

ircamAlign

Système de segmentation de signaux de parole

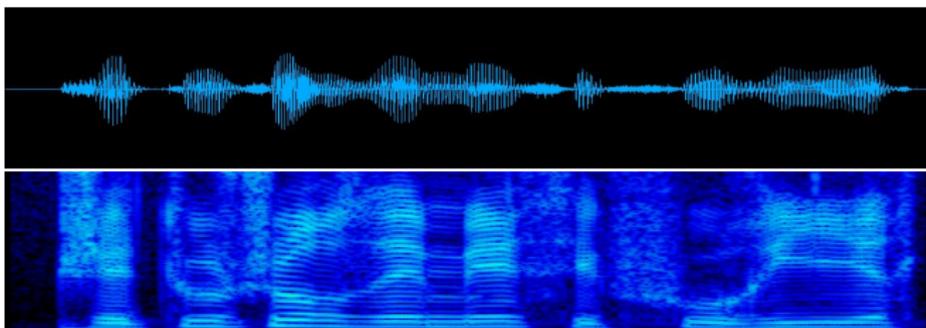
P. Lanchantin, A. C. Morris, X.Rodet

Département Analyse/Synthèse

19 septembre 2007

Introduction

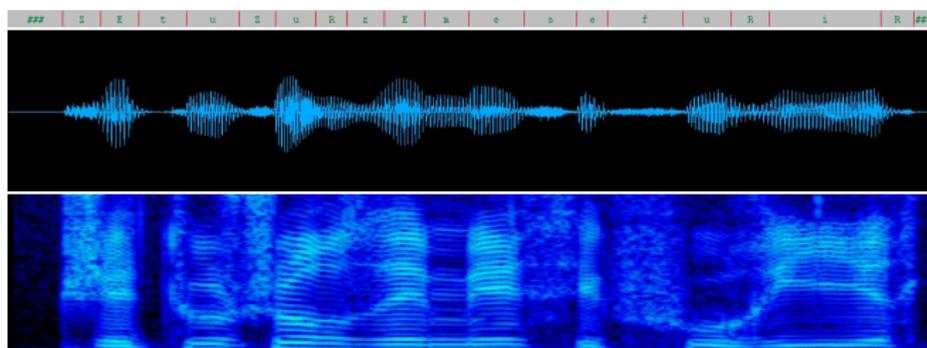
- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

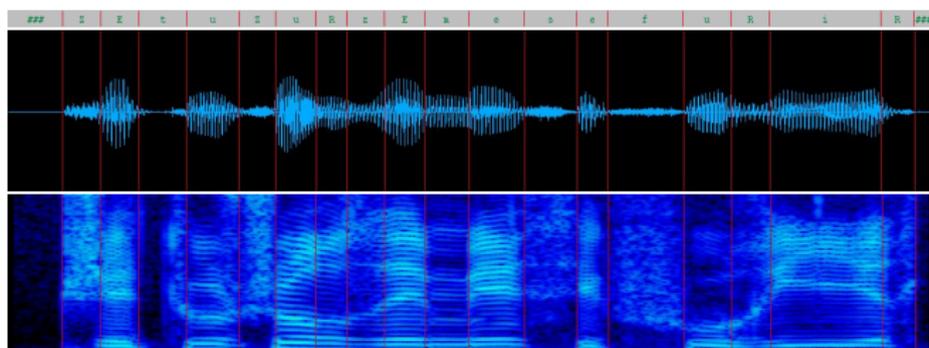
- Segmentation d'un signal de parole en phones :
 - **Etiquetage** : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

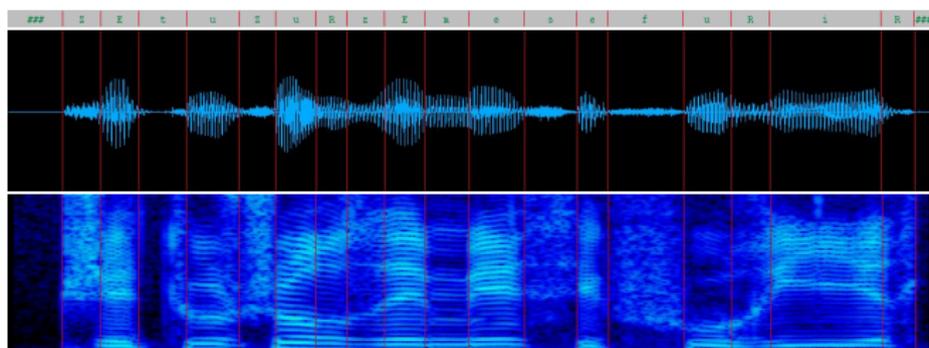
- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - **Marquage** : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

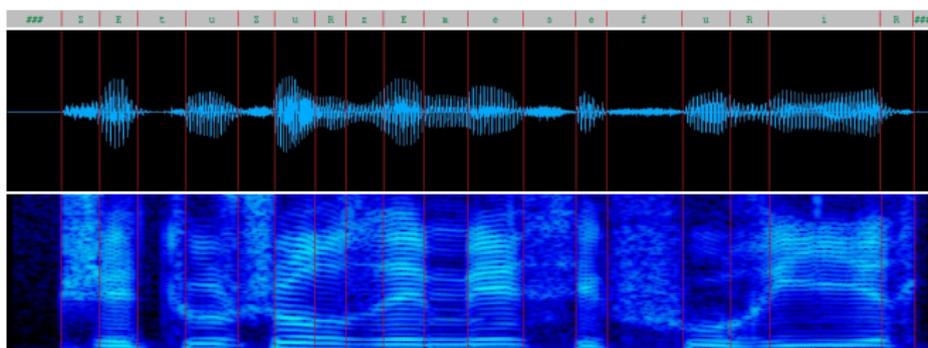
- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

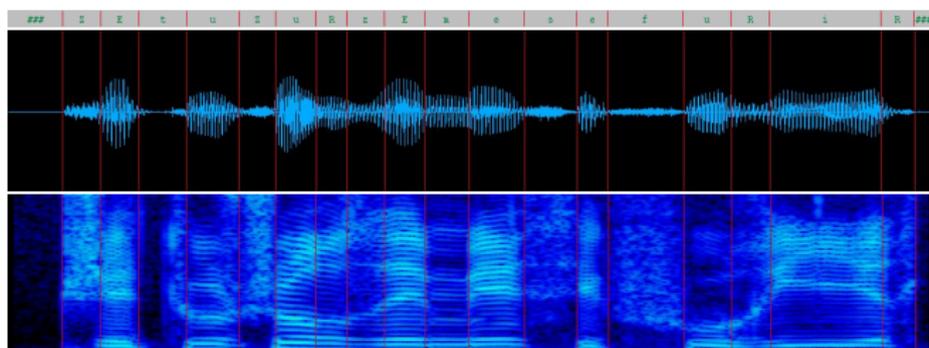
- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

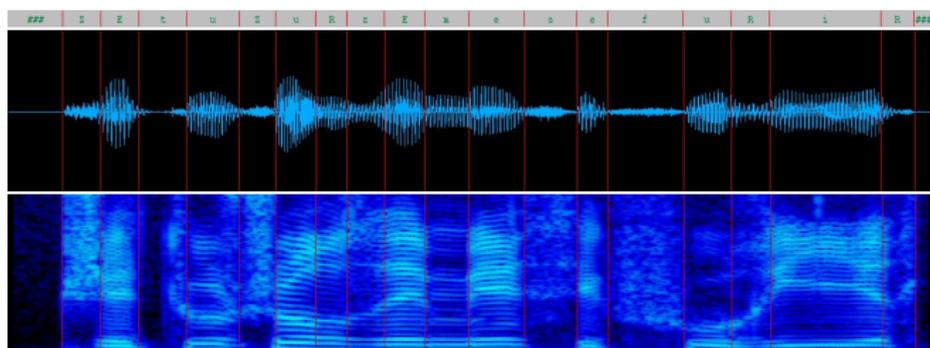
- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

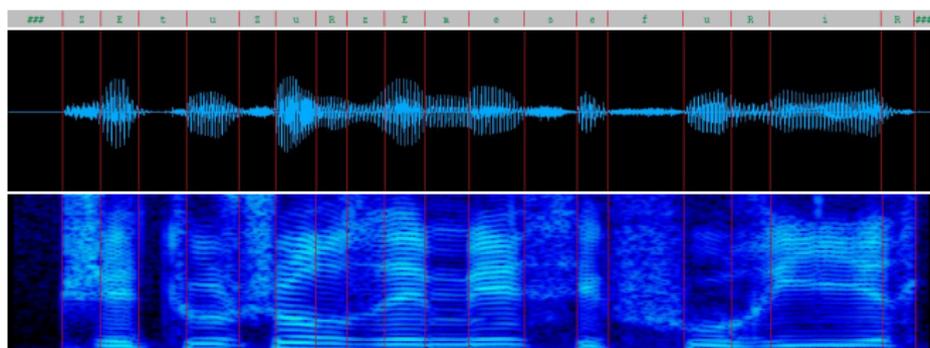
- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

Introduction

- Segmentation d'un signal de parole en phones :
 - Etiquetage : reconnaître la suite des phones réalisés
 - Marquage : trouver les marques de début et de fin des phones



- Applications :
 - synthèse par concaténation d'unités
 - analyse de corpus
 - traitements (transformations ...)
 - suivi temps-réel

- 1 HMMs
- 2 Apprentissage
- 3 ircamAlign
- 4 Démonstrations

- 1 HMMs
- 2 Apprentissage
- 3 ircamAlign
- 4 Démonstrations

Modélisation probabiliste

Modélisation probabiliste

- $Y = \{Y_n\}_{n=1}^N$: processus observé (coefficients MFCC)
- $X = \{X_n\}_{n=1}^N$: processus caché (étiquettes des phonèmes)
- Liens statistiques modélisés par la loi de (X, Y) donnée par $p(x, y|\phi)$

Objectif : **Estimer une réalisation \hat{x} à partir de y .**

Estimateur de Maximum *a posteriori* (MAP)

$$\hat{x}_{MAP}(y) = \arg \max_{x' \in \mathcal{X}^N} p(x'|y; \phi) \quad (1)$$

Modélisation probabiliste

Modélisation probabiliste

- $Y = \{Y_n\}_{n=1}^N$: processus observé (coefficients MFCC)
- $X = \{X_n\}_{n=1}^N$: processus caché (étiquettes des phonèmes)
- Liens statistiques modélisés par la loi de (X, Y) donnée par $p(x, y|\phi)$

Objectif : **Estimer une réalisation \hat{x} à partir de y .**

Estimateur de Maximum *a posteriori* (MAP)

$$\hat{x}_{MAP}(y) = \arg \max_{x' \in \mathcal{X}^N} p(x'|y; \phi) \quad (1)$$

Chaîne de Markov cachée

HMM

- Processus caché : chaîne de Markov (CM)
- Processus observé : observations indépendantes conditionnellement à X
- Quantités nécessaires à l'estimation calculables récursivement (passes forward/backward)
- Estimateur du MAP : calculé par l'algorithme de Viterbi
- Paramètres ϕ estimés par Baum-Welch (EM)
- $p(x, y|\phi) = p(x)p(y|x; \phi)$
 - $p(x)$: modèle de langage
 - $p(y|x; \phi)$: modèle acoustique

Chaîne de Markov cachée

HMM

- Processus caché : chaîne de Markov (CM)
- Processus observé : observations indépendantes conditionnellement à X
- Quantités nécessaires à l'estimation calculables récursivement (passes forward/backward)
- Estimateur du MAP : calculé par l'algorithme de Viterbi
- Paramètres ϕ estimés par Baum-Welch (EM)
- $p(x, y|\phi) = p(x)p(y|x; \phi)$
 - $p(x)$: modèle de langage
 - $p(y|x; \phi)$: modèle acoustique

Chaîne de Markov cachée

HMM

- Processus caché : chaîne de Markov (CM)
- Processus observé : observations indépendantes conditionnellement à X
- Quantités nécessaires à l'estimation calculables récursivement (passes forward/backward)
- Estimateur du MAP : calculé par l'algorithme de Viterbi
- Paramètres ϕ estimés par Baum-Welch (EM)
- $p(x, y|\phi) = p(x)p(y|x; \phi)$
 - $p(x)$: modèle de langage
 - $p(y|x; \phi)$: modèle acoustique

HMM: modèle de langage

- Connaissance *a priori* modélisée par une CM
- Texte non disponible : modèle de langage
- Texte disponible : CM gauche-droite

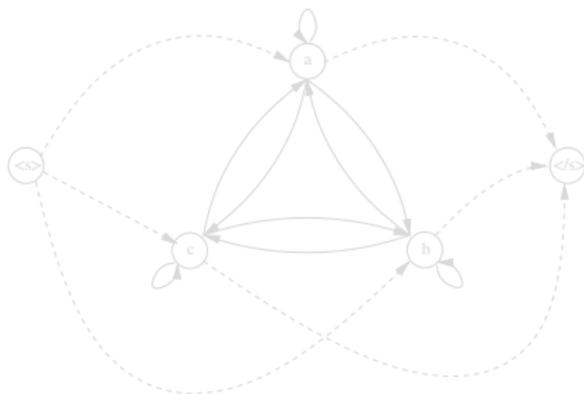


Figure: Reconnaissance (modèle de langage)

HMM: modèle de langage

- Connaissance *a priori* modélisée par une CM
- **Texte non disponible** : modèle de langage
- Texte disponible : CM gauche-droite

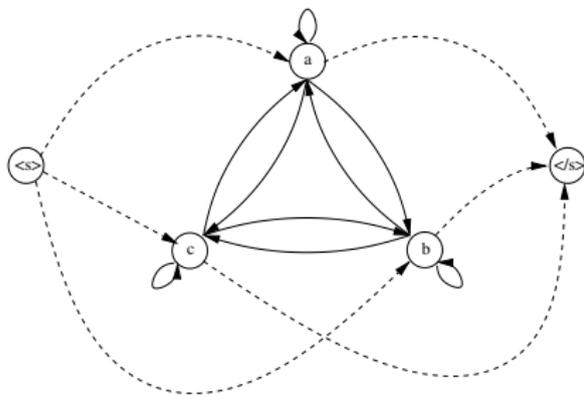


Figure: **Reconnaissance** (modèle de langage)

HMM: modèle de langage

- Connaissance *a priori* modélisée par une CM
- Texte non disponible : modèle de langage
- **Texte disponible** : CM gauche-droite

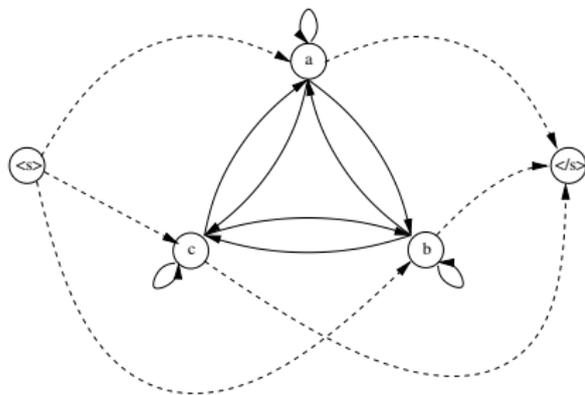


Figure: Reconnaissance (modèle de langage)

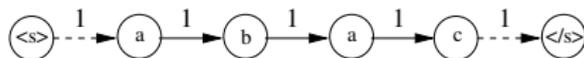
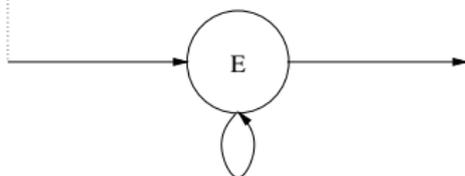
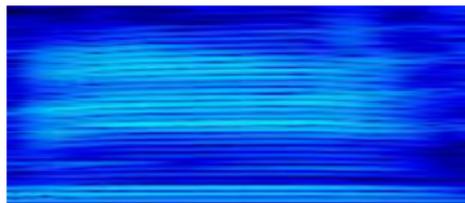


Figure: **Alignement** (la suite phonétique est connue)

HMM: modèle acoustique

Modèle acoustique : modéliser différentes variabilités

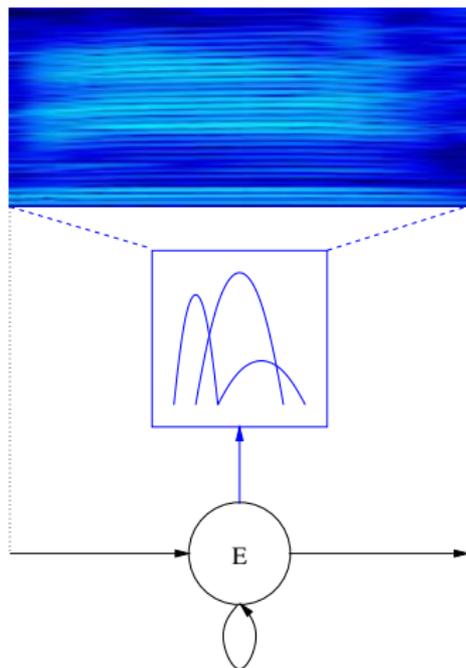
- Variabilité intra et inter-locuteur :
 - mélange de gaussiennes
- Evolution des caractéristiques spectrales au cours du temps :
 - sous-états: modèle de Bakis
- Variabilité en fonction du voisinage
 - Triphones
 - $t-E+n$, $t-E+N$, ...
 - $38^3 = 54872$ modèles HMMs à estimer !!!



HMM: modèle acoustique

Modèle acoustique : modéliser différentes variabilités

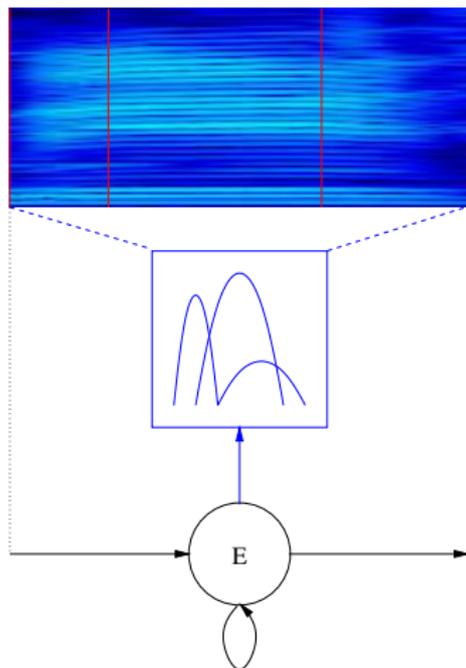
- Variabilité intra et inter-locuteur :
 - mélange de gaussiennes
- Evolution des caractéristiques spectrales au cours du temps:
 - sous-états: modèle de Bakis
- Variabilité en fonction du voisinage
 - Triphones
 - t-E+n, t-E+N, ...
 - $38^3 = 54872$ modèles HMMs à estimer !!!



HMM: modèle acoustique

Modèle acoustique : modéliser différentes variabilités

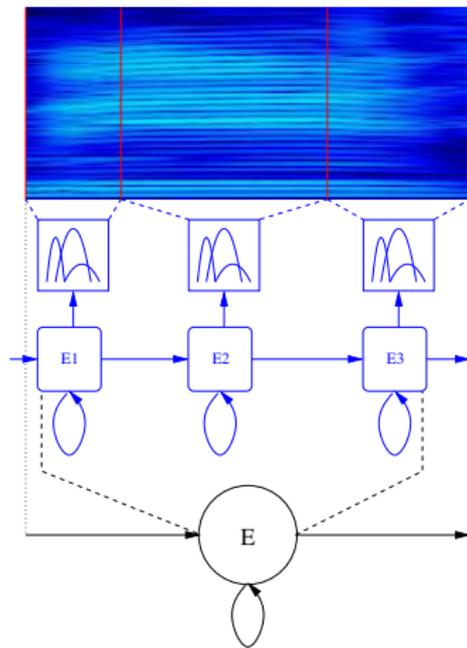
- Variabilité intra et inter-locuteur :
 - mélange de gaussiennes
- Evolution des caractéristiques spectrales au cours du temps :
 - sous-états: modèle de Bakis
- Variabilité en fonction du voisinage
 - Triphones
 - $t-E+n$, $t-E+N$, ...
 - $38^3 = 54872$ modèles HMMs à estimer !!!



HMM: modèle acoustique

Modèle acoustique : modéliser différentes variabilités

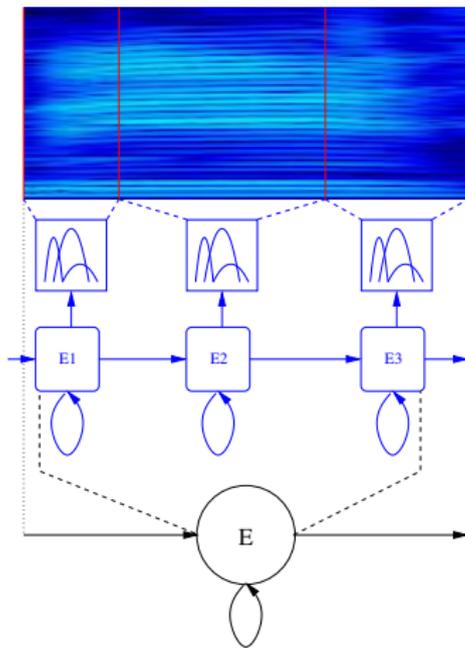
- Variabilité intra et inter-locuteur :
 - mélange de gaussiennes
- Evolution des caractéristiques spectrales au cours du temps :
 - sous-états: modèle de Bakis
- Variabilité en fonction du voisinage
 - Triphones
 - $t-E+n$, $t-E+N$, ...
 - $38^3 = 54872$ modèles HMMs à estimer !!!



HMM: modèle acoustique

Modèle acoustique : modéliser différentes variabilités

- Variabilité intra et inter-locuteur :
 - mélange de gaussiennes
- Evolution des caractéristiques spectrales au cours du temps :
 - sous-états: modèle de Bakis
- Variabilité en fonction du voisinage
 - Triphones
 - $t-E+n$, $t-E+N$, ...
 - $38^3 = 54872$ modèles HMMs à estimer !!!



Triphones

- Triphones = grand nombre de paramètres

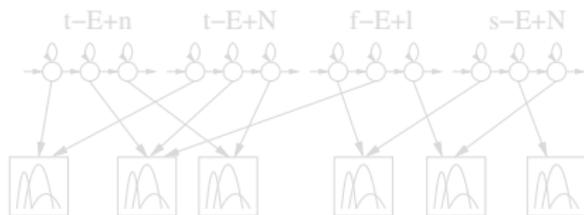


Figure: Tying

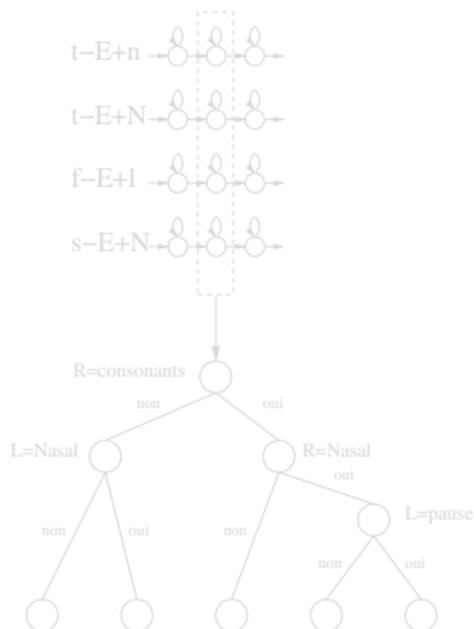


Figure: Arbre de décision

Triphones

- Triphones = grand nombre de paramètres

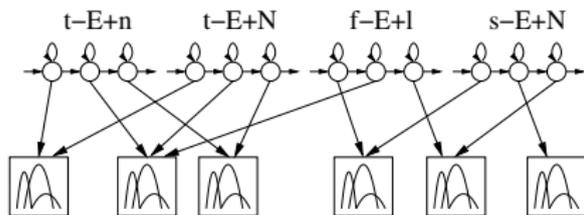


Figure: Tying

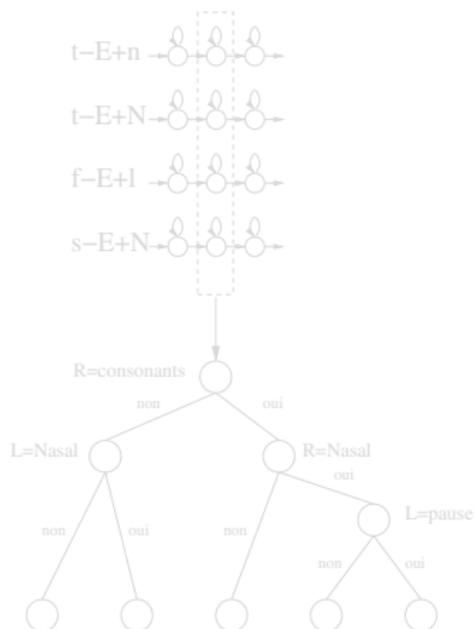


Figure: Arbre de décision

Triphones

- Triphones = grand nombre de paramètres

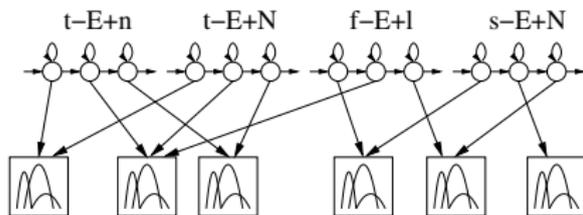


Figure: Tying

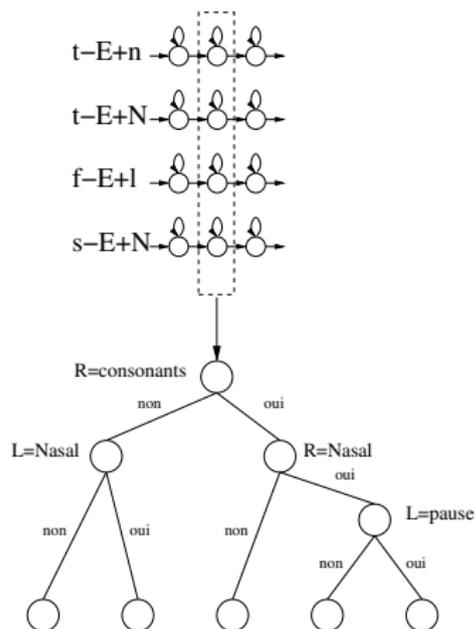


Figure: Arbre de décision

- 1 HMMs
- 2 Apprentissage**
- 3 ircamAlign
- 4 Démonstrations

Description des différents corpus

Lien wiki : <http://wiki.rd/index.php?title=Corpus>

Conversion de voix (F. Villavicencio)

- Différents locuteurs (12 hommes, 1 femme, 1 enfant)
- 200 phrases phonétiquement équilibrées

Transformation de voix (S. Farner)

- Différents types (enfant, femme, vieillard...) et natures (rauque, soufflée, douce, chuchotée, ...) de voix

Voix expressives (G. Beller)

- Différentes expressivités (peur, joie, tristesse, colère...)

Description des différents corpus

Lien wiki : <http://wiki.rd/index.php?title=Corpus>

Conversion de voix (F. Villavicencio)

- Différents locuteurs (12 hommes, 1 femme, 1 enfant)
- 200 phrases phonétiquement équilibrées

Transformation de voix (S. Farner)

- Différents types (enfant, femme, vieillard...) et natures (rauque, soufflée, douce, chuchotée, ...) de voix

Voix expressives (G. Beller)

- Différentes expressivités (peur, joie, tristesse, colère...)

Description des différents corpus

Lien wiki : <http://wiki.rd/index.php?title=Corpus>

Conversion de voix (F. Villavicencio)

- Différents locuteurs (12 hommes, 1 femme, 1 enfant)
- 200 phrases phonétiquement équilibrées

Transformation de voix (S. Farner)

- Différents types (enfant, femme, vieillard...) et natures (rauque, soufflée, douce, chuchotée, ...) de voix

Voix expressives (G. Beller)

- Différentes expressivités (peur, joie, tristesse, colère...)

Description des différents corpus

Lien wiki : <http://wiki.rd/index.php?title=Corpus>

Conversion de voix (F. Villavicencio)

- Différents locuteurs (12 hommes, 1 femme, 1 enfant)
- 200 phrases phonétiquement équilibrées

Transformation de voix (S. Farner)

- Différents types (enfant, femme, vieillard...) et natures (rauque, soufflée, douce, chuchotée, ...) de voix

Voix expressives (G. Beller)

- Différentes expressivités (peur, joie, tristesse, colère...)

Description des différents corpus

Cocteau

- Discours de réception à l'Académie française
- Réverbéré
- *Application* : Test

Monolocuteur : X.Rodet

- 6 heures
- Chambre anéchoïque, 16 bits, 44.1kHz
- *Application* : Synthèse de parole

Multilocuteur : BREF80

- 80 locuteurs: 44 femmes, 36 hommes.
- *Application* : Segmentation

Description des différents corpus

Cocteau

- Discours de réception à l'Académie française
- Réverbéré
- *Application* : Test

Monolocuteur : X.Rodet

- 6 heures
- Chambre anéchoïque, 16 bits, 44.1kHz
- *Application* : Synthèse de parole

Multilocuteur : BREF80

- 80 locuteurs: 44 femmes, 36 hommes.
- *Application* : Segmentation

Description des différents corpus

Cocteau

- Discours de réception à l'Académie française
- Réverbéré
- *Application* : Test

Monolocuteur : X.Rodet

- 6 heures
- Chambre anéchoïque, 16 bits, 44.1kHz
- *Application* : Synthèse de parole

Multilocuteur : BREF80

- 80 locuteurs: 44 femmes, 36 hommes.
- *Application* : Segmentation

Apprentissage : Calcul réparti

- Modèles appris : **monolocuteur** (Xavier) et **multilocuteur** (BREF80)
- Méthode d'apprentissage : embedded Training
- Répartition sur les machines de l'équipe
 - Calculs des quantités nécessaires à l'estimation décomposables

Apprentissage : Calcul réparti

- Modèles appris : **monolocuteur** (Xavier) et **multilocuteur** (BREF80)
- Méthode d'apprentissage : embedded Training
- Répartition sur les machines de l'équipe
 - Calculs des quantités nécessaires à l'estimation décomposables

Apprentissage : Calcul réparti

- Modèles appris : **monolocuteur** (Xavier) et **multilocuteur** (BREF80)
- Méthode d'apprentissage : embedded Training
- Répartition sur les machines de l'équipe
 - Calculs des quantités nécessaires à l'estimation décomposables

Optimisation des modèles monocuteurs

Match Accuracy

$$M_{Acc} = 100 \times \frac{H}{H + S + D + I} \quad (2)$$

H : Hits, S : Substitutions, I : Insertions, D : Délétions

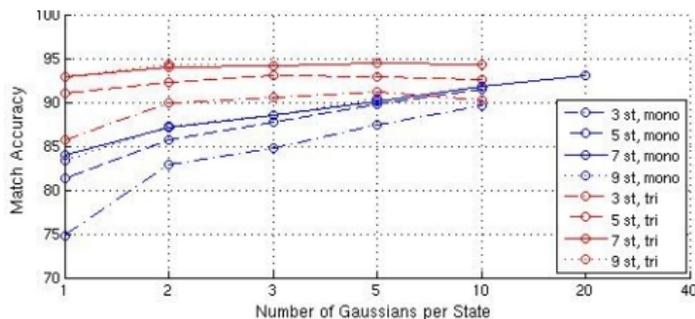


Figure: Match accuracy en fonction du nombre de gaussiennes par état pour différents nombres d'états par HMM (6 itérations à chaque ajout de gaussiennes au mélange, mono et triphones)

Topologie retenue

Topologie des HMMs la plus adaptée au corpus considéré

- Modèles de triphones
- 7 états par phonème
- 5 gaussiennes par état

$M_{Acc} = 93.20\%$ en reconnaissance (sans le texte)

	s	a	n	E	e	b	R	9	y	k	u	s	t	o	i	z	l	z	f	a	d	w	o	v	p	e	j	g	H	Z	S	Del	[%c / %e]					
sil	308	0				
i																																		0				
l																																			0			
p																																				0		
sp		40	3	[97.6/0.0]		
a			170	0		
m				81	0		
E					93	12	2	[88.6/0.4]		
e						8	115	1	[93.5/0.3]		
b							68	0	[98.6/0.0]		
R								192	2			
9									67	2			
n										79	0			
y											72	1	[92.3/0.2]		
k												79	0			
u													76	0			
s														113	1			
t															146	0	[98.0/0.1]		
o																41	1	[87.2/0.2]		
i																	141	3	[97.2/0.1]		
z																		50	2	[96.2/0.1]		
l																			2	3	[99.5/0.0]		
z																																		0	[80.6/0.4]			
f																																			0	[95.2/0.1]		
a																																			0			
d																																				0	[98.8/0.0]	
w																																				0	[88.5/0.1]	
o																																				0	[88.5/0.2]	
v																																				1		
p																																				0	[99.0/0.0]	
e																																				0	[95.1/0.2]	
9																																				0	[81.8/0.1]	
e																																				0	[88.0/0.2]	
j																																				2	[90.0/0.2]	
g																																				1	[97.3/0.0]	
o																																				2	[95.8/0.1]	
H																																				1		
Z																																					1	
S																																				25	0	
Ins	0	7	9	1	3	3	0	1	1	0	0	1	1	2	3	1	2	0	4	1	0	2	1	0	1	0	9	29	0	0	2	0	0	0	1			

Figure: Matrice de confusion, $H = 3000$, $D = 31$, $S = 103$, $I = 85$.

- 1 HMMs
- 2 Apprentissage
- 3 ircamAlign**
- 4 Démonstrations

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- Monolocuteur : appris sur le corpus *X.Rodet*
- Multilocuteur : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les variantes de prononciations et permettant les sauts et les répétitions de mots.

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- **Monolocuteur** : appris sur le corpus *X.Rodet*
- **Multilocuteur** : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les variantes de prononciations et permettant les sauts et les répétitions de mots.

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- **Monolocuteur** : appris sur le corpus *X.Rodet*
- **Multilocuteur** : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les variantes de prononciations et permettant les sauts et les répétitions de mots.

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- **Monolocuteur** : appris sur le corpus *X.Rodet*
- **Multilocuteur** : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les variantes de prononciations et permettant les sauts et les répétitions de mots.

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- **Monolocuteur** : appris sur le corpus *X.Rodet*
- **Multilocuteur** : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les **variantes de prononciations** et permettant les **sauts** et les **répétitions de mots**.

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- **Monolocuteur** : appris sur le corpus *X.Rodet*
- **Multilocuteur** : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les variantes de prononciations et permettant les sauts et les répétitions de mots.

Caractéristiques d'ircamAlign

Système de segmentation de signaux de parole au niveau *phonème* fondé sur la modélisation HMM.

Modèles de phonèmes disponibles (HMMs)

- **Monolocuteur** : appris sur le corpus *X.Rodet*
- **Multilocuteur** : appris sur le corpus *BREF80*

Décodage

- Sans texte : *modèle de langage* au niveau phonème.
- Avec texte : *graphe phonétique* prenant en compte les **variantes de prononciations** et permettant les **sauts** et les **répétitions de mots**.

Fonctionnement

Ligne de commande

- `ircamAlign -i foo.wav -m -w`
- `-m` : modèle multilocuteur ou `-x64` : modèle monolocuteur
64 gaussiennes
- `-w` : affichage sous wavesurfer
- Manuel disponible : `man ircamAlign`

Fonctionnement

Ligne de commande

- `ircamAlign -i foo.wav -m -w`
- `-m` : modèle multilocuteur ou `-x64` : modèle monolocuteur
64 gaussiennes
- `-w` : affichage sous wavesurfer
- Manuel disponible : `man ircamAlign`

Fonctionnement

Ligne de commande

- `ircamAlign -i foo.wav -m -w`
 - `-m` : modèle multilocuteur ou `-x64` : modèle monolocuteur
64 gaussiennes
 - `-w` : affichage sous wavesurfer
- Manuel disponible : *man ircamAlign*

Fonctionnement

Ligne de commande

- `ircamAlign -i foo.wav -m -w`
- `-m` : modèle multilocuteur ou `-x64` : modèle monolocuteur
64 gaussiennes
- `-w` : affichage sous wavesurfer

- Manuel disponible : *man ircamAlign*

- 1 HMMs
- 2 Apprentissage
- 3 ircamAlign
- 4 Démonstrations**

Monocuteur

- Locuteur : X. Rodet
 - Modèle Utilisé : Monocuteur 64 gaussiennes (option -x64)
 - Caractéristiques: délétion et répétition de mots

Texte donné en entrée d'ircamAlign

"J'ai toujours aimé ces fous rires qui montrent l'âme grande ouverte"

Texte prononcé

- "J'ai toujours aimé ces fou-rires qui montrent l'âme grande ouverte"
- "J'ai toujours aimé (...) rires (...) l'âme grande ouverte"
- "J'ai toujours *toujours* aimé (...) rires l'âme grande ouverte"

Monocuteur

- Locuteur : X. Rodet
 - Modèle Utilisé : Monocuteur 64 gaussiennes (option -x64)
 - Caractéristiques: délétion et répétition de mots

Texte donné en entrée d'ircamAlign

"J'ai toujours aimé ces fous rires qui montrent l'âme grande ouverte"

Texte prononcé

- "J'ai toujours aimé ces fou-rires qui montrent l'âme grande ouverte"
- "J'ai toujours aimé (...) rires (...) l'âme grande ouverte"
- "J'ai toujours *toujours* aimé (...) rires l'âme grande ouverte"

Monocuteur

- Locuteur : X. Rodet
 - Modèle Utilisé : Monocuteur 64 gaussiennes (option -x64)
 - Caractéristiques: délétion et répétition de mots

Texte donné en entrée d'ircamAlign

"J'ai toujours aimé ces fous rires qui montrent l'âme grande ouverte"

Texte prononcé

- "J'ai toujours aimé ces fou-rires qui montrent l'âme grande ouverte"
- "J'ai toujours aimé (...) rires (...) l'âme grande ouverte"
- "J'ai toujours *toujours* aimé (...) rires l'âme grande ouverte"

Monocuteur

- Locuteur : X. Rodet
 - Modèle Utilisé : Monocuteur 64 gaussiennes (option -x64)
 - Caractéristiques: délétion et répétition de mots

Texte donné en entrée d'ircamAlign

"J'ai toujours aimé ces fous rires qui montrent l'âme grande ouverte"

Texte prononcé

- "J'ai toujours aimé ces fou-rires qui montrent l'âme grande ouverte"
- "J'ai toujours aimé (...) rires (...) l'âme grande ouverte"
- "J'ai toujours *toujours* aimé (...) rires l'âme grande ouverte"

Monocuteur

- Locuteur : X. Rodet
 - Modèle Utilisé : Monocuteur 64 gaussiennes (option -x64)
 - Caractéristiques: délétion et répétition de mots

Texte donné en entrée d'ircamAlign

"J'ai toujours aimé ces fous rires qui montrent l'âme grande ouverte"

Texte prononcé

- "J'ai toujours aimé ces fou-rires qui montrent l'âme grande ouverte"
- "J'ai toujours aimé (...) rires (...) l'âme grande ouverte"
- "J'ai toujours *toujours* aimé (...) rires l'âme grande ouverte"

Multilocuteur: Accent hispanique

- Locuteur : F. Villavicencio
 - Modèle Utilisé : Multilocuteur (option -m)
 - Caractéristique : léger accent hispanique

Texte donné en entrée d'ircamAlign et prononcé

"Ecoutez, Ecoutez, Ecoutez ce chant binaire."

Multilocuteur: Accent hispanique

- Locuteur : F. Villavicencio
 - Modèle Utilisé : Multilocuteur (option -m)
 - Caractéristique : léger accent hispanique

Texte donné en entrée d'ircamAlign et prononcé

"Ecoutez, Ecoutez, Ecoutez ce chant binaire."

Enregistrement ancien

- Locuteur : J.Cocteau
 - Modèle Utilisé : Multilocuteur (option -m)
 - Caractéristique : enregistrement ancien, paragraphe.

Texte donné en entrée d'ircamAlign et prononcé

"Messieurs, Rémy de Gourmont disait que chez Edmond Rostand la chance est une des formes du génie. Rostand fut porté sur ce siège par des fées rapides et dans un tumulte d'ailes qu'il évoque autour de la naissance d'Henri de Bornier. Toutes les portes qui se ferment devant les guerriers noirs des Lettres dont Kleist reste l'exemple, s'ouvriraient toutes seules devant ses armes blanches et son blanc panache"

Enregistrement ancien

- Locuteur : J.Cocteau
 - Modèle Utilisé : Multilocuteur (option -m)
 - Caractéristique : enregistrement ancien, paragraphe.

Texte donné en entrée d'ircamAlign et prononcé

"Messieurs, Rémy de Gourmont disait que chez Edmond Rostand la chance est une des formes du génie. Rostand fut porté sur ce siège par des fées rapides et dans un tumulte d'ailes qu'il évoque autour de la naissance d'Henri de Bornier. Toutes les portes qui se ferment devant les guerriers noirs des Lettres dont Kleist reste l'exemple, s'ouvriraient toutes seules devant ses armes blanches et son blanc panache"

Parole expressive

- Locuteur : L. De Funès
 - Modèle Utilisé : Multilocuteur (option -m)
 - Caractéristique : voix expressive

Texte donné en entrée d'ircamAlign et prononcé

"Je vous conseille de ne pas faire le mariolle, hein! Souvenez vous qui vous êtes et où vous êtes!"

Parole expressive

- Locuteur : L. De Funès
 - Modèle Utilisé : Multilocuteur (option -m)
 - Caractéristique : voix expressive

Texte donné en entrée d'ircamAlign et prononcé

"Je vous conseille de ne pas faire le mariolle, hein! Souvenez vous qui vous êtes et où vous êtes!"

Voix chantée

- Locuteur : Cristina Branco
 - Modèle utilisé : Multilocuteur (option -m)
 - Caractéristique : Voix chantée en portugais

Texte original

Sonhei que estava um dia em Portugal
 À toa, num Carnaval em Lisboa, além da poesia
 Meu sonho voa, além da poesia
 E encontra o poeta em Pessoa

Phonétisation manuelle utilisée (*option -k*)

soJEi k istavo dija ## a~m pORtygao ##
 atOa ## num kaRnava e nisbOa ###)
 ### meo swaju vOa anee n da pOezjija ##
 V~ EnkOmtRa u pOEta e tesOa ###

- adaptation sur quelques phrases.

Voix chantée

- Locuteur : Cristina Branco
 - Modèle utilisé : Multilocuteur (option -m)
 - Caractéristique : Voix chantée en portugais

Texte original

Sonhei que estava um dia em Portugal
 À toa, num Carnaval em Lisboa, além da poesia
 Meu sonho voa, além da poesia
 E encontra o poeta em Pessoa

Phonétisation manuelle utilisée (*option -k*)

soJEi k istavo dija ## a~m pORtygao ##
 atOa ## num kaRnava e nisbOa ###)
 ### meo swaju vOa anee n da pOezjija ##
 V~ EnkOmtRa u pOEta e tesOa ###

- adaptation sur quelques phrases.

Voix chantée

- Locuteur : Cristina Branco
 - Modèle utilisé : Multilocuteur (option -m)
 - Caractéristique : Voix chantée en portugais

Texte original

Sonhei que estava um dia em Portugal
 À toa, num Carnaval em Lisboa, além da poesia
 Meu sonho voa, além da poesia
 E encontra o poeta em Pessoa

Phonétisation manuelle utilisée (*option -k*)

soJEi k istavo dija ## a~m pORtygao ##
 atOa ## num kaRnava e nisbOa ###)
 ### meo swaju vOa anee n da pOezjija ##
 V~ EnkOmtRa u pOEta e tesOa ###

- adaptation sur quelques phrases.

Voix chantée

- Locuteur : Cristina Branco
 - Modèle utilisé : Multilocuteur (option -m)
 - Caractéristique : Voix chantée en portugais

Texte original

Sonhei que estava um dia em Portugal
 À toa, num Carnaval em Lisboa, além da poesia
 Meu sonho voa, além da poesia
 E encontra o poeta em Pessoa

Phonétisation manuelle utilisée (*option -k*)

soJEi k istavo dija ## a~m pORtygao ##
 atOa ## num kaRnava e nisbOa ###)
 ### meo swaju vOa anee n da pOezjija ##
 V~ EnkOmtRa u pOEta e tesOa ###

- adaptation sur quelques phrases.

Perspectives

- Langue anglaise
- Adaptation au locuteur
- Indice de confiance
- Voix chantée (utilisation de la F0)- collaboration avec le suivi de partition

Perspectives

- Langue anglaise
- Adaptation au locuteur
- Indice de confiance
- Voix chantée (utilisation de la F0)- collaboration avec le suivi de partition

Perspectives

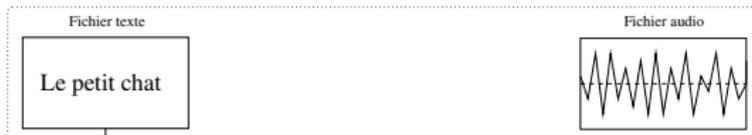
- Langue anglaise
- Adaptation au locuteur
- Indice de confiance
- Voix chantée (utilisation de la F0)- collaboration avec le suivi de partition

Perspectives

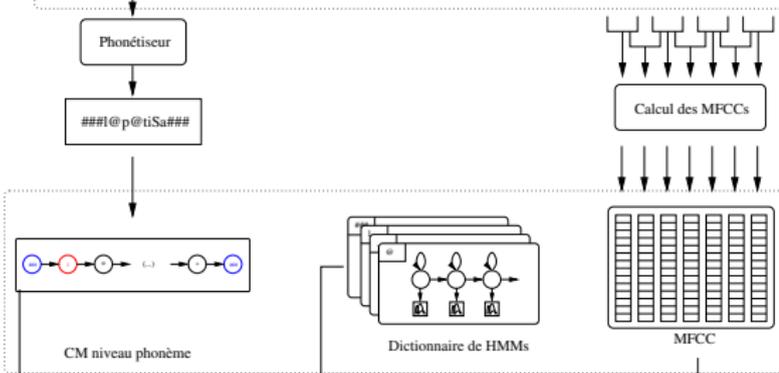
- Langue anglaise
- Adaptation au locuteur
- Indice de confiance
- Voix chantée (utilisation de la F0)- collaboration avec le suivi de partition

Annexe : Alignement

Données



Prétraitements



Alignement

