travail de l'étendre pour calculer la distance entre deux mélanges de gaussiennes.

Dans ce travail, nous proposons une nouvelle vision de la classification des phonèmes. Au contraire des travaux mentionnés auparavant, les phonèmes sont caractérisés par des vecteurs estimés par l'approche d'analyse factorielle (Kenny *et al.*, 2005). Un intérêt indéniable de cette représentation vectorielle est la possibilité de traiter les états par des techniques d'analyse de données, telle que l'analyse factorielles des correspondances. Le fait de représenter les états par des vecteurs permet aussi de mesurer efficacement la similarité entre eux, en utilisant des distance adaptées (distance euclidienne ou distance de Mahlanobis). Il est important de noter que cette tâche (mesure de similarité) était très complexe (voire très approximative) lorsque les états n'avaient pour représentants que les GMMs. Nous proposons ici l'utilisation de cette représentation vectorielle dans la modélisation acoustique pour la reconnaissance de la parole. Nous nous basons sur les vecteurs pour réaliser la procédure de regroupement d'états des MMCs. Nous appelons les vecteurs représentatifs des états des MMCs *facteurs d'états*.

Dans la section 2.1 nous exposons la méthode d'estimation des *facteur d'états*. Nous détaillerons dans la section 2.2 les différentes étapes de la procédure de regroupement des états. Nous exposons ensuite les résultats obtenus sur la langue française. Dans la section 4 nous montrons l'intérêt de l'application de cette méthode dans la modélisation acoustique pour les langues peu dotées en ressources.

2 Représentation vectorielle des états

Dans cette section nous exposons le modèle utilisé ainsi que la procédure de regroupement d'états pour une mutualisation des paramètres dans le cadre d'une modélisation acoustique.

2.1 Analyse factorielle pour une représentation vectorielle d'états de MMC

Dans la modélisation acoustique basée sur l'analyse factorielle, tous les GMMs associés aux états des MMC sont dérivés à partir d'un seul modèle générique appelé "modèle du monde" (GMM-UBM). C'est un mélange de gaussiennes modélisant tout l'espace acoustique de la parole.

1. www.aiaccess.net/french/Glossaires/GlosMod/f_gm_kullbak.htm
2. http ://www.cse.yorku.ca

Les moyennes et les poids des états sont obtenus à partir de l'UBM, alors que les variances restent inchangés par rapport à l'UBM. Soit $\mathbf{m}$ le super-vecteur du modèle du monde obtenu par la concaténation des moyennes de ses gaussiennes. Le super-vecteur $\mathbf{m}_e$ de l'état e est obtenu par l'équation suivante :

$$\mathbf{m}_e = \mathbf{m} + \mathbf{U}\mathbf{x}_e \tag{1}$$

Dans ce modèle, la matrice $\mathbf{U}$, de faible dimension R, représente le sous-espace de la variabilité inter-états. Elle est estimée sur toutes les données de tous les états des MMC. $\mathbf{x}_e$ est un vecteur de dimension R estimé sur les données propres à l'état e. C'est un vecteur caractéristique de l'état que nous appelons *facteur d'état*. L'algorithme d'estimation de $\mathbf{U}$ et des $\mathbf{x}_e$ est décrit en détail dans (Matrouf *et al.*, 2007).

Nous proposons d'adopter les *facteur d'état* comme une représentation vectorielle des états des MMC. Les *facteurs d'états* simplifient le calcul des distances entre les états des MMC. En effet, nous remplaçons les mélanges des gaussiennes par un simple vecteur de faible dimension (maximum 100 paramètres). Les distances utilisées généralement dans la classification phonétique sont remplacées par un simple calcul de distance entre vecteurs. Avec la représentation vectorielle, la classification phonétique devient un problème de classification classique dans un espace R^d. Par ailleurs, cette représentation vectorielle permettra d'exploiter les résultats scientifiques obtenus au cours de plusieurs années de recherche dans le domaine de l'analyse de données. Ces résultats peuvent servir dans l'analyse de la variabilité acoustique et de la variation de l'articulation phonétique. Également, les *facteurs d'états* peuvent être utilisés dans autres applications comme la phonétique clinique, la détection de dialecte ou l'identification automatique de la langue.

2.2 Procédure de regroupement d'états basée sur des facteurs d'états

Dans la littérature, le regroupement est souvent réalisé par un algorithme de classification utilisant les arbres de décision (Reichl et Chou, 1998). Plusieurs inconvénients sont inhérents à l'utilisation de cette méthode : elle nécessite notamment des connaissances linguistiques (pour construire le jeu de questions), qui peuvent ne pas être disponibles pour certains langues. De plus, le temps de calcul est long à cause de l'évaluation des vraisemblances pour chaque question à chaque noeud de la structure de l'arbre. Comme alternative nous proposons d'utiliser les *facteurs d'états* dans le regroupement des états des MMC. Avec les *facteurs d'états* le calcul de la probabilité est remplacé par un simple calcul de distance entre vecteurs et le regroupement des états peut être formulé comme un problème de classification classique dans R^d. Cette méthode ne nécessite plus de connaissance phonétique ou linguistique, ce qui nous permet de contourner le problème d'insuffisance ou l'absence de ce types d'informations pour les langues peu dotées. Dans la suite nous exposons les différentes étapes de réalisation de partage d'états en utilisant les *facteurs d'états*.

Nous présentons dans cette section les étapes nécessaires pour réaliser le regroupement d'états en utilisant les facteurs d'états.

– Etape-1 : utilisation d'un système de reconnaissance de la parole existant pour faire la segmentation par rapport aux états : alignement forcé. Dans cette étape nous utilisons l'algorithme de Viterbi. Dans cette segmentation, chaque phonème indépendant du contexte

est représenté par trois états.

– Etape-2 : contextualisation des phonèmes dans la segmentation obtenue dans l'étape-1. En effet, nous associons à chaque phonème le phonème précédent et le phonème suivant. L'objectif de cette étape est de trouver tous les phonèmes contextuels dans notre corpus d'apprentissage.

– Etape-3 : estimation des *facteurs d'états* $\mathbf{x}_s$ pour chaque état avec l'équation 1. Une bonne estimation des $\mathbf{x}_s$ nécessite un nombre minimum de trames pour chaque état. Pour cela nous ignorons les états qui sont rarement observés dans le corpus d'apprentissage et nous calculons les *facteurs d'états* que pour les états qui ont suffisamment de trames (dans nos expériences nous avons traité les états qui ont plus de 50 trames).

– Etape-4 : utilisation des *facteurs d'états* obtenus précédemment pour classifier les états en utilisant l'algorithme de classification non-supervisée k-means [3]. Cette classification permet de regrouper les états dépendants du contexte qui sont acoustiquement proches. Nous appelons les classes obtenues, *classe-états*.

– Etape-5 : modélisation des états appartenant à la même classe par un seul GMM appelé *GMMs-classe-états*. Ces GMMs sont dérivés à partir du GMM-UBM en utilisant une adaptation par maximum a posteriori (MAP, voir section 1.5.1) au moyen des données appartenant à chaque *classe-état* (obtenu dans l'étape-4).

– Etape-6 : affectation des états non classifiés dans l'étape-4 (qui ont moins de 50 trames) à la classe la plus proche. Pour ce faire, nous calculons la vraisemblance des trames de chaque état non classifiés sur les *GMMs-classe-état* obtenu dans l'étape-5. Après, nous assemblons ses trames avec les trames de classe la plus vraisemblable.

– Etape-7 : adaptation de chaque ancienne *GMMs-classe-état* obtenu dans l'étape-5 sur les données de la *classe-états* correspondante incluant les données les états nouvellement classifiés.

– Etape-8 : construction des MMC en utilisant les *GMMs-classe-état* obtenu dans l'étape présidente.

– Etape-9 : Pour améliorer les performances de notre modèle, nous appliquons la procédure récursive standard du ré-alignement/ré-estimation des paramètres.

3 Résultats expérimentaux

Dans ces expériences, nous avons utilisé le système SPEERAL avec un processus de transcription en deux passes (voir section 3.3.2). Les nouveaux modèles acoustiques sont estimés et évalués sur le corpus d'évaluation ESTER (décrit dans la section 3.3.3). Les modèles acoustiques de base sont des MMC gauche-droit avec 13316 phonèmes contextuelles. Pour modéliser tous ces phonèmes

3. l'algorithme k-means permet une classification non-hiérarchique en minimisant la variance intra-classe basée sur la distance Euclidienne ; cette procédure est un moyen simple pour classifier un ensemble de données dans un certain nombre de classes (noté k) préalablement fixé

contextuels indépendamment, nous avons besoin de 39.948 états dans les MMCs. Ce nombre est réduit à 5050 états par l'application du regroupement d'états basé sur des arbres de décision descendantes. Cette approche utilise des questions linguistiques pour construire les arbres de décision.

Dans cette expérience nous avons fixé le nombre de paramètres des *facteurs d'états* à 60. Nous rappelons que les GMMs d'états des MMC sont dérivés à partir d'un seul modèle du monde avec une adaptation MAP. Pour améliorer les performances, nous proposons de modéliser indépendamment chaque état en utilisant l'algorithme EM (Espérance-Maximisation). Dans notre procédure de modélisation, une fois la classification terminée, nous estimons un GMM pour chaque classe d'états avec l'algorithme EM. Les résultats sont exposés dans le tableau 1.

	modèle de base	Modèle MAP	modèle EM
F.Inter	30.88	30.10	29.38
RFI	16.62	16.09	15.84
TVME	25.74	24.90	24.77
AFRICA	29.42	28.89	27.47
Moyenne	27.50	26.61	25.66
gain absolu	-	0.89	1.84

TABLE 1 – *Les performances du modèle MAP et modèle EM comparées avec le modèle de base.*

4 Modèle indépendant du contexte pour les langues peu dotées basé sur les facteurs d'états

Parmi les 6 912 langues parlées dans le monde, seul un tout petit nombre d'entre-elles possède les ressources nécessaires pour implémenter des technologies issues du traitement du langage naturel. Au niveau de la modélisation acoustique, plusieurs travaux sont proposés dans la littérature pour contourner les difficultés d'absence partielle ou totale de ressources linguistiques et informatiques. La solution la plus utilisée aujourd'hui est le *bootstraping* (Osterholtz *et al.*, 1992). Cette solution consiste à obtenir un tableau de correspondances phonétiques (phone mapping) entre une ou plusieurs langues sources et la langue cible (la langue peu dotée). Nous distinguons deux classes de méthodes pour réaliser cette solution : la premiere classe comprend les méthodes manuelles à base de connaissances linguistique et phonétiques (Le et Besacier, 2009). Elles consistent à chercher les couples de phonèmes source/cible les plus proches dans le tableau d'Alphabet Phonétique International(API). Dans une deuxième classe, nous trouvons les méthodes automatiques (Anderson *et al.*, 1994). Ces méthodes utilisent un modèle de la langue source et un corpus vocal étiqueté de la langue cible pour calculer la matrice de confusion entre les phonèmes et trouver le tableau de correspondances phonétiques. Ces méthodes sont utilisées pour construire un modèle acoustique indépendant du contexte. Pour construire des modèles acoustiques contextuels, des travaux comme (Beulen et Ney, 1998) (Singh *et al.*, 1999) proposent d'utiliser la procédure classique basée sur les arbres de décision. Les questions nécessaires à la construction des arbres sont générées automatiquement. D'autres travaux proposent de combiner les modèles acoustiques (au niveau des phonèmes contextuels) de différentes langues sources afin d'obtenir un modèle contextuelle de la langue cible (Beyerlein, 1998) (Schultz et Waibel,

2001).

La plupart des travaux de modélisation acoustique contextuelle, cités auparavant, sont basés sur l'idée d'association entre une ou plusieurs langues source et la langue cible. Les différentes applications montrent que l'association à base de connaissances linguistiques donnent de meilleures résultats que les méthodes automatiques basées sur le calcul de distance entre modèles. Mais ces solutions (à base des connaissances linguistiques) se retrouvent toujours face au problème de la couverture phonétique. De plus, elles nécessitent des connaissances linguistiques et phonétiques des deux langues source et cible, ce qui n'est pas toujours disponible (en quantité et en qualité) pour la langue cible. Afin de contourner ces difficultés, nous proposons d'appliquer la méthode de modélisation acoustique contextuelles proposée dans la section 2.2. Dans la suite nous montrons les résultats obtenus sur la langue vietnamienne.

4.1 Résultats expérimentaux : application à la langue vietnamienne

La langue vietnamienne appartient au groupe Viet-Muong, qui est un membre de la branche mon-khmer qui fait partie de la branche austro-asiatique [4]. Elle est parlée par environ 82 millions de personnes, principalement au Vietnam. Le vietnamien est une langue tonale qui possède six tons avec des caractères accentués pour les tons. Le vietnamien possède 41 phonèmes, dont 23 consonnes, 13 voyelles simples, 3 diphtongues et 2 semi-voyelles, représentées par 29 lettres dans l'alphabet (N. Thi, 2006).

Corpus de la parole : dans ces expériences nous avons utilisé une partie du corpus VNSPEECHCORPUS (Tran, 2003). C'est un corpus de 39 heures de la parole enregistrées au centre MICA [5]. En 2005, il contenait 39 locuteurs, 19 femmes et 20 hommes, venant des régions nord, centre et sud du Vietnam. Nous utilisons uniquement les enregistrements de locuteurs d'une langue standard (nord du Vietnam). Au total, environ 9 heures de parole sont enregistrées, ce qui correspond à 18 locuteurs : 10 hommes et 8 femmes. Nous avons utilisé 7 heures pour l'apprentissage des modèles (8 hommes et 6 femmes) et 2 heures pour le corpus de test (2 hommes et 2 femmes).

Corpus de texte : le corpus de texte vietnamien, utilisé pour estimer le modèle de langage, est collecté exclusivement à partir du web et des journaux numériques. Ce corpus est constitué de 2,7 millions de phrases avec 45 millions de syllabes.

Modèle de base : Arbre de décision

Le modèle de phonème indépendant du contexte est obtenu avec le *bootstraping* (Osterholtz *et al.*, 1992). Un tableau de correspondances phonétiques est construit entre les phonèmes vietnamiens et les phonèmes du français en utilisant l'Alphabet Phonétique International. Le modèle de phonème contextuel est obtenu par un arbre de décision basé sur des questions générées automatiquement (Singh *et al.*, 1999). Nous avons construit plusieurs modèles acoustiques de différentes tailles afin de trouver le nombre optimal des paramètres du modèle. Nous avons fait varier les nombres d'états par MMC et le nombre de gaussiennes par état. Dans le tableau 2, nous exposons les résultats en termes de taux d'erreur mot des différents modèles. Chaque colonne exposent les nombres d'états dans les MMC. Les lignes exposent les nombres de gaussiennes par état. Le modèle qui est composé de 600 états avec 64 gaussiennes a les meilleure performances (32,70% WER). Ce modèle sera notre modèle de base.

4. http ://www.ethnologue.com
5. http ://www.mica.edu.vn

	200s	400s	600s	800s	1200s
32g	33.11	33.51	34.70	34.90	36.10
64g	32.81	32.73	**32.70**	33.00	34.90
128g	33.83	33.80	34.40	35.30	36.40
256g	34.00	33.97	36.10	35.40	38.40

TABLE 2 – *Performances des modèles acoustiques contextuels de différentes tailles où le regroupement des états contextuels est basé sur un arbre de décision.*

***modèle contextuels du vietnamien basé sur les* facteurs d'états**

Dans ces expériences, nous appliquons la méthode de modélisation contextuelle proposée sur le vietnamien. Dans une première expérience, nous avons estimé quatre groupes de *facteurs d'états* qui ont respectivement 20, 40, 80 et 120 coefficients. Par la suite, nous utilisons chaque groupe de vecteurs pour regrouper les états dans 1000, 1200, 1400, 1600, 1800, 2000 et 2200 classes respectivement. Le nombre de gaussiennes par état est fixé à 128 gaussiennes. Dans le tableau 3, nous exposons le WER obtenu par les différents modèles. Chaque ligne du tableau correspond au nombre d'états des modèles. Dans les colonnes nous exposons les tailles des *facteurs d'états* que nous avons utilisées dans la phase du regroupement.

	20	40	60	80	120
1000	28.80	29.50	36.15	38.50	57.10
1200	28.00	28.80	35.10	38.40	60.70
1400	28.10	28.20	33.80	35.70	52.30
1600	27.20	27.80	33.20	35.10	49.90
1800	26.80	27.20	30.70	34.20	50.50
2000	**26.60**	27.50	30.90	33.80	49.00
2200	26.90	27.50	31.10	33.55	49.24

TABLE 3 – *Performances des modèles de différents nombres d'états MMC, avec 128 gaussiennes par états.*

Le meilleur modèle donne un gain absolu de 6,10% par rapport au modèle de base. Les résultats obtenus montrent que les modèles ayant le plus grand nombre d'états obtiennent de meilleurs résultats. En outre, nous observons qu'à partir de 1800 états le gain devient stationnaire. Ces résultats montrent que, à cause de la quantité limitée de données d'apprentissage disponible pour chaque état, les meilleurs *facteurs d'états* sont ceux avec la plus faible dimension (20 coefficients).

Dans une deuxième expérience, nous avons essayé de trouver le nombre optimal de gaussiennes par états. Le nombre de paramètres des *facteurs d'états* est fixé à 20. Nous fixons le nombre d'états de 1600, 1800 et 2000. Pour chaque modèle, le nombre des gaussiennes varie de 64, 128 et 256 gaussiennes.

Le tableau 4 montre que le meilleur modèle est celui avec 2000 états et 256 gaussiennes par état, avec un gain de 7,8% absolu par rapport au modèle de base. Pour améliorer les performances du système, nous avons testé une approche de regroupement guidé. Nous rappelons que chaque

	64g	128g	256g
1600	30.60	27.80	26.00
1800	29.80	27.20	25.50
2000	29.30	26.60	**24.90**

TABLE 4 – *Performances des modèles en fonction du nombres d'états et du nombres de gaussiennes.*

phonème contextuel *ph* est modélisé par trois états. Dans une première étape, nous regroupons ensemble les *facteurs d'états* qui représente les différents contextes du même phoneme. nous obtenons 38 classes. Dans une deuxième étape nous utilisons les *facteurs d'états* pour classifier les états à l'intérieurs de chaque classe phonème *ph*. Le nombre de gaussiennes par état est de 256 gaussiennes. Les résultats présentés dans le tableau 5 montrent que le regroupement guidé nous permet un gain de 0,7% absolu par rapport au regroupement global. Donc nous obtenons un gain absolu 8,50% par rapport au modèle de base.

	1600s	1800s	2000s	2200s	2400s
WER	25.00	24.80	24.90	24.30	**24.20**

TABLE 5 – *Results de regroupement guidé.*

5 Conclusion

Dans cet article, nous avons proposé une nouvelle représentation vectorielle des l'états des MMC. Cette représentation est obtenue avec le paradigme d'analyse factorielle. Dans la modélisation acoustique contextuelle, nous avons exploité cette représentation vectorielle dans la réalisation de procédure de regroupement d'états des MMC. La classification d'états dans cette procédure est basée seulement sur l'information portée par les vecteurs *facteurs d'états*. L'application de cette méthode sur la langue française a monté une amélioration significative des performances. Nous avons étendu l'utilisation des *facteurs d'états* dans la modélisation acoustique pour les langues peu dotées. La modélisation acoustique pour cette classe de langues souffrent du manque de ressources informatiques et linguistiques. Avec les *facteurs d'états* nous sommes arrivés à contourner ces contraintes pour construire des modèles acoustiques dépendants du contexte ayant de bonnes performances (avec un gain relatif de 23% pour la langue vietnamienne par rapport au système de référence obtenu par la technique de l'arbre de décision standard). De plus, nous avons montré, que grâce à la représentation vectorielle, il devient possible d'avoir une visualisation graphique des états (ou des phonèmes). Cette visualisation pourrait être utile pour l'analyse de différents types de variabilités affectant les prononciations des unités phonétique. Elle peut être utilisée par de phonéticiens qui désirent étudier certains phénomènes phonologiques particuliers. Elle peut, par exemple, être utilisée dans le domaine de la phonétique clinique.

Références

ANDERSON, O., DALSGAARD, P. et BARRY, W. (1994). On the use of data-driven clustering techniques for language identification of poly and mono-phonemes for four european languages. pages 121–124. ICASSP, Adelaide.

BEULEN, K. et NEY, H. (1998). Automatic question generation for decision tree based state tying. *In Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, pages 805–809, Seattle, WA, USA.

BEYERLEIN, P. (1998). Discriminative model combination. *In ICASSP*, volume 1. Institute of Electrical Engineers (IEE).

KENNY, P., BOULIANNE, G., OUELLET, P. et DUMOUCHEL, P. (2005). Factor Analysis Simplified. *In Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP 2005)*, volume 1, pages 637–640.

LE, V. B. et BESACIER, L. (2009). Automatic speech recognition for under-resourced languages : Application to vietnamese language. *IEEE Transactions on Audio, Speech and Language Processing*, 17(8):1471–1482.

MAK, B. et BARNARD, E. (1996). Phone clustering using the bhattacharyya distance. *In ICSLP*. ISCA.

MATROUF, D., SCHEFFER, N., FAUVE, B. G. B. et BONASTRE, J.-F. (2007). A straightforward and efficient implementation of the factor analysis model for speaker verification. *In INTERSPEECH*, pages 1242–1245.

N. THI, M. H. (2006). *Outils et ressources linguistiques pour l'alignement de textes multilingues français-vietnamiens*. Thèse de doctora, Université Henri Poincaré, Nancy, France.

NONYME, A. (2013a). Un article de revue. *Une grande revue*, 2(1):20–40.

NONYME, A. (2013b). *Un livre*. Éditeur du livre.

NONYME, A. (2014). Un magnifique article. *In XXX^{e} Journées d'Études sur la Parole (JEP 2014)*, Le Mans, France.

OSTERHOLTZ, L., AUGUSTINE, C., MCNAIR, A., ROGINA, I., SAITO, H., SLOBODA, T., TEBELSKIS, J. et WAIBEL, A. (1992). Testing generality in janus : A multi-lingual speech translation system.

REICHL, W. et CHOU, W. (1998). Decision tree state tying based on segmental clustering for acoustic modeling. *In icassp*, volume 2, pages 801–804.

SCHULTZ, T. et WAIBEL, A. (2001). Language-independent and language-adaptive acoustic modeling for speech recognition. *Speech Communication*, 35(1-2):31–51.

SINGH, R., RAJ, B. et M.STERN, R. (1999). Automatic clustering and generation of contextual questions for tied states in hidden markov models. volume vol. 1, pages 117–120. International Conference on Spoken Language Processing (ICSLP).

TRAN, D.-D. (2003). Building a large vietnamese speech database. Rapport de master tic, Vietnam.

YOUNG, S. J. (1992). The general use of tying in phoneme-based hmm speech recognisers. *In ICASSP*, pages 569–572.

YOUNG, S. J. et WOODLAND, P. C. (1993). The use of state tying in continuous speech recognition. *In EUROSPEECH*. ISCA.