

TP OP3 - Traitement Automatique de la Parole

Master 1 informatique

jerome.farinas@irit.fr

Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier

5 mars 2009



Présentation des TPs

- 6 séances de TP de 2h
- Evaluation :
 - ▶ mini dossier (moins de 10 pages) synthétisant ce que vous avez appris lors de ces deux parties)

Détail des parties :

- 1 Rappels (prérequis M1 info TC4)
- 2 TP1 : Analyse du signal de parole (sous windows+matlab)
Manipulation du signal sous matlab, compréhension des mécanismes de base de la paramétrisation audio
- 3 TP2-4 : Modélisation GMM (reconnaissance du locuteur) (sous linux+MISTRAL) : Modélisation acoustique des locuteurs et mise en place d'un outil de vérification.
- 4 TP5-6 : Modélisation HMM (décodage acoustique) (sous linux+wavesurfer+HTK) : Réalisation d'un système de reconnaissance de phonème, élément de base des système de reconnaissance de la parole grand vocabulaire

Première partie I

Rappels

Plan

1 Numérisation du signal de parole

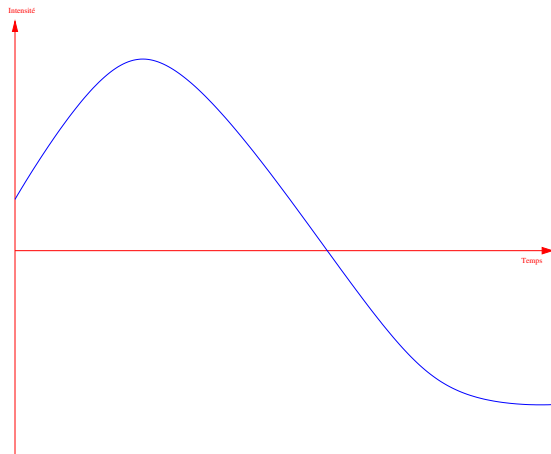
- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Numérisation du signal de parole

- Objectif : reproduire un signal sonore analogique dans un fichier informatique numérique



Plan

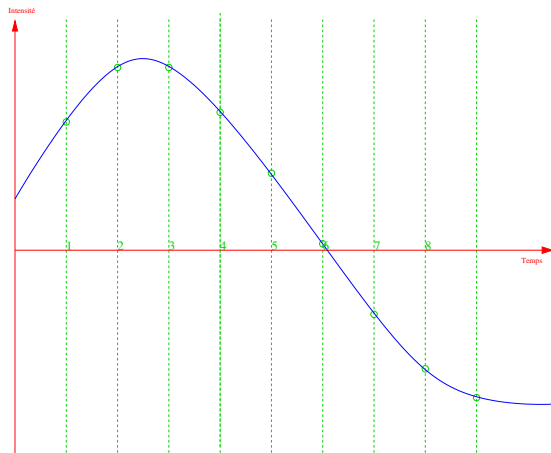
1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Fréquence d'échantillonnage



- Le signal est échantillonné avec un pas de mesure $T = 1/F_{ech}$

Fréquence d'échantillonnage

Théorème de Shannon

La fréquence d'échantillonnage assurant un non repliement du spectre doit être supérieure à 2 fois la fréquence haute du spectre du signal analogique.

$$F_{ech} = 2 * F_{max}$$

- Signal audio pleine bande à un spectre s'étalant de 20 Hz à 20 kHz
- Bandes de fréquences couramment utilisées pour réduire F_{ech} :

	Spéctre du signal	F_{ech}	Applications
Qualité téléphonique	[300-3400 Hz]	8 kHz	Téléphonie
Qualité "bande élargie"	[50-7000 Hz]	16 ou 22 kHz	PC, audio-conférence (ADPCM)
Haute qualité en radiodiffusion	[50 - 15 000 Hz]	32 kHz	DAB, NICAM
Qualité "Hi-Fi"	[20 - 20 000 Hz]	44.1 ou 48 kHz	CD Audio, Studio numérique, DAT

Plan

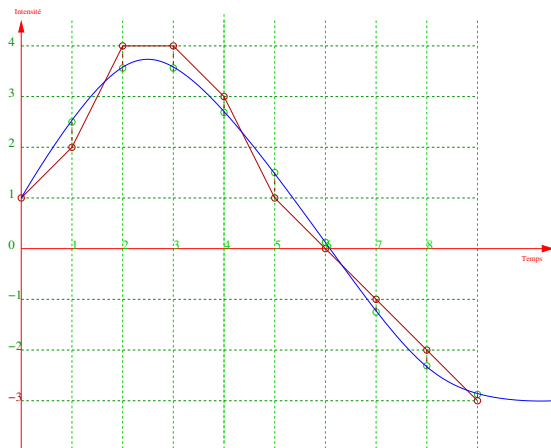
1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- **Quantification**
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Quantification



- Le signal numérisé est représenté en rouge, l'erreur de quantification en pointillé rouge

Quantification

- Nombre n de bits utilisés pour le codage des échantillons
- Erreur de quantification maximale proportionnelle à $\frac{1}{2^n}$
- Dans le cas d'une quantification linéaire (pas de quantification constant sur toute la plage de conversion), on exprime l'erreur due à la quantification sous la forme d'un rapport Signal à Bruit (SNR : Signal to Noise Ratio) dont l'expression est la suivante¹ :

$$SNR_{dB} = 6.02 * n + 1.76dB$$

	Q	SNR_{dB}
Qualité "Hi-Fi"	16-18 bits	95
Codage la parole, NICAM	14 bits	80
Codage son PC	8 bits	40

¹Calcul SNR : http://fr.wikipedia.org/wiki/Rapport_signal_sur_bruit

Plan

1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Fichiers informatiques

- Il existe de nombreux formats de fichiers son
- Parmi les plus courant :
 - ▶ raw : fichier binaire nu
 - ▶ wav : introduit par Microsoft
 - ▶ aiff : utilisé sur Silicon Graphics et Macintosh
 - ▶ au : format SUN
- En plus du fichier brut, ces formats encapsulent des informations complémentaires, comme la valeur de la fréquence d'échantillonnage, la quantification utilisée, le nombre de canaux...

Plan

1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Plan

1 Numérisation du signal de parole

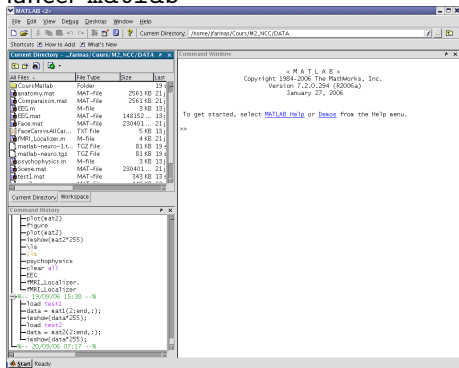
- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Lancer Matlab

- se connecter sur Windows
- lancer matlab



Plan

1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- **Charger un fichier audio**
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Charger un fichier audio

```
cd 'T :\M10P3\' ;  
[signal,fs,q,options] = wavread('Signal/essai.wav') ;
```

- Quelles sont les valeurs de `fs` et `q`? Que contient `options`?
- Affichez les 10 premières valeurs de `signal`.
- Quel est l'intervalle de variation des échantillons du `signal`? Comment ramener le vecteur à des valeurs entières?

Plan

1 Numérisation du signal de parole

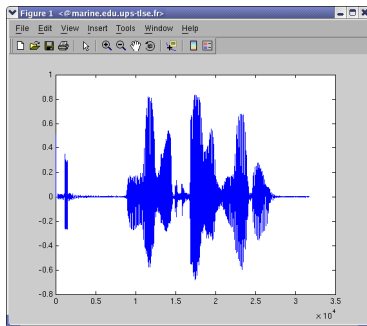
- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- **Afficher un signal audio**
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Afficher un signal audio

```
plot(signal) ;
```



- Calculez la longueur en secondes du signal
- Affichez dans une nouvelle fenêtre le vecteur converti en entier
- Affichez la portion de signal comprise entre les échantillons 11345 et 11375 en bleu
- Figez l'affichage et affichez avec des croix rouge la même portion

Plan

1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- **Spectre du signal**
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Spectre d'un signal

- 1 Charger le fichier 'Signal/note.wav' dans la variable 'note'.
- 2 Affichez le signal de `note` avec l'axe des abscisses en secondes.
- 3 Affichez le spectre de `note` (avec la fonction `fft`) avec l'axe des abscisses en Hertz. Que remarquez-vous ?
- 4 Quelle est la fréquence de base de cette note de musique ?
- 5 Calculez la transformée de Fourier rapide de l'extrait du signal de parole (`signal`) et affichez le résultat sur une demi-période, avec les abscisses en Hertz.
- 6 Comparez le spectre de `note` et celui de `signal`.

Plan

1 Numérisation du signal de parole

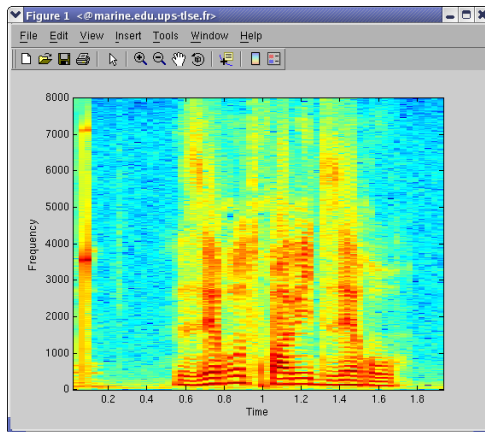
- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- **Spectrogramme du signal**
- Echelle MEL
- Application : spectre du signal
- Application : analyse cepstrale

Spectrogramme du signal

```
figure ;  
specgram(signal,1024,fs,hamming(1024)) ;
```



Plan

1 Numérisation du signal de parole

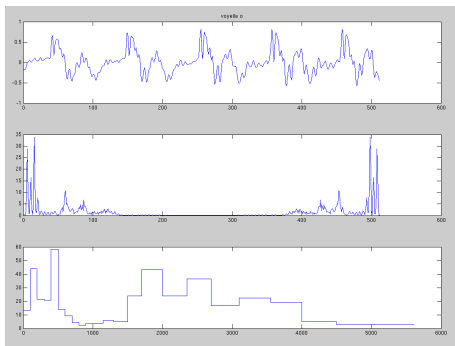
- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- **Echelle MEL**
- Application : spectre du signal
- Application : analyse cepstrale

Echelle MEL

```
en = canaux(signal,11300,512,16,24,'voyelle o');
```



- 1 fenêtre d'analyse (512 points de signal)
- 2 Module du spectre sur 256 points
- 3 Canaux d'énergies dans les 24 bandes réparties selon l'échelle MEL

Plan

1 Numérisation du signal de parole

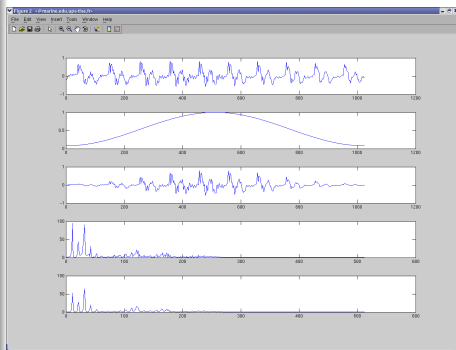
- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- **Application : spectre du signal**
- Application : analyse cepstrale

Calcul du spectre et fenêtrage de Hamming

```
figure ;  
extrait = signal(11300 :12323) ;  
subplot(511) ;  
plot(extrait) ;  
subplot(512) ;  
plot(hamming(1024)) ;  
subplot(513) ;  
plot(hamming(1024).*extrait) ;  
subplot(514) ;  
f=abs(fft(extrait)) ;  
plot(f(1 :512)) ;  
subplot(515) ;  
h=abs(fft(extrait.*hamming(1024))) ;  
plot(h(1 :512)) ;
```



- Expliquez ces différents tracés.
- Quel est l'intérêt de l'utilisation d'une fenêtre de Hamming ?

Plan

1 Numérisation du signal de parole

- Fréquence d'échantillonnage
- Quantification
- Fichiers informatiques

2 Analyse du signal de parole

- Lancer Matlab
- Charger un fichier audio
- Afficher un signal audio
- Spectre du signal
- Spectrogramme du signal
- Echelle MEL
- Application : spectre du signal
- **Application : analyse cepstrale**

Calcul du cepstre

- ➊ Réalisez le calcul du cepstre sur l'extrait de signal de parole.
- ➋ Affichez l'extrait, son spectre et son cepstre dans la même fenêtre.
- ➌ Affichez le spectre lissé du signal.
- ➍ Pourquoi la valeur du premier coefficient cepstral est-elle aussi élevée ?

Deuxième partie II

Modélisation GMM (locuteur)

Plan

- 3 Enregistrement fichiers audio
- 4 Analyse triangle vocalique des locuteurs
- 5 Modélisation GMM des locuteurs
 - Théorie
 - Mise en pratique

Enregistrement fichiers audio

- En utilisant wavesurfer (sous linux) vous allez enregistrer 3 versions différentes d'un texte équilibré phonétiquement (« la bise et le soleil ») :
 - ▶ une version sera lue normalement (nom du fichier : nom-prenom_2009_normal.wav),
 - ▶ une version lue rapidement (nom-prenom_2009_rapide.wav),
 - ▶ une version lente et hyper-articulée (nom-prenom_2009_lent.wav).

La bise et le soleil

La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort, quand ils ont vu un voyageur qui s'avancait, enveloppé dans son manteau. Ils sont tombés d'accord que celui qui arriverait le premier à faire ôter son manteau au voyageur serait regardé comme le plus fort. Alors, la bise s'est mise à souffler de toute sa force mais plus elle soufflait, plus le voyageur serrait son manteau autour de lui et à la fin, la bise a renoncé à le lui faire ôter. Alors le soleil a commencé à briller et au bout d'un moment, le voyageur, réchauffé a ôté son manteau. Ainsi, la bise a du reconnaître que le soleil était le plus fort des deux.

Plan

- 3 Enregistrement fichiers audio
- 4 Analyse triangle vocalique des locuteurs
- 5 Modélisation GMM des locuteurs
 - Théorie
 - Mise en pratique

Analyse du triangle vocalique

- Chaque locuteur doit extraire de l'enregistrement lent les valeurs des formants F1 et F2 afin de représenter leur propre triangle vocalique
- Les phonèmes à isoler sont : /a/ /i/ et /u/ (« ou »)
- Afin d'isoler ces valeurs, le plus facile est d'activer le suivi de formants sous wavesurfer. Vous pouvez également utiliser l'affichage d'une fenêtre LPC pour isoler les formants
- Comparez les triangles obtenus par les différents locuteurs : que concluez-vous ?
- Réfaites cette manipulation en utilisant la version lue à vitesse normale (sur la même partie du texte). Est-ce que le triangle est modifié ?

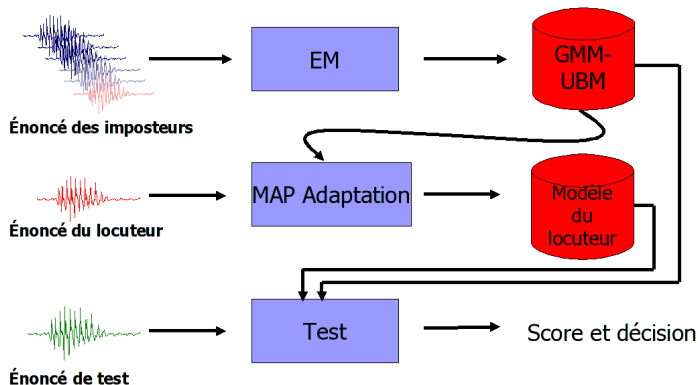
Plan

- 3 Enregistrement fichiers audio
- 4 Analyse triangle vocalique des locuteurs
- 5 Modélisation GMM des locuteurs**
 - Théorie
 - Mise en pratique

Plan

- 3 Enregistrement fichiers audio
- 4 Analyse triangle vocalique des locuteurs
- 5 Modélisation GMM des locuteurs**
 - Théorie
 - Mise en pratique

Principe général de la modélisation de locuteurs



Source : exposé Najim Dehac

Modélisation du monde

- l'ensemble des locuteurs qui vont servir a l'apprentissage vont former un modèle du monde (appelé GMM-UBM : Universal Background Model, cf. [Reynolds QuatieriDunn 2000])
- utilité du modèle GMM-UBM :
 - ▶ permettra l'adaptation à un locuteur
 - ▶ permettra de normaliser les scores afin de permettre des comparaisons entre systèmes
- Il s'agit d'un modèle de mélange de lois Gaussiennes (GMM) appris par l'algorithme EM (Expectation-Maximisation). En général 5 itérations EM sont nécessaires.

Modélisation du locuteur

- le modèle du monde peut-être adapté au locuteur par :
 - ▶ adaptation MAP (Maximum A Posteriori) : adaptation fondée sur le maximum de vraisemblance Bayésien. Cela revient à modifier les variances des modèles.
 - ▶ adaptation MLLR (Maximum Likelihood Linear Regression) : utilisation des données de test pour modifier le modèle. Cela revient à modifier les moyennes des modèles.

Reconnaissance du locuteur

- Calcul du score :
 - ▶ rapport de log vraisemblance (LLR) :

$$S_x(Y) = \frac{1}{N} \sum_{n=1}^N \log \frac{p(y_m/X)}{p(y_m/\Omega)} \leq \theta$$

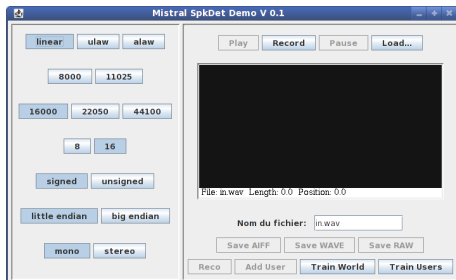
- ★ avec X modèle à évaluer et Ω modèle du monde
- ★ Y représente l'observation qui se décompose en N paramètres que l'on suppose indépendants.
- ★ le seuil θ est défini lors de l'apprentissage (sur l'ensemble de développement)

Plan

- 3 Enregistrement fichiers audio
- 4 Analyse triangle vocalique des locuteurs
- 5 Modélisation GMM des locuteurs**
 - Théorie
 - **Mise en pratique**

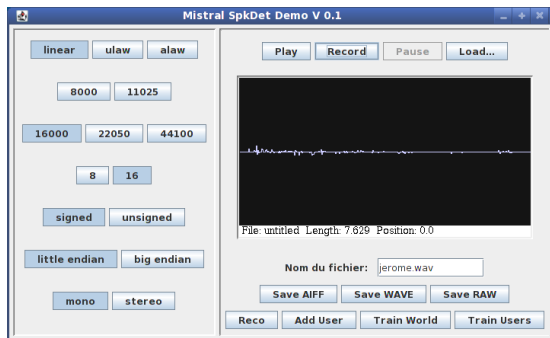
Installation Mistral IHM

- Récupérez l'archive contenant le programme (Mistral démo pour Lia_Spk_Det) et décompressez-la : `tar xzvf tpspk.tar.gz`
- placez-vous dans le repertoire tpgmm
- lancez le programme : `java -jar SpkIHM`



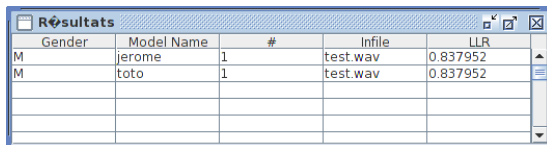
Enregistrement locuteur

- un locuteur correspond a un seul enregistrement audio
- il faut au moins 2 minutes de parole pour obtenir un bon modèle
- appuyez sur “Add User”
- saisissez un nom de locuteur (le fichier actuellement en cours d’édition sera renommé nom_du_locuteur.wav)



Modélisation et reconnaissance

- Après avoir rajouté tous les locuteurs voulus, lancez la modélisation du modèle du monde : “Train World”
- Puis lancez la modélisation des locuteurs : “Train Users”
- Vous pouvez ensuite enregistrer un fichier de test et appuyer sur Reco : le genre est détecté (Male/Female), et la vraisemblance est calculée (LLR)



Gender	Model Name	#	Infile	LLR
M	jerome	1	test.wav	0.837952
M	toto	1	test.wav	0.837952

Troisième partie III

Modélisation HMM (décodage acoustique)

Plan

- 6 Variabilité du signal de parole
- 7 Paramétrisation de la collection audio
- 8 Modélisation par Modèles de Markov Cachés

Variabilité du signal de parole

Sous linux, installer les fichiers et l'arborescence :

```
~farinas/M10P3/install
```

En vous aidant de la transcription du signal de parole, analysez un fichier d'une heure de radio avec wavesurfer (vérifiez que wavesurfer ne charge pas tous les fichiers en mémoire : Préférences -> Sound I/O -> Sound storage = keep on disk).

- Quels sont les facteurs qui contribuent à rendre variable le signal de parole ?
- Est-ce lié :
 - ▶ au locuteur ?
 - ▶ au sexe de la personne qui parle ?
 - ▶ aux bruits environnants ?
- Affichez le signal de parole avec un étiquetage en phonème et tracez le triangle vocalique du chroniqueur « Pierre Le Marc ». Regardez dans le répertoire Listes/ quels sont les fichiers qui correspondent à ce critère.

Plan

- 6 Variabilité du signal de parole
- 7 Paramétrisation de la collection audio**
- 8 Modélisation par Modèles de Markov Cachés

Paramétrisation de la collection audio

- Vous allez travailler sur les fichiers audio découpé en utilisant la boîte à outils HTK.
- Les labels phonétiques se trouvent rassemblés dans le fichier « Labels/LABELSalignes.mlf »
- Utiliser le script « HTKparametrisation » pour réaliser la paramétrisation (les fichiers résultats sont stockés sur votre compte dans le répertoire Parametres/)
- Questions :
 - ▶ Quelle est la paramétrisation appliquée ?
 - ▶ Où pourriez-vous agir pour modifier la paramétrisation ?
 - ▶ Comment est structuré le fichier de labels ? Quelle autre représentation des labels aurait-été possible ?

Plan

- 6 Variabilité du signal de parole
- 7 Paramétrisation de la collection audio
- 8 Modélisation par Modèles de Markov Cachés

Modélisation par Modèles de Markov Cachés

- A quoi va correspondre un modèle de Markov caché ?
- Réalisez l'apprentissage des modèles avec le programme `HTKinitialisation` (utilisez les fichiers du 14 au 16 avril 2003)
- Testez votre apprentissage (fichier `HTKreco.mlf`) avec le fichier du 17 avril 2003.
- Comment pourriez-vous améliorer les résultats ? Quelles modélisations pourriez-vous utiliser ? Quels paramètres ?