

---

---

# CHAPITRE X

## Résultats du post-traitement segmental

---

---

### 1 Introduction

Nous présentons dans ce chapitre les résultats obtenus sur trois corpus de données : les chiffres isolés, le Trégor et les nombres à deux chiffres. Le nombre de paramètres du post-traitement étant important, contexte/hors contexte, silence/hors silence, durée absolue/durée relative, multigaussienne (nombre de gaussiennes pour les modèles corrects et les modèles incorrects), nombre de coefficients... ; nous allons présenter les résultats les plus marquants et éliminer au fur et à mesure les combinaisons les moins prometteuses.

On va s'intéresser plus particulièrement à 5 cas de figures que nous présentons dès le début de ce chapitre. Ces 5 cas correspondent aux situations suivantes :

- 1/ contexte indépendant, sans silence et avec durée relative.
- 2/ contexte indépendant, avec silence et durée relative.
- 3/ contexte dépendant, avec silence et durée relative.
- 4/ contexte dépendant, avec silence et durée absolue.
- 5/ contexte dépendant, sans silence et avec durée absolue.

Pour mettre en évidence l'apport des modèles incorrects, on va d'abord présenter les résultats obtenus en ne tenant compte que des modèles corrects. Puis, on présente les résultats obtenus en utilisant simultanément les modèles corrects et les modèles incorrects. Par la suite, on étudie le cas d'une modélisation multigaussienne pour les modèles corrects et incorrects et on dresse un tableau comparatif pour le corpus du Trégor, entre l'utilisation des modèles corrects uniquement et l'utilisation conjointe des modèles corrects et incorrects. A la fin de ce chapitre, on résume les résultats obtenus, sur les différents corpus de données, avec les différentes configurations citées plus haut.

Les  $N$  meilleures solutions ont été développées à partir d'un modèle de Markov utilisant une modélisation par allophones (phonèmes en contexte). Le nombre de solutions générées est égal à 5.

Tous les résultats sont présentés comme suit :

- Les performances du modèle de Markov seul (HMM), quelque soit la configuration testée.
- Les performances du post-traitement seul (PTS) et les performances de la combinaison du HMM avec le PTS. Ces résultats sont fonction de la configuration testée.

Les taux d'erreur sont exprimés en % et sont présentés sur les corpus de données d'apprentissage et de test. On présente d'une part les histogrammes et d'autre part les valeurs numériques dans un tableau. Le paramètre  $\alpha$  est quand à lui optimisé sur le corpus d'apprentissage et utilisé par la suite dans le test. Pour tous les tests présentés au début de ce chapitre, on a associé à chaque modèle statistique **une** seule gaussienne à matrice de covariance diagonale.

## 2 Les 5 types d'expérience

### 2.1 Contexte indépendant, sans silence et avec durée relative

On utilise une modélisation indépendante du contexte pour chaque segment de parole (correspondant à un phonème). Les segments correspondants aux silences de début et de fin d'enregistrement ne sont pas pris en considération.

La figure X.1 illustre l'association des étiquettes aux divers segments des solutions proposées.

Pour un mot prononcé donné, on a représenté les 3 meilleures solutions, "5", "6" et "7". L'alignement a permis de définir les différents segments mis en jeu. En contexte indépendant, on va regrouper tous les vecteurs associés à l'étiquette /S/ (de même pour /E/ et toutes les autres étiquettes) dans une seule classe. Dans l'espace des étiquettes, on a représenté les différentes étiquettes utilisées (donc les segments associés) pour définir les différents modèles statistiques employés par la suite. On note qu'on a supprimé de cet espace les étiquettes correspondant aux silences (début et fin).

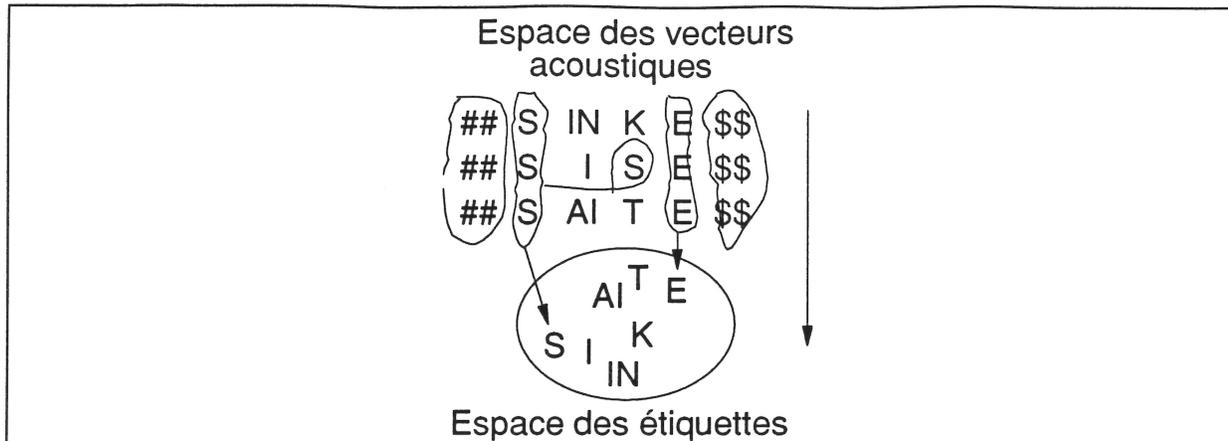


Figure X.1 Contexte indépendant, sans silence

Le score d'une solution ne tient donc pas compte du score des segments de silence. La durée de chaque segment d'une solution est normalisée par la durée globale de la solution.

## 2.2 Contexte indépendant, avec silence et durée relative

On utilise encore une modélisation indépendante du contexte pour chaque segment de parole. Par contre les segments de silence sont pris en compte.

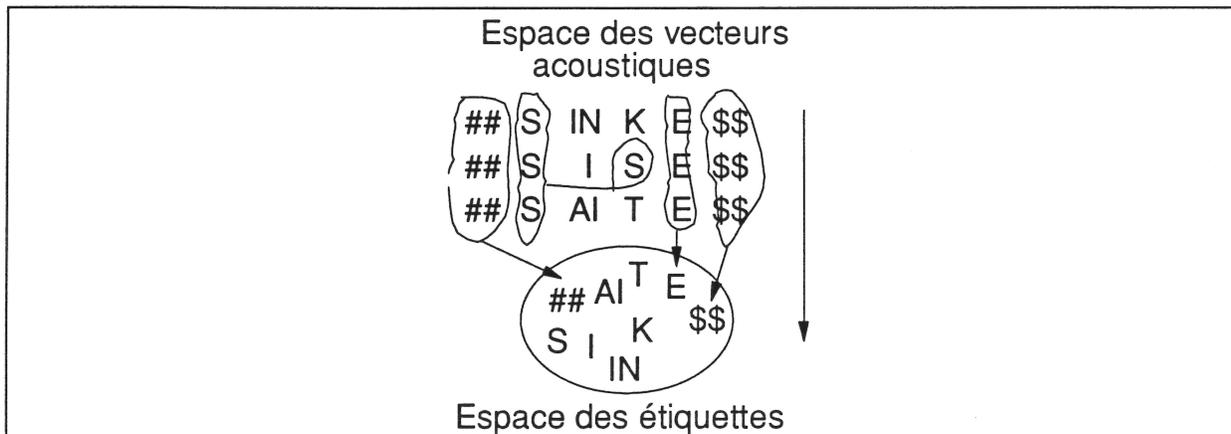


Figure X.2 Contexte indépendant, avec silence

Ainsi le calcul du score d'une solution va tenir compte du score des segments de silence. La durée de chaque segment d'une solution est normalisée par la durée globale de la solution.

### 2.3 Contexte dépendant, avec silence et durée relative

Les modèles dépendants du contexte sont choisis pour chaque segment en fonction du contexte droit et gauche. On prend en compte dans cette configuration les segments de silence. La durée relative d'un segment est elle aussi prise en compte.

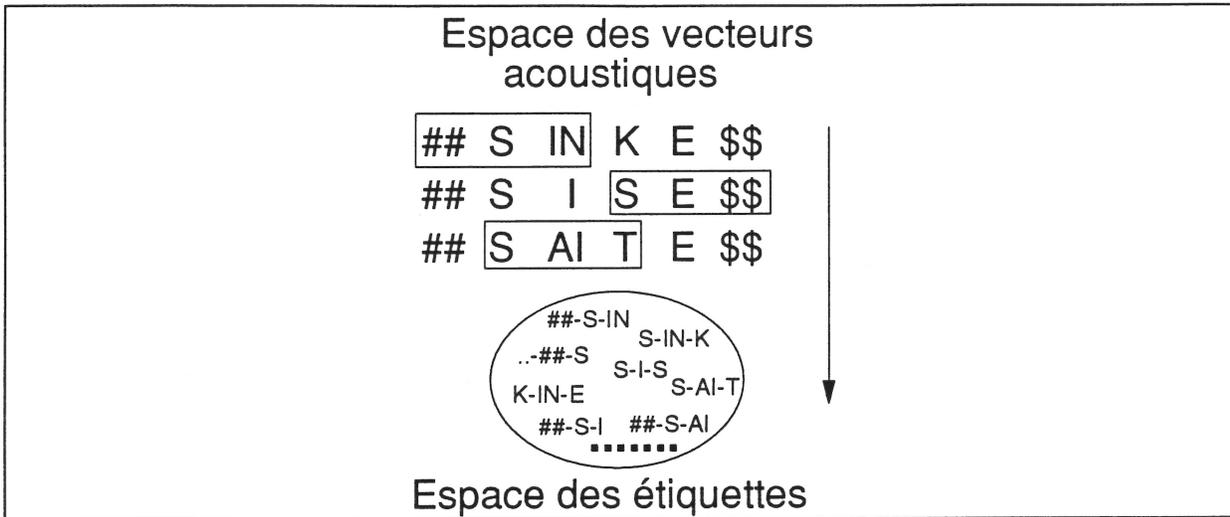


Figure X.3 Contexte dépendant, avec silence

La figure X.3 présente le cas de la prise en compte du contexte et du silence. On n'a regroupé que quelques contextes possibles pour ne pas charger la figure.

### 2.4 Contexte dépendant, avec silence et durée absolue

On conserve le contexte et le silence comme décrit précédemment. On change uniquement le paramètre durée. En effet, dans ce cas, la durée d'un segment n'est pas normalisée par la durée globale d'une solution.

### 2.5 Contexte dépendant, sans silence et avec durée absolue

Pour mesurer l'influence du silence on a choisi de garder les modèles dépendants du contexte. Par contre, on supprime les segments correspondants aux silences. Même si le silence n'est pas utilisé pour calculer le score d'une solution, il est conservé dans la définition des contextes.

### 3 Seuil minimum sur l'écart type des gaussiennes

A chaque modèle statistique est associée une fonction de densité de probabilité gaussienne. Pour éviter que ces gaussiennes soient trop sélectives (trop pointues, surapprentissage ou manques de données), on impose un seuil minimum sur l'écart type de chaque gaussienne. Ainsi, si l'écart type  $\sigma_u$  d'une gaussienne est inférieur à l'écart type global  $\sigma_{global}$  (calculé sur tout l'espace) multiplié par un seuil, la nouvelle valeur de l'écart type sera fixée à :

$$\sigma_u = \text{seuil} * \sigma_{global} \quad (\text{X.1})$$

Néanmoins, élargir trop l'écart type d'une gaussienne implique une faible précision dans la modélisation de l'ensemble des vecteurs qui lui sont associés.

Le seuil minimum a été fixé à 50%. Cette valeur a été obtenue en comparant les performances du PTS obtenues sur tous les corpus des données d'apprentissage pour divers cas de figures et pour divers seuils compris entre 25% et 75%.

### 4 Vecteur des coefficients

Le vecteur des coefficients décrit dans le chapitre IX, est issu d'un découpage en trois parties fournies par les alignements. La première partie représente le début du phonème (influencée par le contexte gauche), la seconde sa partie centrale et la troisième la fin du phonème (influencée par son contexte droit).

Une seconde technique de segmentation a été testée. Elle consistait à diviser le segment en cinq parties égales et à calculer une approximation des dérivées premières et secondes par des formules de régression linéaire sur les 5 vecteurs. Ces tests ont été réalisés avec des modèles corrects uniquement et les résultats obtenus montrent une légère détérioration par rapport à une segmentation réalisée par le modèle de Markov. En conséquence l'ensemble des tests présentés dans ce chapitre ont été réalisés en utilisant la segmentation en trois parties fournie par le modèle de Markov.

## 5 Prise en compte du modèle correct uniquement

### 5.1 Score d'une solution

Lorsque l'on utilise uniquement les modèles corrects, la probabilité d'une solution est obtenue en multipliant les probabilités obtenues pour les différents segments de la solution :

$$P\left(X_{i_j} / M_{E_{correct} w_j}\right) = \prod_{k=1}^{K_{i_j}} P\left(V_{i_j}(k) / M_{E_{correct} s_{i_j,k}}\right) \quad (X.2)$$

où  $V_{i_j}(k)$  est le vecteur acoustique du  $k^{\text{ième}}$  segment de l'observation  $X_{i_j}$ ,  $M_{E_{correct} s_{i_j,k}}$  est le modèle qui représente les statistiques correctes du segment  $s_{i_j,k}$  et  $K_{i_j}$  le nombre de segments dans l'observation  $X_{i_j}$ .

Afin de permettre la comparaison de solutions ayant des segmentations différentes, et donc des nombres de segments différents, on utilise pour la mesure du score de chaque solution le score moyen par segment, c'est-à-dire :

$$Q_{i_j} = \frac{1}{K_{i_j}} \sum_{k=1}^{K_{i_j}} \log \left[ P\left(V_{i_j}(k) / M_{E_{correct} s_{i_j,k}}\right) \right] \quad (X.3)$$

### 5.2 Paramètre de combinaison $\alpha$

Pour tous les tests effectués, la valeur optimale du paramètre de combinaison  $\alpha$  a été déterminée sur le corpus d'apprentissage. Bien souvent cette valeur ne permettait pas d'avoir les meilleures performances sur le corpus de test. C'est un problème de généralisation des modèles. La figure ci-dessous présente, sur le corpus du Trégor en configuration contexte dépendant, sans silence et durée absolue, l'influence du paramètre de combinaison  $\alpha$  sur les performances globales obtenues.

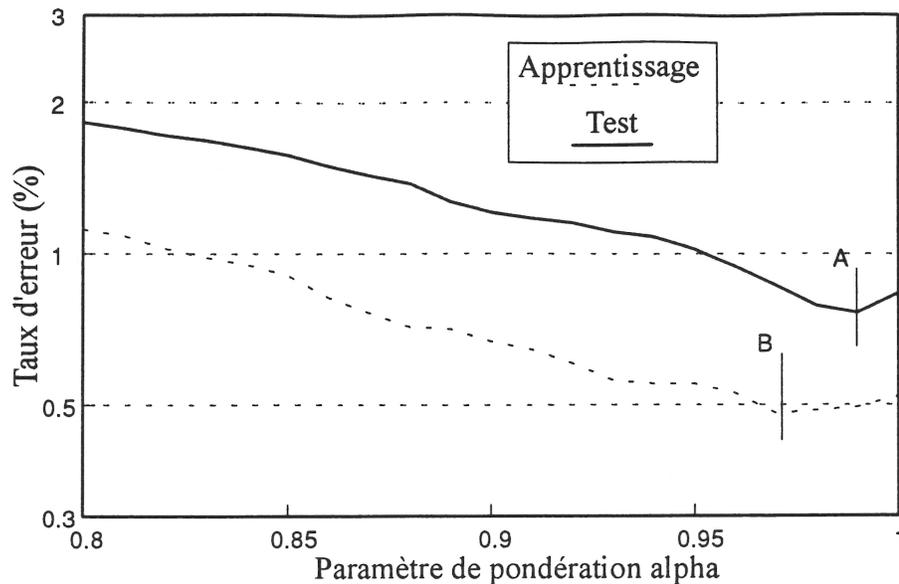


Figure X.4 Influence du paramètre de combinaison  $\alpha$  sur le taux d'erreur du corpus du Trégor ; prise en compte des modèles corrects uniquement

	$\alpha$ optimal sur l'apprentissage		$\alpha$ optimal sur le test		
	$\alpha$	HMM+PTS		$\alpha$	HMM+PTS
		APP	TST		
Contexte dépendant, sans silence et durée absolue	0.97	0.48%	0.86%	0.99	0.76%

Tableau X.1 Valeurs de  $\alpha$  et performances obtenues sur le Trégor (apprentissage et test) ; prise en compte des modèles corrects uniquement

La valeur optimale du paramètre  $\alpha$  sur le corpus d'apprentissage était de 0.97 et sur celui de test, elle était de 0.99. Cette légère différence introduit une perte de performances puisque dans ce cas avec  $\alpha=0.97$  les performances sur le corpus du test sont : 0.86 %. Il faut cependant noter que, pour un intervalle de confiance de 95 %, ces deux valeurs sont équivalentes. Mais comme l'estimation du paramètre  $\alpha$  doit se faire sur un corpus de données d'apprentissage, il est important que la valeur estimée soit proche de l'optimal pour les données de test.

### 5.3 Contexte indépendant sans silence et avec durée relative

L'histogramme ci-dessous présente les performances de reconnaissance obtenues sur les divers corpus (partie test) et le tableau représente les taux d'erreur sur l'ensemble d'apprentissage (APP) et de test (TST).

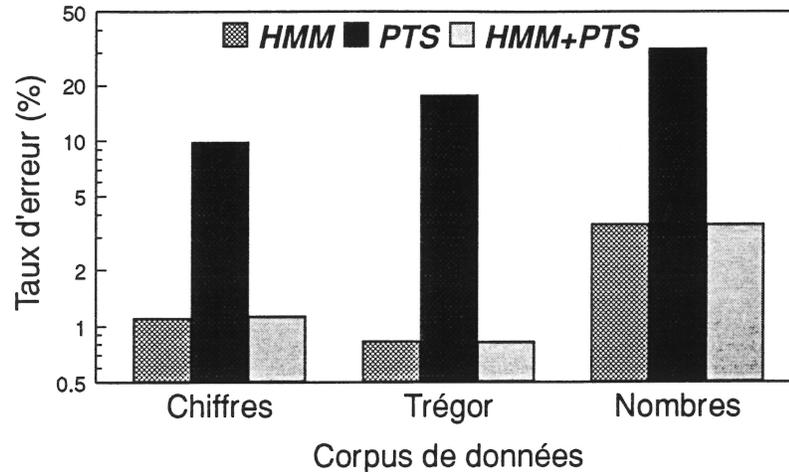


Figure X.5 Performances en configuration contexte indépendant, sans silence et avec durée relative ; prise en compte des modèles corrects uniquement

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	10.02%	9.82%	16.56%	17.53%	30.95%	31.53%
HMM+PTS	0.65%	1.13%	0.50%	0.82%	2.38%	3.51%

Tableau X.2 Performances en configuration contexte indépendant, sans silence et avec durée relative ; prise en compte des modèles corrects uniquement

On remarque que le taux d'erreur sur le PTS est très élevé pour les trois corpus de données sur l'apprentissage et sur le test. A la combinaison des scores on a obtenu un gain uniquement sur les corpus d'apprentissage des chiffres isolés. Dans tous les autres cas le taux d'erreur reste constant.

## 5.4 Contexte indépendant avec silence et durée relative

On présente ci-dessous les performances obtenues en tenant compte des segments de silence et en conservant des modèles indépendants du contexte avec la prise en compte de la durée relative des segments. L'histogramme présente les performances obtenues sur les corpus des chiffres isolés, du Trégor et des nombres à deux chiffres sur la partie test. Dans le tableau, on a présenté pour les trois corpus de données les performances obtenues sur l'apprentissage et le test.

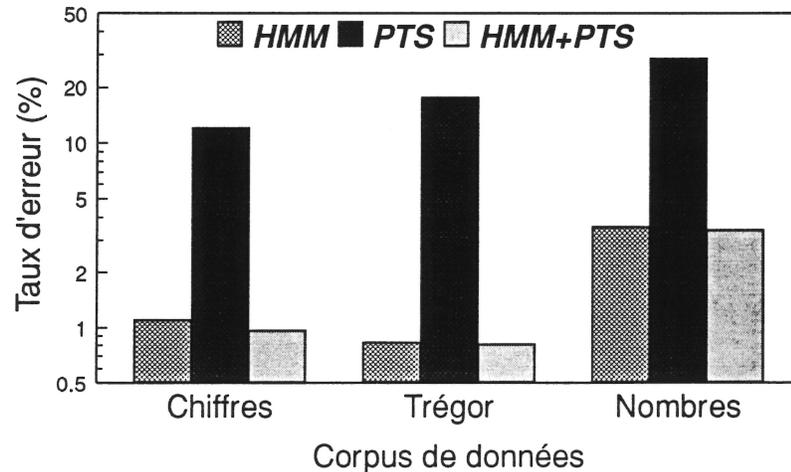


Figure X.6 Performances en configuration contexte indépendant, avec silence et durée relative ; prise en compte des modèles corrects uniquement

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	12.72%	12.04%	16.96%	17.60%	27.32%	28.43%
HMM+PTS	0.59%	0.96%	0.52%	0.81%	2.37%	3.37%

Tableau X.3 Performances en configuration contexte indépendant, avec silence et durée relative ; prise en compte des modèles corrects uniquement

Au niveau du post-traitement segmental, on a eu une légère dégradation entre cette configuration et la configuration précédente sur le corpus des chiffres isolés. Par contre on a obtenu une légère amélioration sur le corpus des nombres.

La combinaison des scores a permis d'améliorer les performances pour les différents corpus de données, par rapport au HMM seul. Les segments de silence apportent donc une information utile.

### 5.5 Contexte dépendant avec silence et durée relative

Pour ces tests on utilise une modélisation contextuelle des segments. L'histogramme regroupe les résultats pour les différents corpus de données (test uniquement) et le tableau détaille les résultats pour les corpus d'apprentissage et de test.

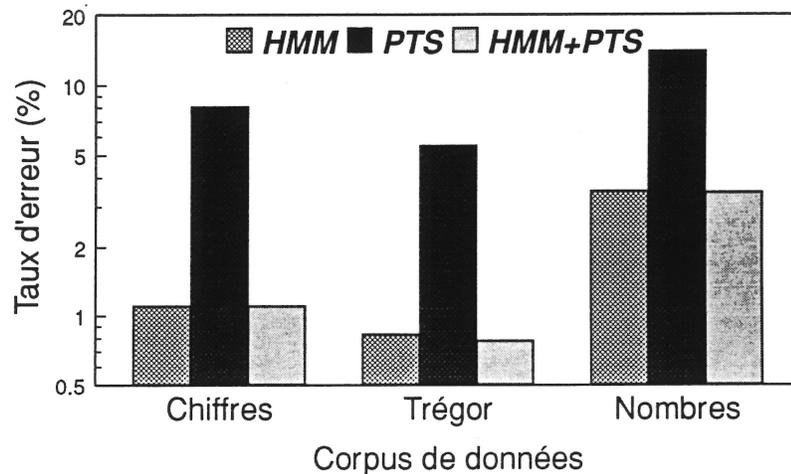


Figure X.7 Performances en configuration contexte dépendant, avec silence et durée relative ; prise en compte des modèles corrects uniquement

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	7.71%	8.06%	4.10%	5.48%	9.96%	13.96%
HMM+PTS	0.65%	1.10%	0.50%	0.78%	2.11%	3.47%

Tableau X.4 Performances en configuration contexte dépendant, avec silence et durée relative ; prise en compte des modèles corrects uniquement

En comparant les résultats obtenus sur le PTS entre cette configuration et la précédente, on observe de fortes améliorations dues à la prise en compte du contexte. Malheureusement cette amélioration du PTS seul ne se retrouve pas au niveau du système global.

## 5.6 Contexte dépendant avec silence et durée absolue

Cette série d'expériences a été effectuée en utilisant la durée absolue des segments c'est-à-dire sans normalisation. On représente de la même manière les performances de notre système (histogramme et tableau).

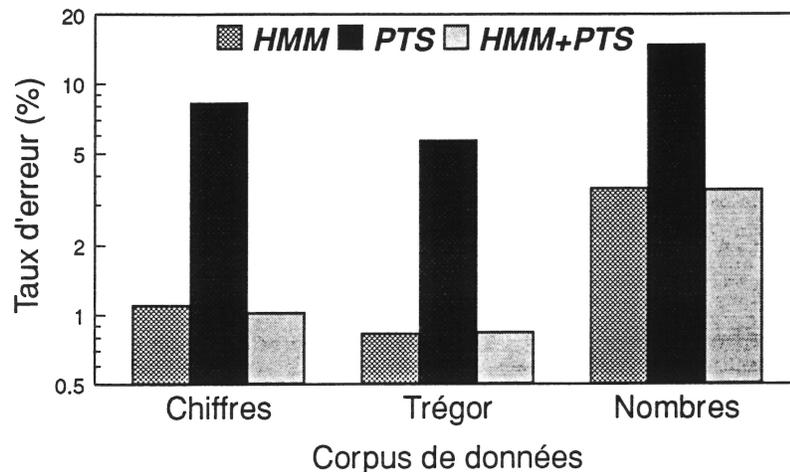


Figure X.8 Performances en configuration contexte dépendant, avec silence et durée absolue ; prise en compte des modèles corrects uniquement

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	7.49%	8.25%	4.35%	5.67%	10.34%	14.69%
HMM+PTS	0.59%	1.02%	0.49%	0.84%	2.11%	3.47%

Tableau X.5 Performances en configuration contexte dépendant, avec silence et durée absolue ; prise en compte des modèles corrects uniquement

On remarque qu'il y a une légère dégradation sur PTS seul, du fait de l'introduction de la durée absolue. Sur la combinaison une légère amélioration est observée uniquement sur les chiffres isolés. Ce type de post-traitement conduit à une légère amélioration, par rapport au HMM seul, pour le corpus des chiffres et le corpus d'apprentissage des nombres.

## 5.7 Conclusion

Nous avons présenté les résultats pour 4 combinaisons des paramètres de post-traitement. Pour ne pas trop tarder dans ce paragraphe nous présentons en annexe 5 les deux configurations : contexte dépendant, sans silence et avec durée absolue ; et contexte dépendant sans silence et avec durée relative. Leurs résultats étaient relativement moins bons que ceux présentés ici sauf pour le PTS en contexte dépendant sans silence et avec durée absolue où on avait une très légère amélioration sur les différents corpus de données.

Globalement l'utilisation des modèles corrects uniquement permettaient d'obtenir une réduction du taux d'erreur inférieure à 10%. Cette réduction n'était pas suffisante pour sortir de l'intervalle de confiance de 95%. Par ailleurs, pour le paramètre de combinaison  $\alpha$ , on a constaté qu'il y avait un problème d'optimalité entre l'apprentissage et le test, et cela même en utilisant un nombre de paramètres très réduits (1 gaussienne par segment). Nous allons présenter dans ce qui suit la prise en compte simultanée des modèles corrects et des modèles incorrects.

## 6 Prise en compte des modèles corrects et incorrects

Etant donné qu'il est difficile, au vu des résultats précédents, de déterminer l'influence de chacun des paramètres de post-traitement, nous allons comparer de nouveau ces configurations en utilisant conjointement les modèles corrects et les modèles incorrects.

Les résultats présentés correspondent à la partie théorique du chapitre IX. Avant de développer les résultats, nous rappelons la définition des segments "incorrects".

### 6.1 Modélisation des segments incorrects

Un segment incorrect est un segment correspondant à une solution incorrecte. Avant de déclarer un segment comme incorrect, on commence par rechercher s'il a des liens avec le mot correctement reconnu (donc avec la solution correcte). On exclut tous les segments incorrects des solutions incorrectes s'ils ont un recouvrement (voir chapitre IX paragraphe 3.1) supérieur à 75% par rapport aux segments corrects. Les segments ainsi obtenus vont servir pour estimer les paramètres des modèles statistiques incorrects associés.

## 6.2 Paramètre de combinaison $\alpha$

Dans presque tous les cas la valeur optimale du paramètre de combinaison  $\alpha$  coïncidait entre le test et l'apprentissage. Ainsi la normalisation par des modèles corrects et des modèles incorrects offre une plus grande robustesse pour le modèle de combinaison.

Nous présentons comme exemple la configuration contexte dépendant avec silence et durée absolue, pour le corpus des nombres.

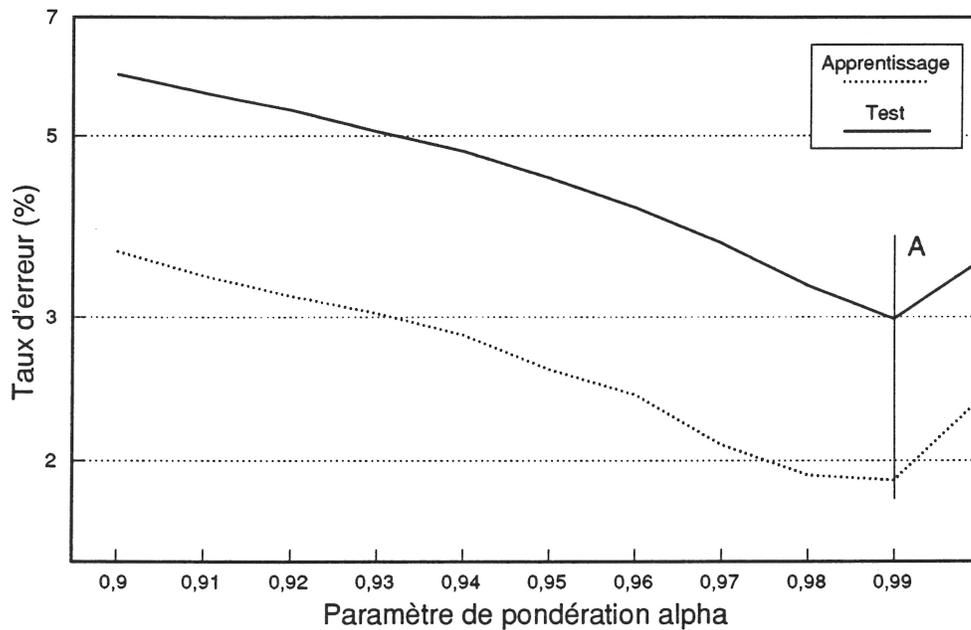


Figure X.9 Influence du paramètre de combinaison  $\alpha$  ;  
prise en compte des modèles corrects et incorrects

Sur la figure le trait A montre la coïncidence de l'optimalité entre l'apprentissage et le test.

### 6.3 Contexte indépendant sans silence et avec durée relative

De même que précédemment, l'histogramme ci-dessous présente les performances de reconnaissance obtenues sur les divers corpus (partie test) et le tableau indique les taux d'erreur sur les ensembles d'apprentissage (APP) et de test (TST).

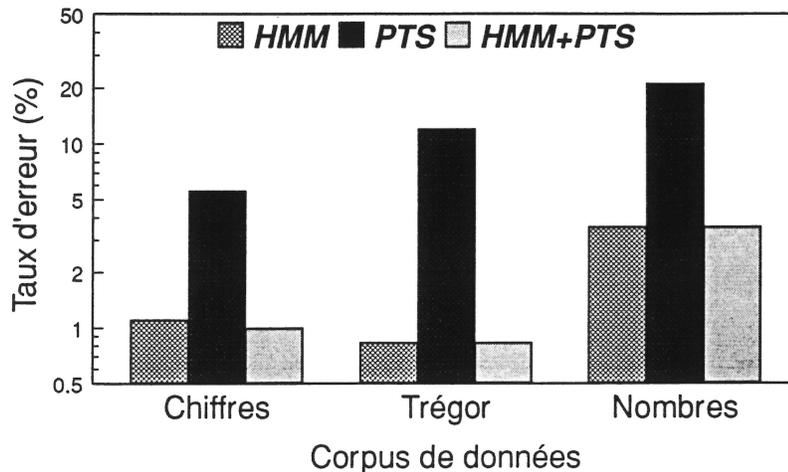


Figure X.10 Performances en configuration contexte indépendant, sans silence et durée relative ; prise en compte des modèles corrects et incorrects

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	4.70%	5.57%	11.16%	11.96%	18.46%	21.00%
HMM+PTS	0.59%	0.99%	0.52%	0.83%	2.38%	3.51%

Tableau X.6 Performances en configuration contexte indépendant, sans silence et avec durée relative ; prise en compte des modèles corrects et incorrects

Les résultats de PTS seul sont encore nettement moins bons que ceux du HMM seul. Toutefois, ils sont nettement meilleurs que ceux obtenus avec l'emploi des modèles corrects uniquement. Après la combinaison, on constate des améliorations sur les chiffres uniquement.

## 6.4 Contexte indépendant avec silence et durée relative

Dans ce cas on tient compte dans le calcul du score d'une solution des scores de silence. On reste cependant en contexte indépendant et on utilise la durée relative.

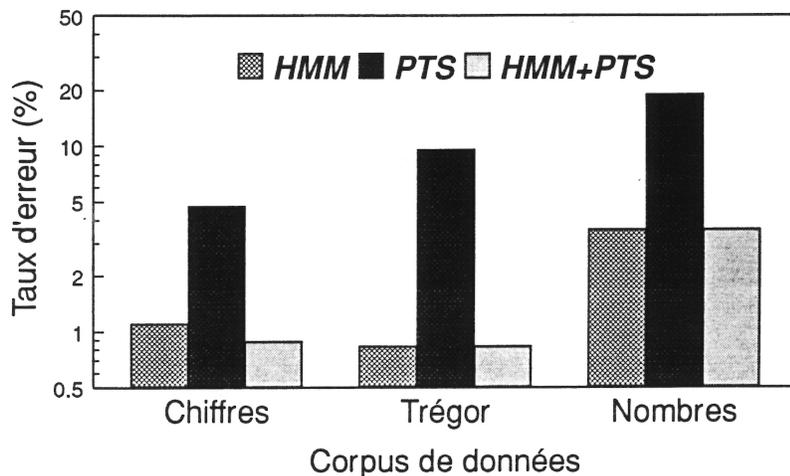


Figure X.11 Performances en configuration contexte indépendant, avec silence et durée relative ; prise en compte des modèles corrects et incorrects

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	4.24%	4.69%	8.72%	9.47%	16.66%	18.95%
HMM+PTS	0.56%	0.88%	0.52%	0.83%	2.38%	3.51%

Tableau X.7 Performances en configuration contexte indépendant, avec silence et durée relative ; prise en compte des modèles corrects et incorrects

On constate des améliorations par rapport à la configuration précédente au niveau du PTS seul. Après la combinaison on a une amélioration sur le corpus des chiffres uniquement mais rien sur les autres corpus de données.

## 6.5 Contexte dépendant avec silence et durée relative

On teste maintenant l'apport du contexte sur le système de post-traitement segmental. On présente les résultats de la même manière que précédemment (histogramme et tableau).

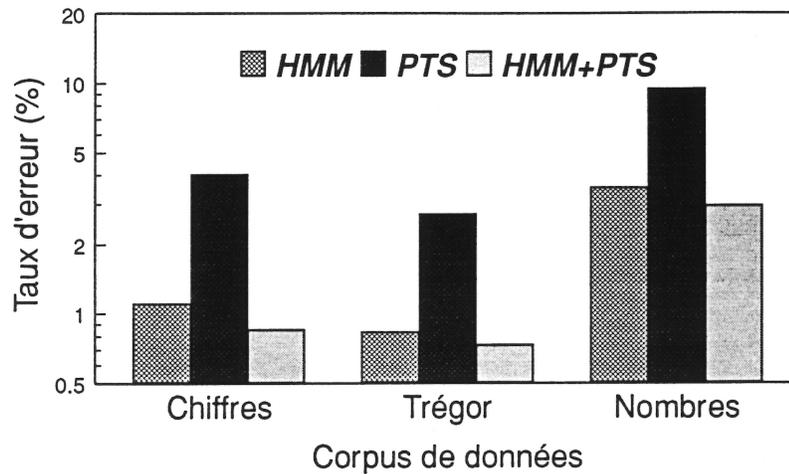


Figure X.12 Performances en configuration contexte dépendant, avec silence et durée relative ; prise en compte des modèles corrects et incorrects

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	3.23%	4.00%	1.95%	2.69%	6.17%	9.45%
HMM+PTS	0.56%	0.85%	0.38%	0.73%	1.89%	2.94%

Tableau X.8 Performances en configuration contexte dépendant, avec silence et durée relative ; prise en compte des modèles corrects et incorrects

On constate cette fois-ci une nette amélioration sur le PTS pour les corpus du Trégor et des nombres. Au niveau de la combinaison des scores, cette configuration permet d'améliorer par rapport au HMM seul.

L'amélioration par rapport à la configuration précédente est beaucoup plus faible sur le corpus des chiffres que pour les autres corpus. Cela s'explique par le fait que le corpus des chiffres implique peu de variantes contextuelles de chaque phonème. La réduction du taux d'erreur par rapport au HMM seul varie entre 12% et 20% suivant les corpus.

## 6.6 Contexte dépendant avec silence et durée absolue

Cette série de tests a pour but de mesurer l'influence de la normalisation de la durée des segments. Les résultats sont toujours présentés sous forme d'histogramme et de tableau.

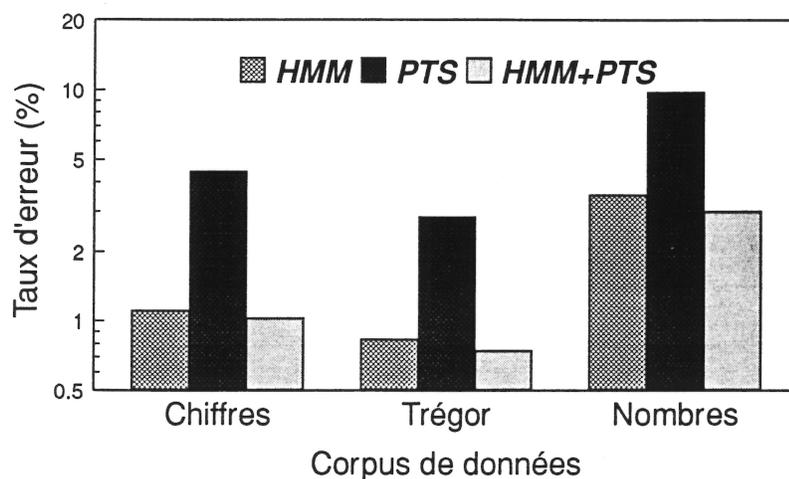


Figure X.13 Performances en configuration contexte dépendant, avec silence et durée absolue ; prise en compte des modèles corrects et incorrects

	Chiffres isolés		Trégor		Nombre à 2 chiffres	
	APP	TST	APP	TST	APP	TST
HMM	0.76%	1.10%	0.52%	0.83%	2.38%	3.51%
PTS	3.40%	4.39%	2.12%	2.81%	6.34%	9.78%
HMM+PTS	0.59%	1.02%	0.38%	0.74%	1.89%	2.98%

Tableau X.9 Performances en configuration contexte dépendant, avec silence et durée absolue ; prise en compte des modèles corrects et incorrects

On constate, par rapport à la configuration précédente, qu'on dégrade à la fois au niveau du PTS seul et de la combinaison HMM+PTS. Il faut noter cependant que l'on conserve une amélioration significative par rapport au HMM seul.

Cette dégradation met en évidence l'intérêt de la normalisation de la durée des segments, dont le but est de rendre les paramètres moins sensibles à la vitesse d'élocution.

## 6.7 Conclusion

Nous avons présenté les résultats pour 4 configurations types. Les résultats obtenus pour les configurations : contexte dépendant hors silence et avec durée relative ; et contexte dépendant hors silence et avec durée absolue, sont présentés en annexe 6. Cependant il est à signaler qu'elles n'apportent pas d'améliorations par rapport aux configurations présentées plus haut. De ce fait le silence, la durée relative et le contexte sont des caractéristiques à prendre en considération pour le post-traitement segmental. Au niveau des performances, les résultats obtenues montrent une réduction du taux d'erreur entre 10 et 20% suivant la configuration testée. Cette réduction nous permet dans certains cas de sortir de l'intervalle de confiance de 95%. Pour le paramètre de combinaison  $\alpha$ , sur l'ensemble des tests effectués, on a constaté que la valeur optimale entre le corpus d'apprentissage et de test coïncidaient dans **TOUS** les cas, et cela bien sûr en associant une gaussienne par segment correct ou segment incorrect, ce qui est très faible pour la modélisation. Nous allons introduire dans ce qui suit plusieurs gaussiennes pour les différents modèles statistiques utilisés, et voir si on obtient des améliorations significatives au sens de l'intervalle de confiance.

## 7 Modélisation avec des multigaussiennes

### 7.1 Technique de clustering

Les multigaussiennes sont obtenues en divisant l'espace des vecteurs acoustiques en un arbre binaire [Gray, 82]. On descend d'un niveau à un autre si le nombre de données par gaussienne est supérieur à un certain seuil. On choisit de minimiser une distance quadratique moyenne entre les vecteurs des coefficients et les vecteurs de code (centroïde de la classe) correspondants afin de s'assurer la convergence du clustering [Duda, 73][Chandon, 81].

La technique de clustering passe par les étapes suivantes :

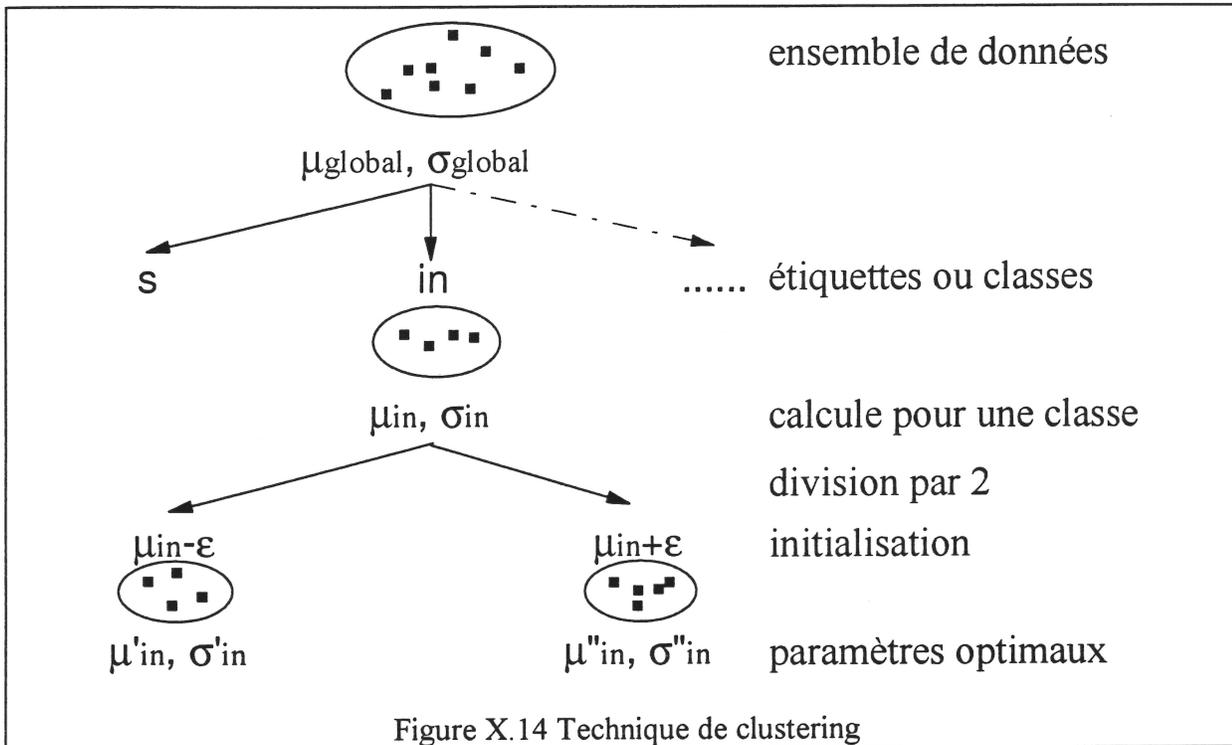
- Calculer la moyenne et l'écart-type de l'ensemble de données, dans notre cas l'ensemble des vecteurs acoustiques.
- Choisir une configuration donnée, par exemple contexte indépendant avec silence et durée relative.
- Pour chaque étiquette, on va regrouper tous les vecteurs des coefficients associés à cette étiquette (appelée aussi une classe). On calcule par la suite pour cette classe sa moyenne et son écart-type.
- Si le nombre de données par classe est suffisant pour contenir plus d'une gaussienne, on associe à la classe deux gaussiennes. La moyenne de chaque gaussienne est initialisée en agitant

d'un  $\epsilon$  de part et d'autre la moyenne initiale (i.e. la moyenne de la gaussienne d'origine) de la classe.

- Il reste à apprendre les paramètres des 2 gaussiennes et cela en minimisant la variance intraclasse.

L'apprentissage s'arrête dès que la différence des moyennes de chaque classe entre deux itérations successives est inférieure à un  $\delta$ .

Un exemple sur la figure X.14 illustre cette technique de clustering.



Nous allons développer les résultats pour le corpus des données de test du Trégor pour deux configurations, d'une part contexte dépendant, sans silence<sup>1</sup> et avec durée absolue et d'autre part contexte indépendant, sans silence et avec durée absolue.

<sup>1</sup> La configuration avec silence et durée relative est présentée en annexe. Nous présentons ses résultats dans un tableau comparatif à la fin de ce chapitre.

## 7.2 Contexte dépendant, sans silence et avec durée absolue

### PTS seul

L'histogramme ci-dessous présente les taux d'erreur obtenus sur le corpus de test du Trégor. L'axe des x correspond au nombre de gaussiennes utilisées pour les modèles corrects, l'axe des y correspond au nombre de gaussiennes employées pour les modèles incorrects et finalement l'axe des z correspond aux taux d'erreur obtenus sur le test (sur l'apprentissage en annexe 7). Cet histogramme présente les résultats du PTS seul.

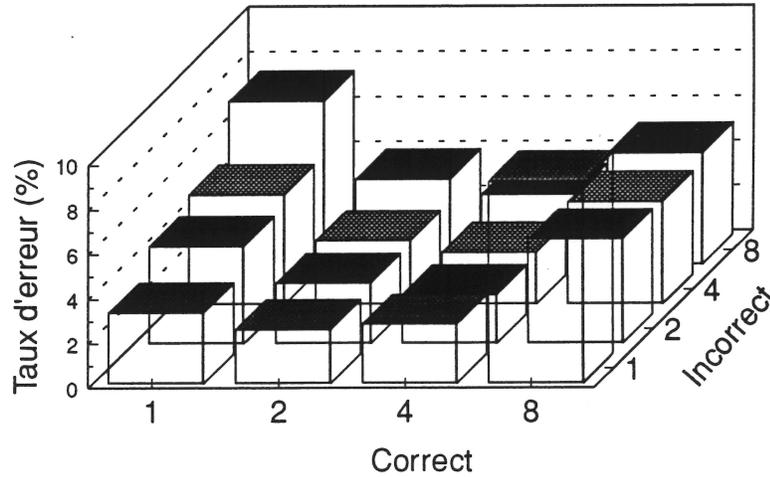


Figure X.15 Performances du PTS sur le corpus de test du Trégor en configuration contexte dépendant, sans silence, durée absolue et multigaussiennes ; prise en compte des modèles corrects et incorrects

Nombre de gaussiennes pour les modèles corrects	Nombre de gaussiennes pour les modèles incorrects			
	1	2	4	8
1	3.18%	4.40%	4.94%	7.35%
2	2.41%	2.73%	2.87%	3.85%
4	2.65%	2.21%	2.34%	3.17%
8	9.03%	4.69%	4.59%	4.97%

Tableau X.10 Performances du PTS sur le corpus de test du Trégor en configuration contexte dépendant, sans silence, durée absolue et multigaussiennes ; prise en compte des modèles corrects et incorrects

En regardant les résultats pour 8 gaussiennes (corrects ou incorrects) nous remarquons une détérioration des performances qui s'explique par l'obtention de gaussiennes très pointues.

Sinon la tendance générale du tableau montre une amélioration lorsque les nombres de gaussiennes sont équilibrés pour les modèles corrects et incorrects. Ainsi nous obtenons les meilleurs résultats avec 4 gaussiennes pour le modèle des corrects et 2 gaussiennes pour le modèle des incorrects. Le taux d'erreur est passé de 3.2% dans le cas d'une gaussienne de part et d'autre à 2.2% soit une réduction du taux d'erreur de 30%, ce qui est significatif pour un intervalle de confiance de 95%<sup>1</sup>.

### **HMM+PTS**

Nombre de gaussiennes pour les modèles corrects	Nombre de gaussiennes pour les modèles incorrects			
	1	2	4	8
1	0.72%	0.77%	0.77%	0.77%
2	0.72%	0.74%	0.74%	0.75%
4	0.71%	0.73%	0.70%	0.75%
8	0.75%	0.76%	0.79%	0.77%

Tableau X.11 Performances du HMM+PTS sur le corpus de test du Trégor, en configuration contexte dépendant, sans silence, durée absolue et multigaussiennes ; prise en compte des modèles corrects et incorrects

Lors de la combinaison HMM+PTS, aucune amélioration significative des performances n'est observée. Toutes les combinaisons semblent équivalentes sur le test.

### **7.3 Contexte indépendant, sans silence et avec durée absolue**

On voulait voir si le fait de prendre une modélisation en contexte indépendant avec des multigaussiennes peut rivaliser du point de vue des performances, sur le corpus test du Trégor, avec une configuration en contexte dépendant.

Les résultats du PTS sont présentés dans un espace à 3 dimensions, les axes ont la même signification que précédemment.

---

<sup>1</sup> Le nombre d'échantillons est de 12842 (corpus test du Trégor). Le taux d'erreur initial est de 3.2. L'intervalle de confiance à 95%, dans ce cas est : [2.91, 3.52]. On constate bien que 2.2 est à l'extérieure de cet intervalle.

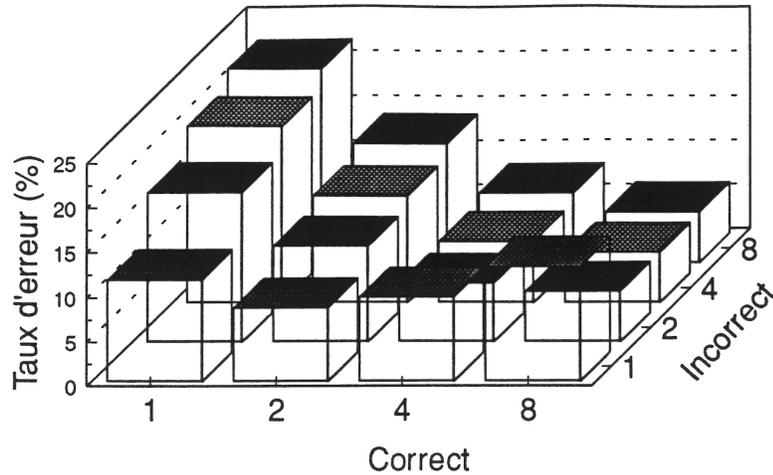
**PTS seul**

Figure X.16 Performances du PTS sur le corpus test du Trégor en configuration contexte indépendant, sans silence, durée absolue et multigaussiennes ; prise en compte des modèles corrects et incorrects

Nombre de gaussiennes pour les modèles corrects	Nombre de gaussiennes pour les modèles incorrects			
	1	2	4	8
1	10.10	16.76	19.77	21.82
2	8.27	10.75	11.92	13.47
4	9.48	6.62	6.87	7.87
8	12.92	5.7	5.63	5.72

Tableau X.12 Performances du PTS sur le corpus test du Trégor, en configuration contexte indépendant, sans silence, durée absolue et multigaussiennes ; prise en compte des modèles corrects et incorrects

Les mêmes remarques s'appliquent ici en ce qui concerne l'équilibre entre les nombres de gaussiennes pour les modèles corrects et les modèles incorrects. Néanmoins quelques analyses supplémentaires sont intéressantes à faire.

Sur ces résultats nous remarquons que 8 gaussiennes pour les deux modèles ne détériorent pas les performances. Ceci s'explique par le nombre d'étiquettes plus réduit que dans le cas précédent et de ce fait un ensemble de données plus large par étiquette. L'augmentation du nombre de gaussiennes des modèles corrects entraîne une amélioration des performances plus importante que l'augmentation du nombre de gaussiennes des modèles incorrects. Ce qui est parfaitement raisonnable vu la nature non contextuelle de la modélisation qui demande une

description plus fine de l'espace des segments corrects. Sur le PTS seul, le minimum est obtenu en associant 8 gaussiennes pour les modèles corrects et 2 à 8 gaussiennes pour les modèles incorrects. Le taux d'erreur dans ce cas, 5.7%, est loin des 2.1% (4 correctes et 2 incorrectes) réalisé dans le cas d'une modélisation dépendante du contexte.

### **HMM+PTS**

Le tableau ci-dessous donne les performances obtenues sur le corpus de test du Trégor en combinant les scores des HMM et ceux obtenus par le PTS.

Nombre de gaussiennes pour les modèles corrects	Nombre de gaussiennes pour les modèles incorrects			
	1	2	4	8
1	0.83*	0.83*	0.83*	0.83*
2	0.86	0.83*	0.83*	0.83*
4	0.84	0.86	0.86	0.85
8	0.81	0.81	0.85	0.81

Tableau X.13 Performances HMM+PTS pour le corpus de test du Trégor, en configuration contexte dépendant, sans silence, durée absolue et multigaussiennes ; prise en compte des modèles corrects et incorrects

A noter que les valeurs marquées par un \* sont obtenues avec un paramètre optimal égal à 1, c'est-à-dire le HMM seul.

Là aussi on n'observe aucune amélioration significative au niveau de la combinaison des scores.

## **7.4 Résultats sur les autres corpus de données**

Sur les autres corpus de données, pour une configuration contexte dépendant hors silence et durée absolue, les mêmes constatations sont observées. Les performances sont résumées dans le tableau suivant. (n,m) indique le nombre optimal de gaussiennes déterminé sur l'ensemble d'apprentissage (n gaussiennes pour des modèles corrects et m gaussiennes pour les modèles incorrects).

	Chiffres isolés	Trégor	Nombres à deux chiffres
HMM	1.10%	0.83%	3.51%
PTS [1 gaus.]	4.80%	3.18%	10.11%
PTS [multi. gaus.]	2.73% (2-2)	2.21% (4-2)	8.01% (2-2)
HMM+PTS [1 gaus.]	0.94%	0.72%	2.96%
HMM+PTS [multi. gaus.]	0.85% (4-4)	0.70% (4-4)	2.96% (2-1)

Tableau X.14 Apport des multigaussiennes sur les performances du système de reconnaissance ; prise en compte des modèles corrects et incorrects

On remarque que l'apport des multigaussiennes sur le PTS seul est significatif puisque sur les trois corpus on a des améliorations de 20 à 40% par rapport à l'utilisation du PTS seul sans multigaussiennes (i.e. 1 seule gaussienne). On constate aussi une amélioration pour la combinaison HMM+PTS, mais cette amélioration est plus faible.

## 8 Corrects vs Corrects/Incorrects

On compare ici les performances obtenues en tenant compte des modèles corrects d'une part et des modèles corrects et incorrects d'autre part. Nous avons pris le corpus des données test du Trégor comme exemple pour cette comparaison. Nous présentons le taux d'erreur obtenu pour chaque modélisation.

### - PTS seul (1 gaus.)

Configuration	correct	correct/incorrect
Contexte indépendant, sans silence et durée relative	17.53%	11.96%
Contexte indépendant, avec silence et durée relative	17.60%	9.47%
Contexte dépendant, avec silence et durée relative	5.48%	2.69%
Contexte dépendant, avec silence et durée absolue	5.67%	2.81%

Tableau X.15 Comparaison correct vs correct/incorrect, pour PTS seul

Les résultats montrent que l'introduction des modèles incorrects apporte une nette amélioration par rapport aux modèles corrects considérés seuls. Pour les configurations contexte dépendant, le taux d'erreur est divisé par 2.

**- Combinaison HMM+PTS (1 gaus.)**

Configuration	correct	correct/incorrect
Contexte indépendant, sans silence et durée relative	0.82%	0.83*%
Contexte indépendant, avec silence et durée relative	0.81%	0.83*%
Contexte dépendant, avec silence et durée relative	0.78%	0.73%
Contexte dépendant, avec silence et durée absolue	0.84%	0.74%

Tableau X.16 Comparaison correct vs correct/incorrect, pour HMM+PTS

\* le paramètre optimal  $\alpha$  a été trouvé pour une valeur égale à 1, c'est-à-dire HMM seul.

Alors que les modèles corrects amélioreraient difficilement les performances globales quelque soit la configuration choisie, l'utilisation des modèles corrects et incorrects permet de réduire le taux d'erreur de plus de 12% (par rapport à l'utilisation du HMM seul) pour les configurations contexte dépendant.

## 9 Conclusion

Nous avons présenté dans la première partie de ce chapitre, les résultats obtenus en utilisant uniquement les modèles corrects. Pour ce type de modélisation, nous avons obtenu une réduction du taux d'erreur de l'ordre de 10% en moyenne. Ce résultat ne nous permettait pas de sortir de l'intervalle de confiance fixé lui à 95% ; ajouté à cela le problème lié à l'optimalité du paramètre  $\alpha$ . Nous avons montré que la valeur optimale du paramètre  $\alpha$  entre l'apprentissage et le test ne coïncidait pas dans beaucoup de cas, malgré l'utilisation d'un nombre de paramètres très réduit (1 gaussienne par segment). De ce fait, ces différents tests, ne nous ont pas permis de dégager une configuration type.

En raison de ces résultats, et parce que nous sommes en présence d'un ensemble d'observations (voir chapitre IX) ; nous avons introduit conjointement les modèles corrects et les modèles incorrects.

Pour une modélisation très réduite (1 gaussienne par segment), les résultats obtenus montrent une réduction du taux d'erreur entre 10 et 20%. Ces résultats nous ont permis de dégager deux types de configurations pour y introduire une modélisation très fine (plus de gaussiennes). Ces deux configurations sont : contexte dépendant avec silence et durée relative ; contexte dépendant sans silence et avec durée absolue.

Le passage aux multigaussiennes nous a permis d'améliorer surtout les performances du corpus du Trégor, à cause de la présence de beaucoup de phénomènes de coarticulation.

Concernant le paramètre de combinaison  $\alpha$ , nous avons constaté, sur les différents test réalisés, que l'optimalité était garantie pour une modélisation très réduite. Dans le cas d'une modélisation très fine, le paramètre de combinaison était sensible au nombre de gaussiennes associées aux modèles statistiques.

On résume dans le tableau suivant les performances obtenues en combinant HMM+PTS sur les trois corpus de données par rapport à l'utilisation du HMM seul. Le tableau rapporte les pourcentages de réduction du taux d'erreur par rapport au HMM seul. La meilleure modélisation conduit à une réduction du taux d'erreur variant de 16% à 23% selon les corpus.

Configurations	Chiffres isolés	Trégor	Nombres à deux chiffres
Contexte indépendant, sans silence et durée relative	10%	0%	0%
Contexte indépendant, avec silence et durée relative	20%	0%	0%
Contexte dépendant, avec silence et durée relative	23%	12%	16%
<b>Contexte dépendant, avec silence durée relative et multigaussiennes</b>	<b>23%</b>	<b>16%</b>	<b>18%</b>
Contexte dépendant, avec silence et durée absolue	7%	11%	15%
Contexte dépendant, sans silence et durée absolue	15%	13%	16%
<b>Contexte dépendant, sans silence, avec durée absolue et multigaussiennes</b>	<b>23%</b>	<b>16%</b>	<b>16%</b>

Tableau X.17 Revue générale des performances du HMM+PTS  
(Réduction du taux d'erreur par rapport au HMM seul) ; prise en compte  
des modèles corrects et incorrects

Dans le cas d'une modélisation très fine, les résultats obtenus entre les deux configurations : contexte dépendant sans silence avec durée absolue et contexte dépendant avec silence durée relative sont sensiblement équivalents. Cependant pour une modélisation multigaussienne, la seconde configuration nécessite moins de paramètres que la première configuration (voir annexe 7 pour plus de détails).