

# **Master 2 AIC & SETI**

## **Université Paris-Saclay**

### **Reconnaissance et interaction vocale**

#### Cours 1 – Introduction



**Claude Barras**  
[claude.barras@limsi.fr](mailto:claude.barras@limsi.fr)

Groupe Traitement du Langage Parlé  
LIMSI CNRS / Université Paris-Sud

(certaines illustrations extraites des livres Calliope et de R. Boîte cités en bibliographie)

# Programme des cours

## ■ Cours le jeudi après-midi (14h-17h15)

	Date	Thème
1	16/11/17	Introduction (C. Barras) ; traitement du signal (G. Richard)
2	23/11/17	Modèles de langue (A. Allauzen)
3	30/11/17	Reconnaissance de la parole (C. Barras)
4	07/12/17	Démarrage du projet
5	14/12/17	Emotion et Parole (L. Devillers)
6	11/01/18	Suite du projet
7	18/01/18	Fouille d'opinion (C. Clavel)
8	01/02/18	Rendu du projet
9	15/02/18	Examen

## ■ Evaluation : compte-rendu projet + examen

# Plan de l'introduction

- Problématique
  - ◆ Communication parlée
  - ◆ Point de vue historique
  - ◆ Champs de recherche
- Production de parole
- Principes en reconnaissance de mots isolés
- Transcription automatique de parole continue
  - ◆ Méthodes
  - ◆ Performances
- Perspectives

# Communication orale

## ■ Intérêt de la communication parlée

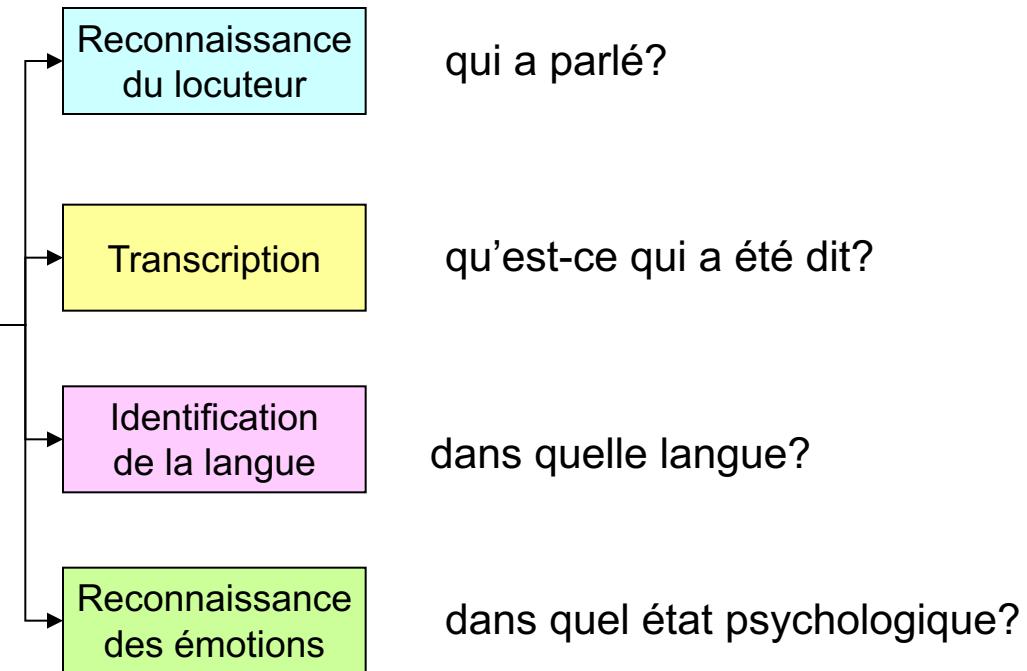
- ◆ moyen de communication entre humains
  - le plus naturel
  - rapide: 150 mots/minute contre 20-50 mots/minute au clavier
- ◆ en communication homme-machine
  - besoins spécifiques: téléphonie, aide aux handicapés
  - modalité supplémentaire au clavier et à la souris

## ■ Applications du traitement automatique de la parole

- ◆ Codage (télécommunications)
- ◆ Synthèse vocale à partir du texte
- ◆ Reconnaissance de la parole

# Que reconnaître dans la parole?

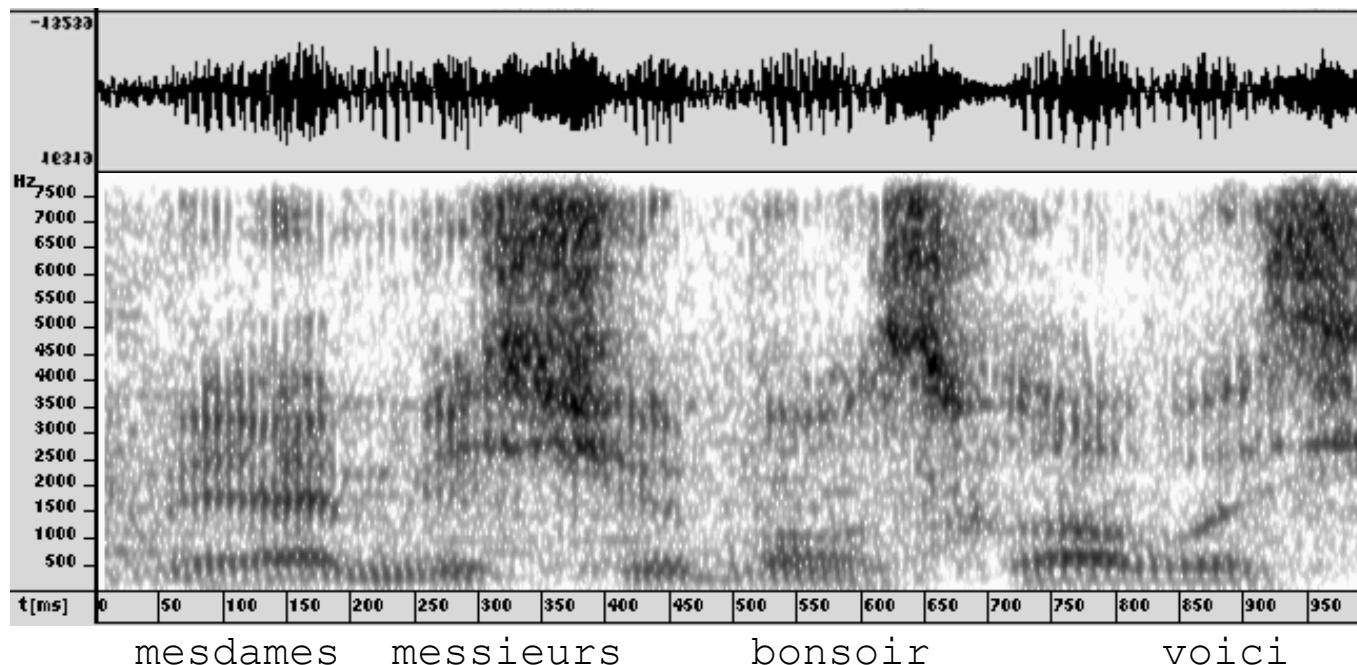
Beaucoup d'informations sont présentes dans un signal de parole :



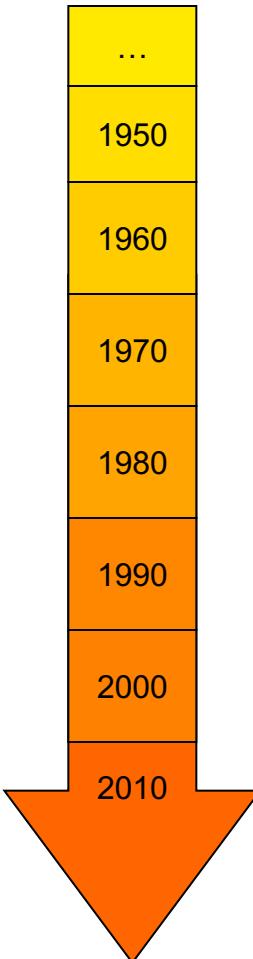
- Aspect non verbal de la voix:
  - ◆ le timbre, la qualité vocale, les disfluences...
  - ◆ la prosodie: rythme + intensité + mélodie

# Complexité de l'oral par rapport à l'écrit

- Signal résultant des contraintes de production, perception et compréhension
  - ◆ Continuité du signal, coarticulation (pas de coupure de mots)
  - ◆ Distorsions temporelles (débit variable)
  - ◆ Variabilité (inter- et intra-locuteurs, conditions acoustiques)
  - ◆ Homophonies (transcriptions différentes, prononciation identique)



# Historique

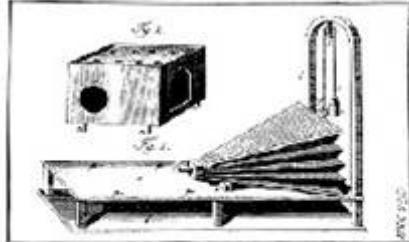


- la "préhistoire"
- apparition des systèmes numériques
- approches par règles 'AI'
- projet ARPA SUR: échec de la compréhension directe
- DTW: reconnaissance de mots isolés, succès de l'approche 'ingénieur'
- approche statistique (Fred Jelinek, Jim Baker à IBM)
- développement des HMM (Rabiner, AT&T)
- campagnes internationales d'évaluation DARPA/NIST
- études sur grands corpus
- DNN: réseaux de neurones et apprentissage profond
- diffusion d'applications grand public (Google, Apple, Nuance)



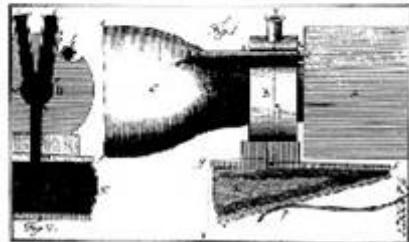
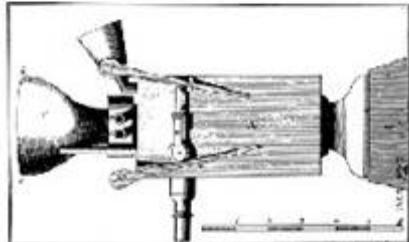
# Les machines parlantes

Von Kempelen  
(1791)



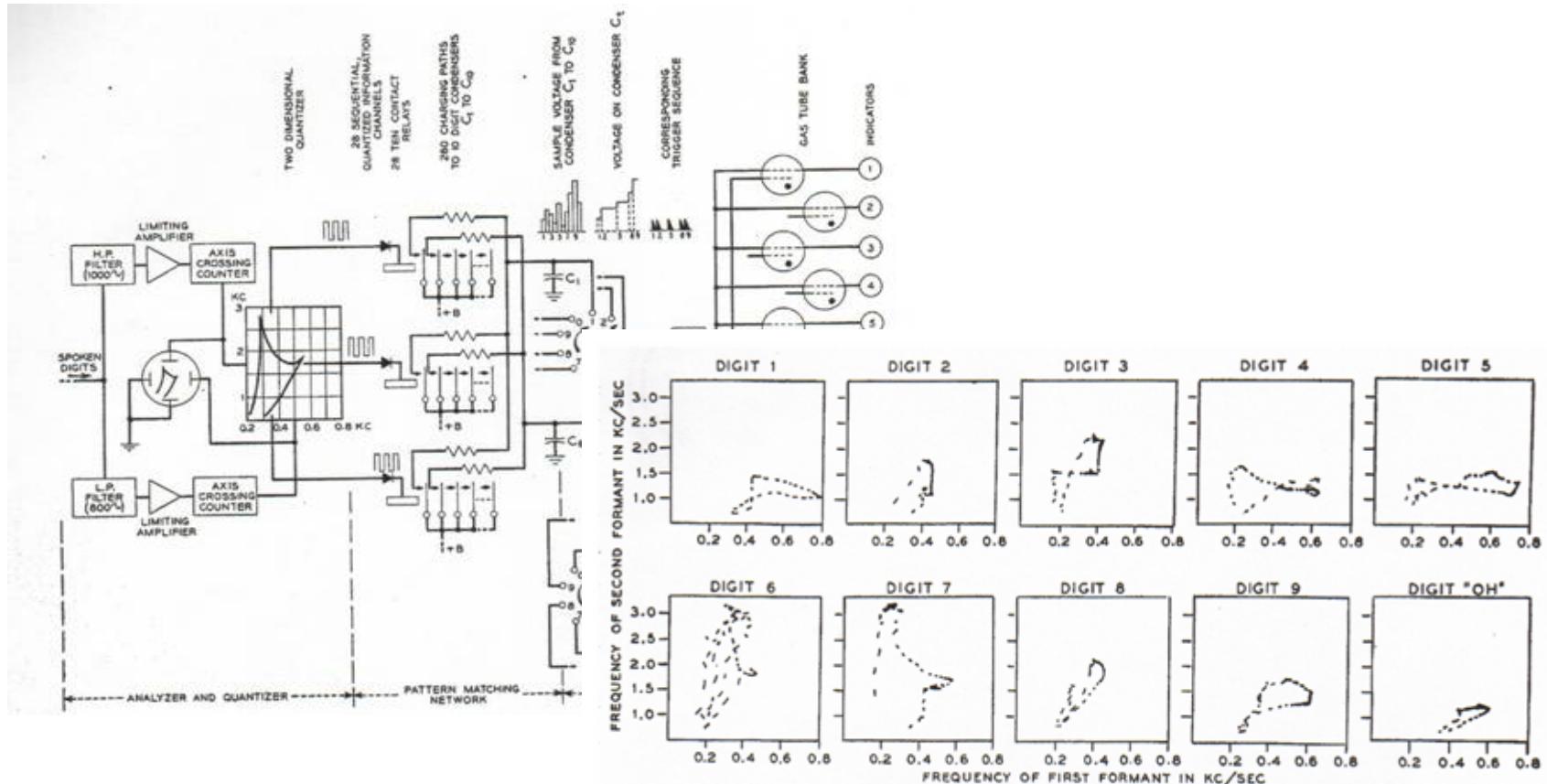
Homer Dudley  
Bell Labs  
(1939)

Joseph Faber  
(1835)



# Reconnaissance de la parole: les premières réalisations

- 1952 – La reconnaissance des chiffres (AUDREY)
- ◆ Davis, Biddulph, Balashek (Bell Laboratories)



# 1960 – Traitement de la parole et machines numériques

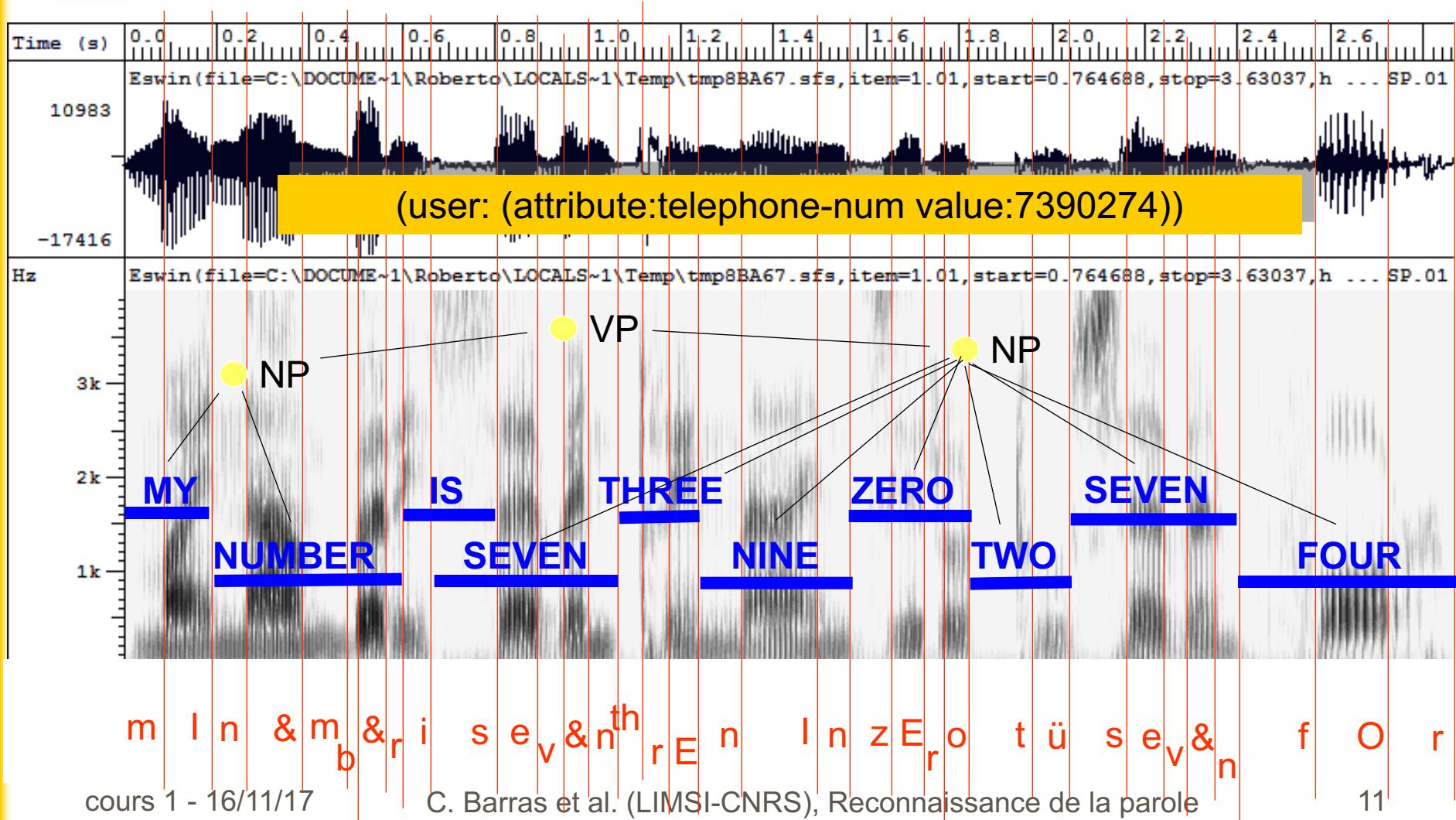
- AnalogicDigital/DigitalAnalogic  
les machines numériques commencent à apparaître dans les laboratoires



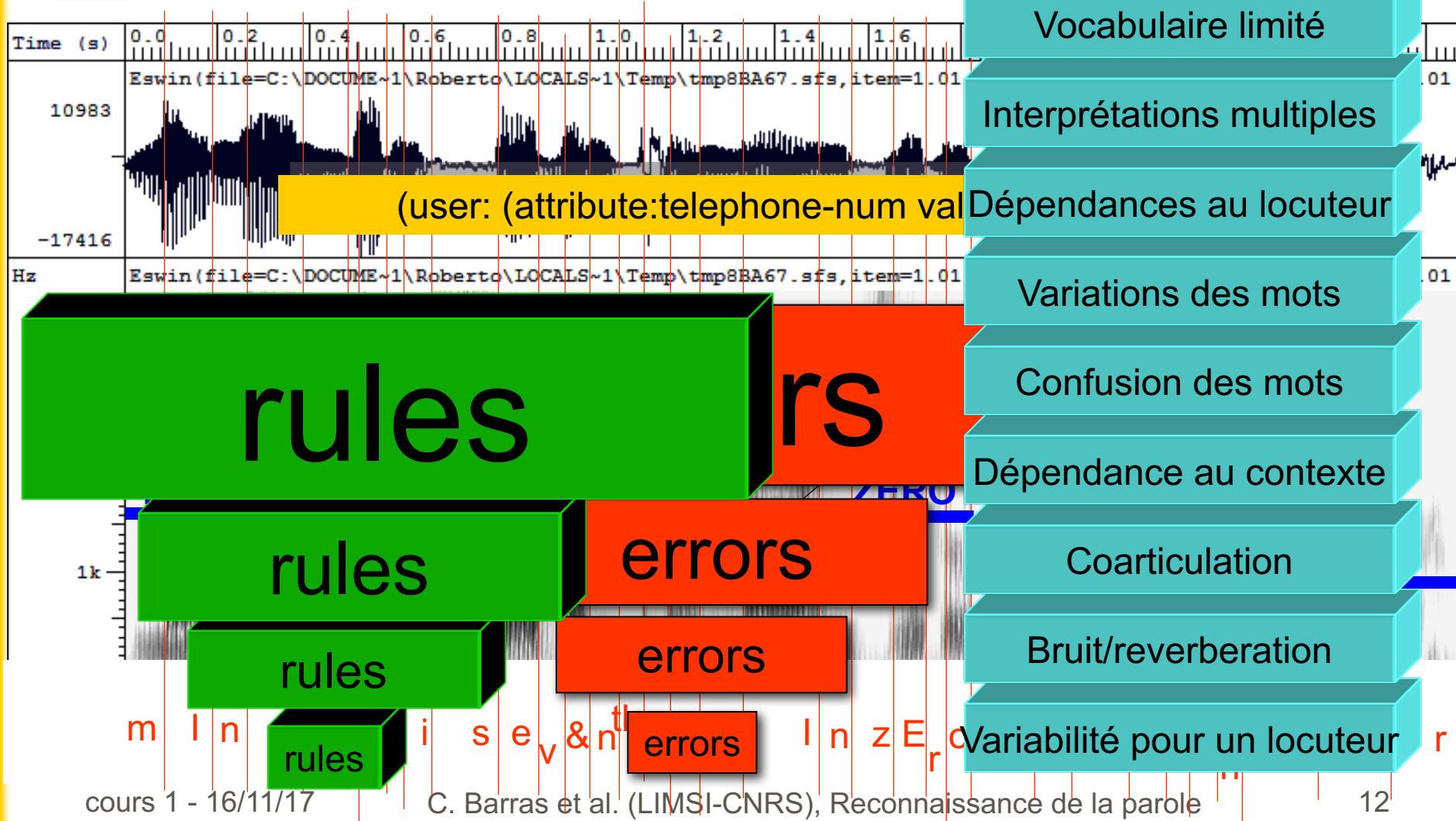
James Flanagan  
Bell Laboratories



# L'illusion de la segmentation...



# Pourquoi c'est si difficile ?



# 1971-1976: Le projet ARPA SUR

- Malgré une campagne anti-reconnaissance de la parole (*Pierce Commission*) ARPA monte un programme de compréhension de la parole sur 5 ans.
- But: 1000 mots de vocabulaire, 90% taux de compréhension, proche du temps réel (machine à 100 MIPS)
- 4 systèmes construits en fin de programme dont
  - ◆ BBN's *HWIM* (44%)
  - ◆ CMU's *Hearsay II* (74%)
  - ◆ CMU's *HARPY* (95% -- 80 times real time!)
- Manque une approche globale de l'évaluation
- Compréhension de la parole: trop tôt... pas de suite

Leçon tirée:

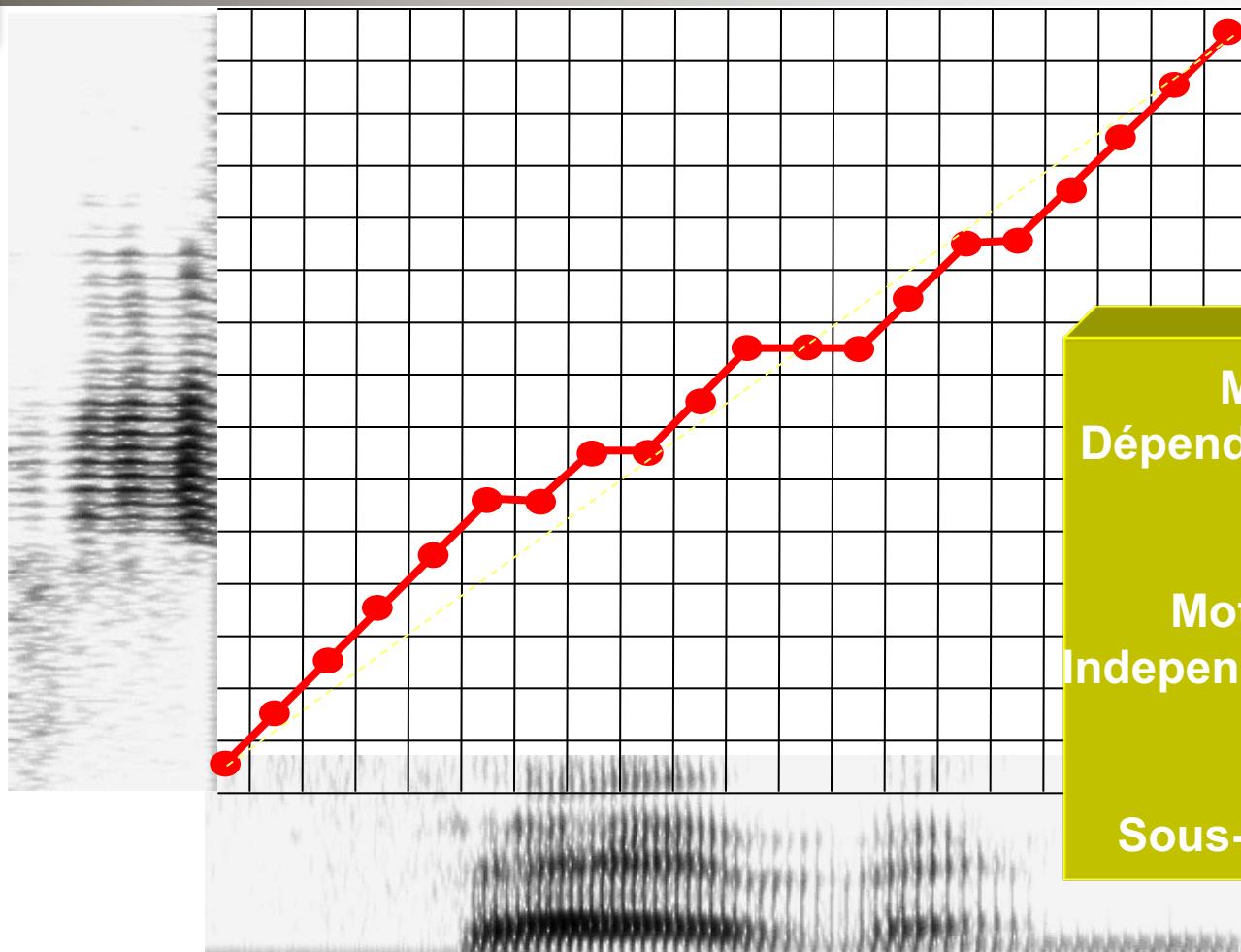
Beaucoup de connaissances “expert”  
Besoin d'un critère d'évaluation globale



Raj Reddy -- CMU

# 1970 – DTW l' approche ingénieur

REFERENCE (MOT 7)



T.K. Vyntsyuk (1969)  
H. Sakoe,  
S. Chiba (1970)

Mots isolés  
Dépendants du locuteur



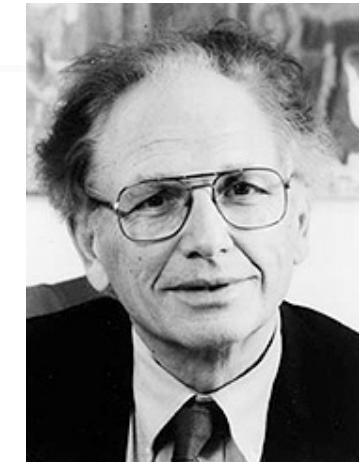
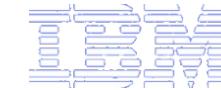
Mots connectés  
Indépendants du locuteur



Sous-unités de mots

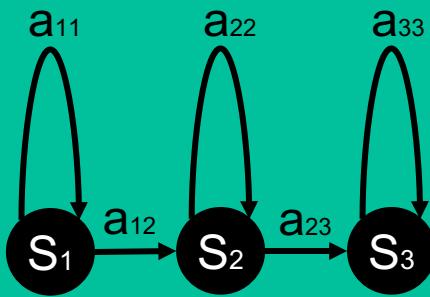
# 1980 – L'approche statistique

- Basée sur les modèles de Markov  
(Leonard Baum at IDA, Princeton à la fin des années 60)
- Fred Jelinek, Jim Baker IBM Watson R
- A la base des systèmes de reconnaissance actuels



Fred Jelinek

## Acoustic HMMs



## Word Tri-grams

$$P(w_t | w_{t-1}, w_{t-2})$$

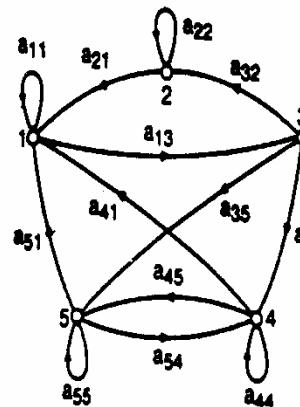
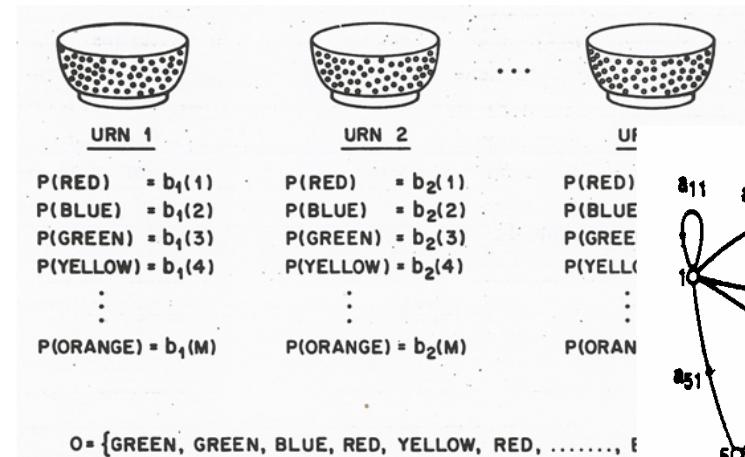


Jim Baker

- No Data Like More Data
- Whenever I fire a linguist, our system performance improves (1988)
- Some of my best friends are linguists (2004)

# 1980-90 Les approches statistiques deviennent omniprésentes

- Lawrence Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceeding of the IEEE, Vol. 77, No. 2, February 1989.



Markov Assumption:

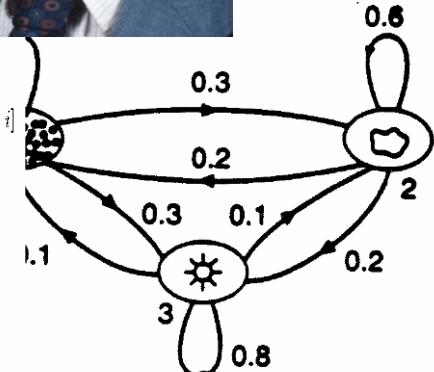
$$P[q_t = j | q_{t-1} = i, q_{t-2} = k, \dots] = P[q_t = j | q_{t-1} = i]$$

Set

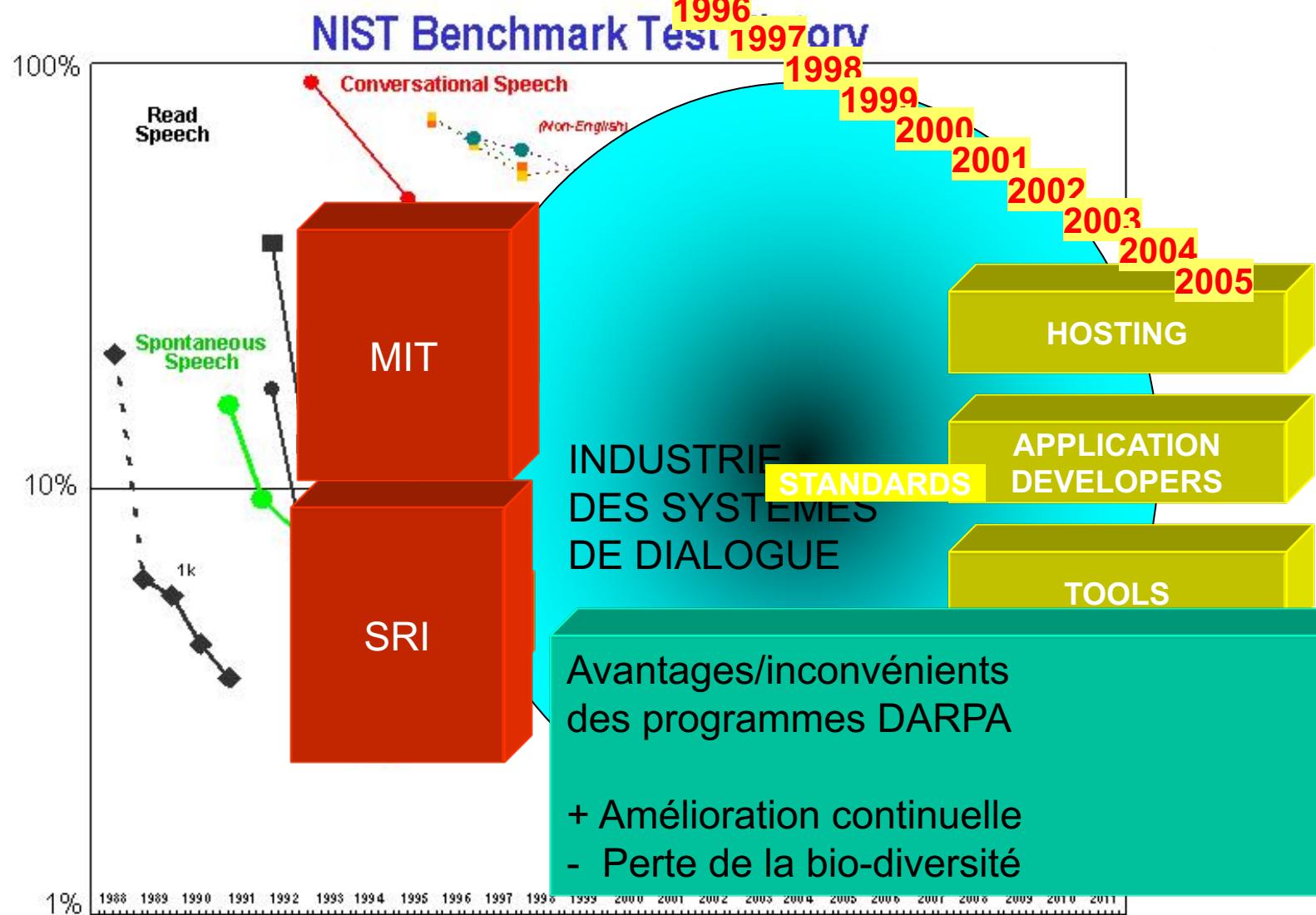
$$a_{ij} = P[q_t = j | q_{t-1} = i] \quad 1 \leq i, j \leq N$$

Such that

$$\begin{aligned} a_{ij} &\geq 0 \quad \forall i, j \\ \sum_{j=1}^N a_{ij} &= 1 \quad \forall i \end{aligned}$$



# 1980s-90s Pouvoir de l'évaluation



# Domaines de recherche concernées

- ◆ physiologie
  - fonctionnement de l'appareil phonatoire et du système auditif
- ◆ acoustique et traitement du signal
  - mécanismes de production/propagation des ondes sonores
  - analyse du signal, en présence de bruit
- ◆ phonétique, phonologie
  - étude des sons du langage
- ◆ autres champs de la linguistique
  - lexique, syntaxe, sémantique, pragmatique...
- ◆ psychologie
  - tests psycholinguistiques et psychoacoustiques
- ◆ théorie de l'information, statistiques, algorithmique
  - mise en œuvre de systèmes automatiques

# Langage / langue / parole

- langage :
  - faculté spécifiquement humaine, universelle
  - système de représentation régi par une grammaire
- langue :
  - réalisation particulière du langage
  - règles et normes partagées par les membres d'une communauté
- parole : usage de la langue, autre que l'écrit

« *La linguistique a un double objet, elle est science du langage et science des langues. Cette distinction (qu'on ne fait pas toujours) est nécessaire : le langage est une faculté humaine, caractéristique universelle et immuable de l'homme et est donc autre chose que les langues qui sont toujours particulières et variables, et en lesquelles le langage se réalise.* »

*Benveniste*

# Les langues dans le monde

## ■ Richesse linguistique

- ◆ 6000 à 7000 langues recensées ([www.ethnologue.com](http://www.ethnologue.com))
  - Continuum entre dialectes et langues

## ■ Diversité d'origine géographique

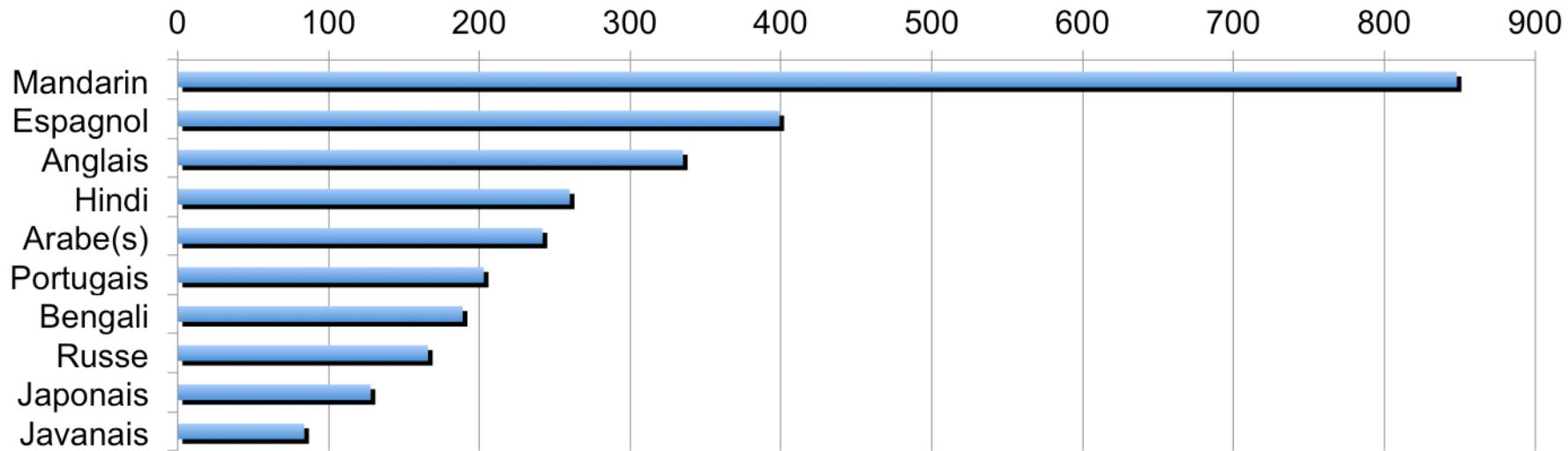
~5%	Europe et Moyen-Orient
~15%	Amériques
~30%	Afrique
~30%	Asie
~20%	Pacifique

## ■ Diversité démographique

- ◆ 6% des langues sont parlées par plus d'un million de personnes et couvrent 94% de la population
- ◆ la moitié sont parlées par moins de 6.000 locuteurs

# Les langues du monde

10 langues maternelles les plus parlées (millions de locuteurs)



source: ethnologue.com, 2015

## ■ Évolution et disparition des langues

- ◆ Atlas de l'UNESCO des langues menacées (3000?)
  - <http://www.unesco.org/culture/languages-atlas/>

# Phonétique et phonologie

<b>Phonologie</b> ex: <i>R</i>	<b>Phonétique</b> ex: <i>R roulé, R grasseyyé</i>
Sciences des <b>phonèmes</b> de la langue	Sciences des <b>sons</b> (phones) des langues (production des sons)
Objet : la langue	Objet: l'acte de parole
méthodes linguistiques	méthodes des sciences naturelles
substance	forme
On décrit les sons fonction des ressemblances et différences phoniques fonctionnelles dans la langue en question	Les sons sont des entités physiques et on les décrits peut importe la langue d'où ils viennent
<b><i>On étudie le statut linguistique des sons (fonctionnel) à l'intérieur d'une langue, d'un système</i></b>	<b><i>Les sons dans leur matérialité, considérés comme des entités physiques et indépendamment de la langue à laquelle ils appartiennent</i></b>
Cherche à établir la fonction des sons dans une langue.	Etudie les sons des langues du monde.

# Variabilité phonologique

- étude sur la base UPSID d'UCLA
  - ◆ ~500 langues représentatives sélectionnées
  - ◆ plus de 900 phonèmes différents
    - (dont 12 partagés par plus de 50% des langues)
  - ◆ en moyenne 25 phonèmes par langue (de 11 à 119!)
    - distribution des voyelles: pic à 5  
(i/a/u/e/o présents dans 1 langue sur 3)
    - distribution des consonnes: pic à 22
    - 10% des langues avec diphongues

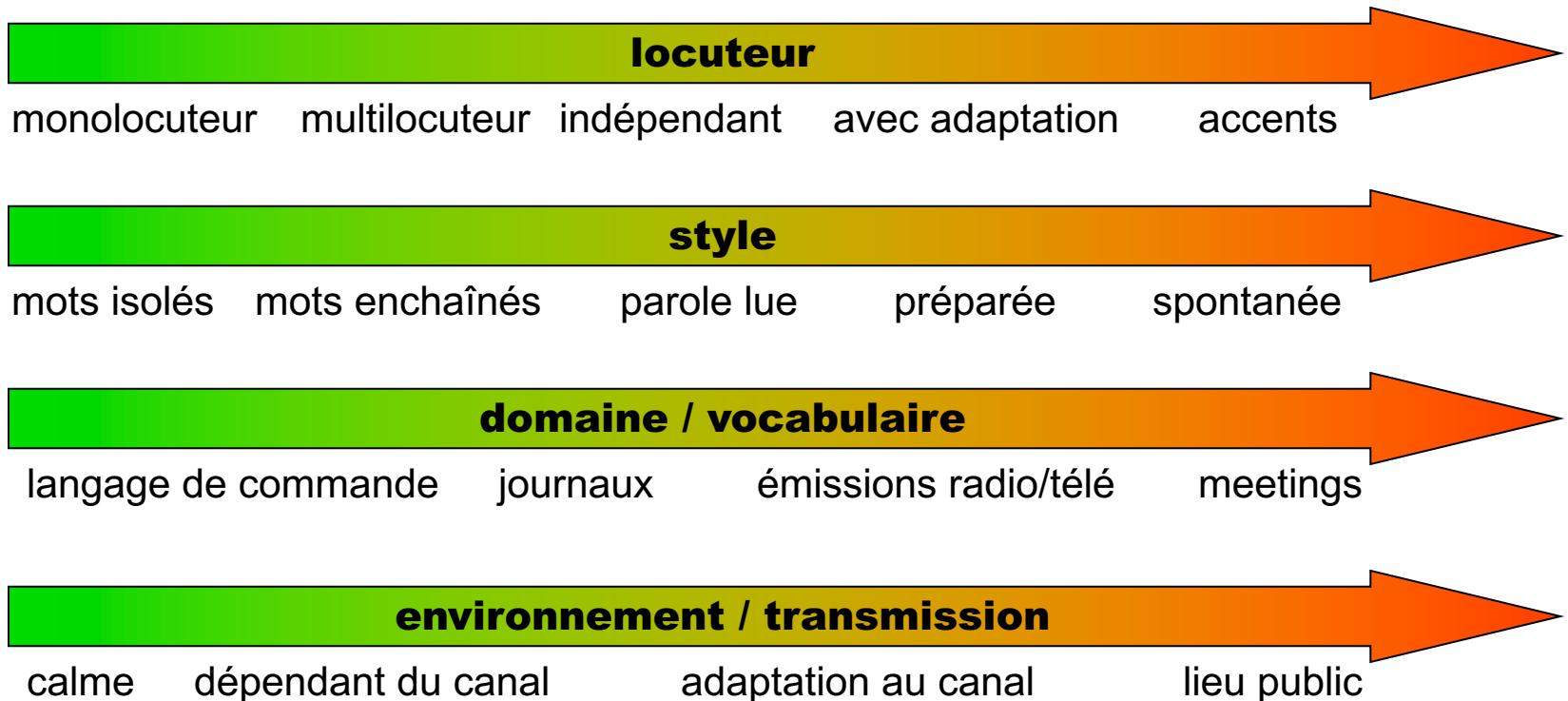
# Niveaux d'analyse

- Propriétés acoustiques et linguistiques :
  - ◆ acoustico-phonétique
    - présence des sons d'une langue
  - ◆ phonotactique
    - fréquence et enchaînement des sons
  - ◆ prosodique
    - intonation, rythme, accentuation
  - ◆ lexical
    - mots possibles d'une langue  
Le petit Larousse : 22000 mots communs, 170000 formes fléchies
  - ◆ syntaxique
    - enchaînement possible de mots dans la langue
  - ◆ sémantique
    - sens de l'enchaînement des mots
  - ◆ pragmatique
    - informations relatives au contexte

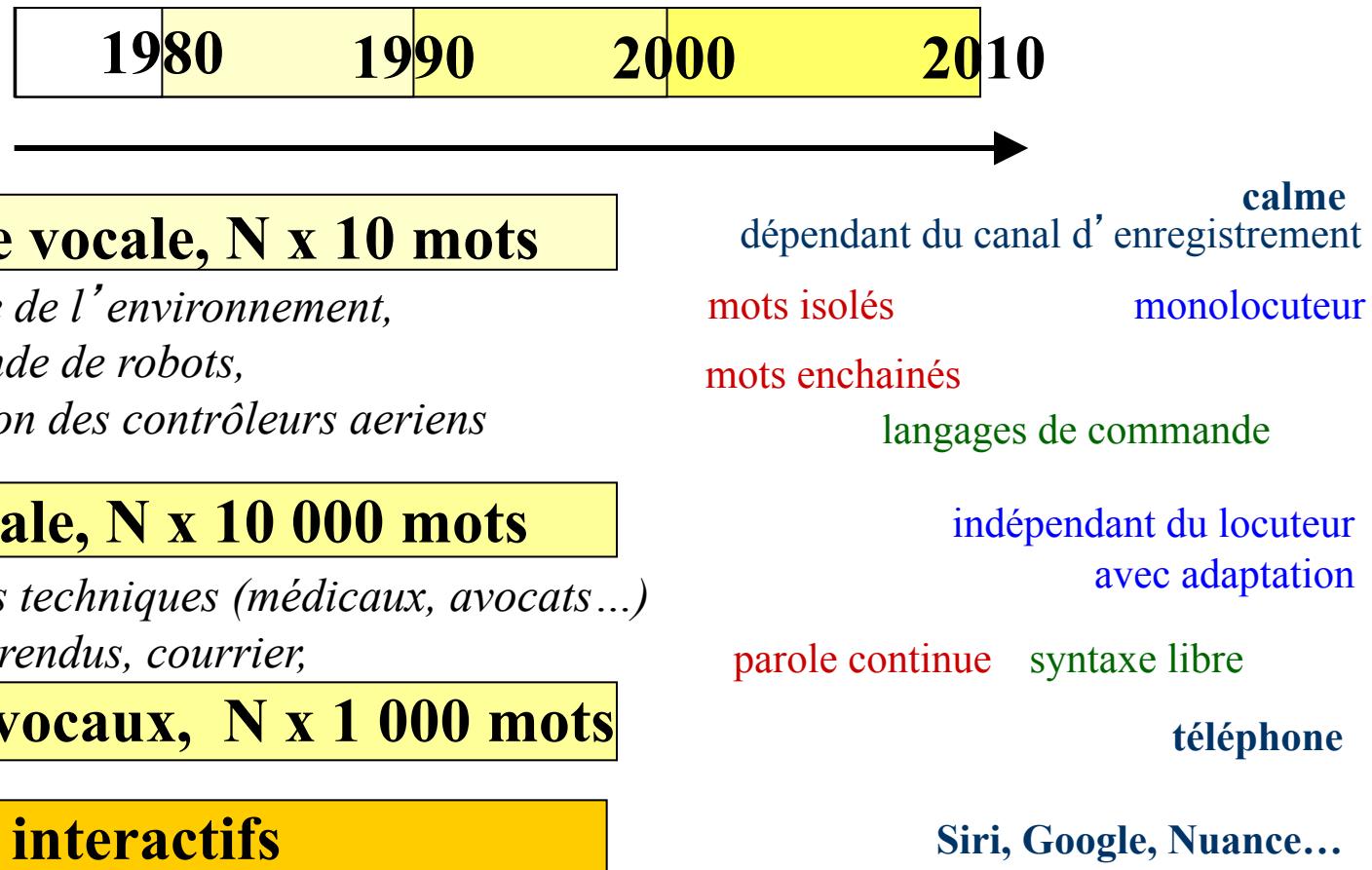
# Difficulté suivant les langues

- Systèmes phonologiques spécifiques
- Structure syllabique et tonale
- Simplicité de la correspondance oral/écrit
- Proportion des formes fléchies, des compositions
  - ◆ allemand - langue compositionnelle
    - /BodenKontrollRadarAnlage/ dispositif de contrôle radar au sol
  - ◆ hongrois – langue agglutinante
    - /vòrlak/ je t'attends, /vòrhattlak/ je pouvais t'attendre
- Découpage en mots - exemples en Français:
  - *les arts, lézards*
  - /õemyrmyrdəmekötätmä/: 340 découpages en mots, 6 syntaxiquement corrects, 1 correct sémantiquement
    - *un murmure de mécontentement*
    - *un murmure de mécontentes ment*
    - *un mur mûr de mes contentements*
    - ...

# Difficultés du contexte applicatif



# Evolution des applications

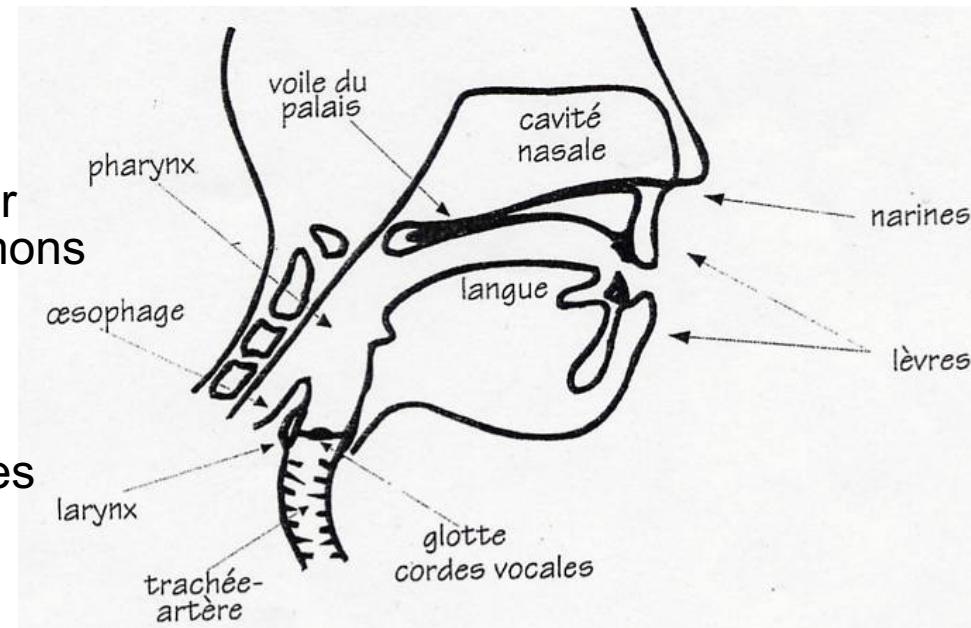


# Production de la parole

## ■ Appareil vocal

- ◆ Poumons et trachée-artère
  - production d'un souffle d'air par compression des poumons
- ◆ Larynx
  - cordes vocales (ouverture/fermeture de la glotte)
  - vibration des cordes vocales  
→ hauteur mélodique
- ◆ Conduit vocal
  - pharynx, cavité buccale, cavité labiale, cavité nasale
  - organes articulateurs influant sur la forme du conduit : mâchoire, lèvres, langue

## ■ Sources sonores résonant dans le conduit vocal



# Cordes vocales

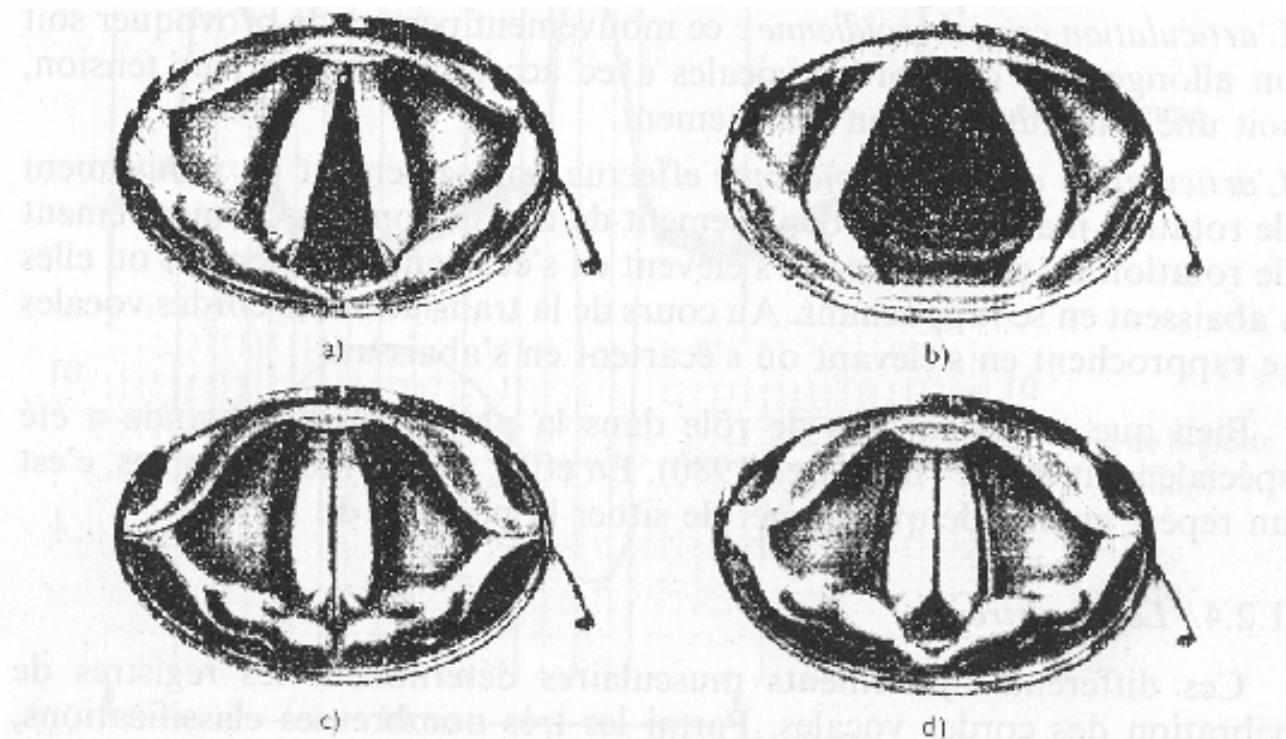
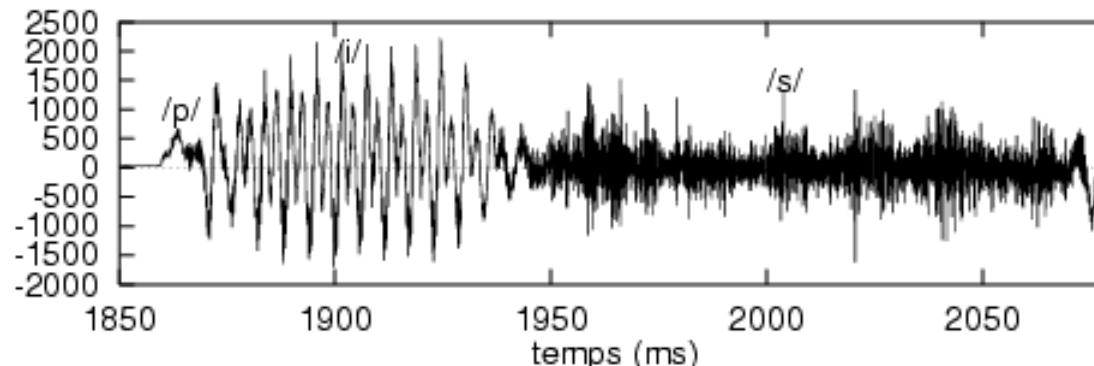


FIG. II.10. — La glotte en position a) de respiration b) d'inspiration profonde, c) de phonation, d) pour la voix chuchotée (PERNKOPF, 1952).

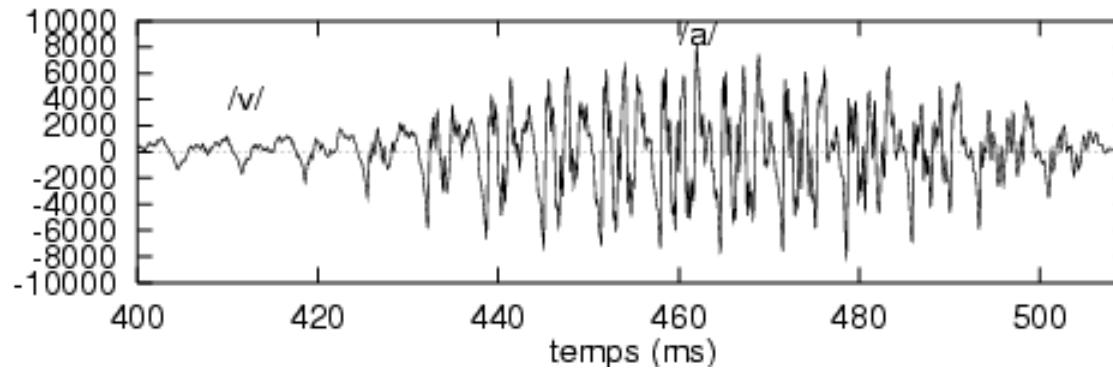
# Sources de production de la parole

- Trois modes d'excitation de la source
  - ◆ Vibrations quasi-périodiques des cordes vocales
    - Fréquence fondamentale ( $F_0$ )
    - Son voisé (voyelles)
  - ◆ Bruits d'écoulement d'air, constriction
    - Son fricatif (fricatives ou constrictives)
  - ◆ Occlusions rapides, impulsions
    - Son avec explosion (occlusives ou plosives)

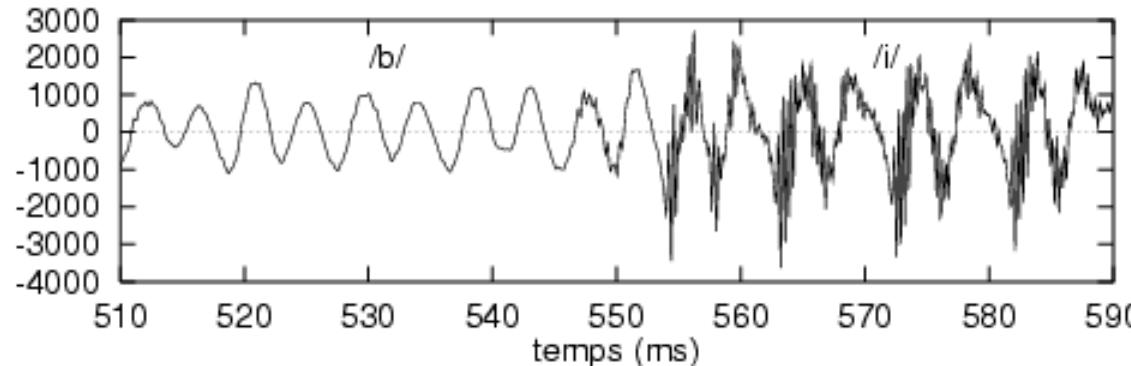


# Sources de production de la parole

- Combinaison des différents modes d'excitation
  - Friction + voisement



- Impulsion + voisement



# Résonateur et formants

- Le signal source produit par l'excitateur se propage dans un volume appelé aussi résonateur
- Les composantes fréquentielles de l'excitation sont affaiblies ou renforcées dans le résonateur
  - ◆ Dépend du volume de la cavité et de ses ouvertures
- Les fréquences de résonances sont appelées **formants**
  - ◆ Caractéristique du **timbre**
- Les premiers formants sont les plus importants
  - ◆ Position des 3 premiers pour caractériser une voyelle
  - ◆ Ex. pour la voyelle [i] : 300, 2200 et 3000 Hz

# Conduit vocal et lieux d'articulation

- Paramètres en phonétique articulatoire
  - ◆ Point d'articulation
    - point où la langue est la plus proche du palais
  - ◆ Aperture
    - Section du conduit vocal au point d'articulation
  - ◆ Labialisation
    - Forme des lèvres
  - ◆ Nasalité
  - ◆ Latéralité
    - Passage de l'air de part et d'autre de la langue
- Typiquement 36 phonèmes en français
  - ◆ Notation avec l'alphabet phonétique international (API)

# Phonèmes du français

## voyelles orales

/i/	pie	/a/	patte
/e/	été	/ɑ/	pâte
/ɛ/	modèle	/ɔ/	auditeur
/y/	puni	/ɔ/	porte
/ø/	deux	/u/	poux
/œ/	peur	/ə/	petite

## voyelles nasales

/ã/	an	/œ̃/	brun
/ɛ̃/	matin	/ɔ̃/	bon

# Phonèmes du français (2)

<b>plosives orales</b>	labiales	alvéolaires	vélaires
sourdes	/p/ : <b>poids</b>	/t/ : <b>toit</b>	/k/ : <b>quoi</b>
voisées	/b/ : <b>bois</b>	/d/ : <b>doigt</b>	/g/ : <b>goût</b>

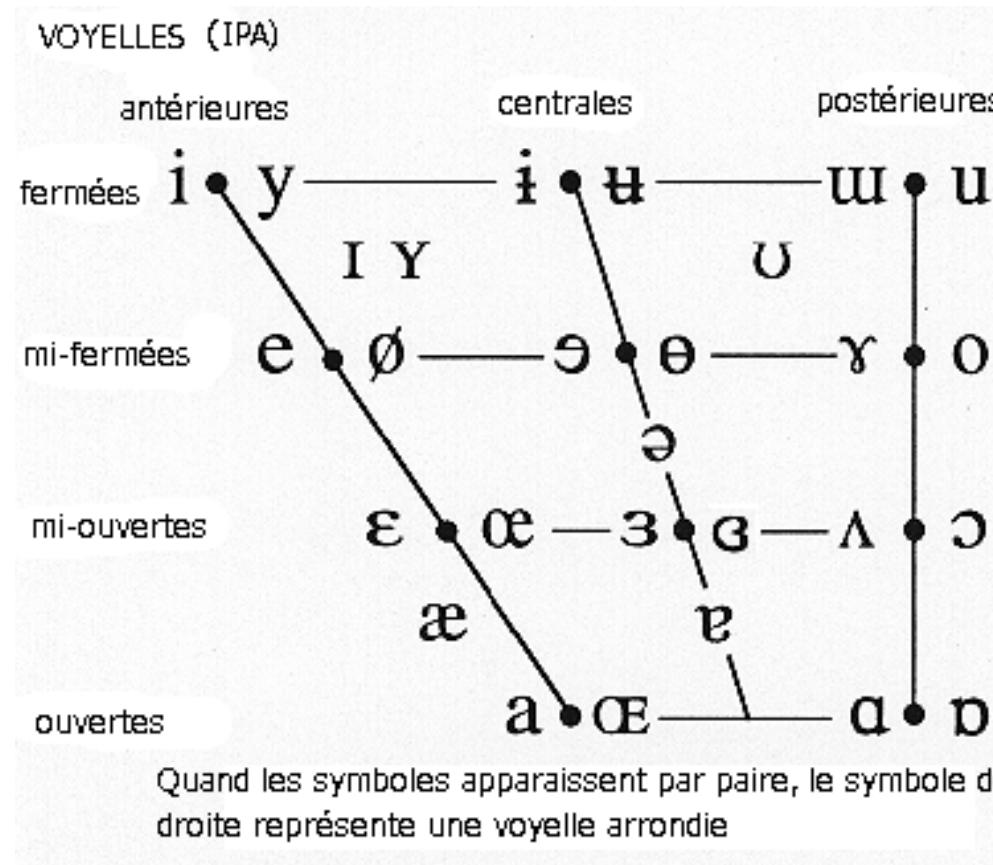
<b>occlusives nasales</b>	labiale	alvéolaire	palatale	
	/m/ : <b>mon</b>	/n/ : <b>nous</b>	/ɲ/ : <b>agneau</b>	/ŋ/ : <b>smoking</b>

<b>fricatives</b>	dentales	alvéolaires	post-alvéolaires
sourdes	/f/ : <b>feu</b>	/s/ : <b>soir</b>	/ʃ/ : <b>poche</b>
voisées	/v/ : <b>voix</b>	/z/ : <b>zéro</b>	/ʒ/ : <b>jeu</b>

<b>liquides</b>	/l/ : <b>long</b>	/ʁ/ : <b>rond</b>
-----------------	-------------------	-------------------

<b>semi-voyelles</b>	/w/ : <b>oui</b>	/j/ : <b>piège</b>	/ɥ/ : <b>lui</b>
----------------------	------------------	--------------------	------------------

# Articulation des voyelles



# Coarticulation et variantes phonologiques

- La coarticulation
  - ◆ Effet d'inertie articulatoire => minimisation de l'effort articulatoire
  - ◆ Modification importante de la réalisation acoustique en fonction du contexte phonétique
- Assimilation de sonorités
  - ◆ schwa (e) élidé en élocution rapide
  - ◆ plosives, fricatives sonores sont dévoisées si la consonne suivante est sourde:
    - méd(e)cin -> /metsin/
  - ◆ plosives, fricatives sourdes sont voisées si la consonne suivante est sonore:
    - pâqu(e)bot -> /pagbo/
  - ◆ Exception: cheveu, cheval

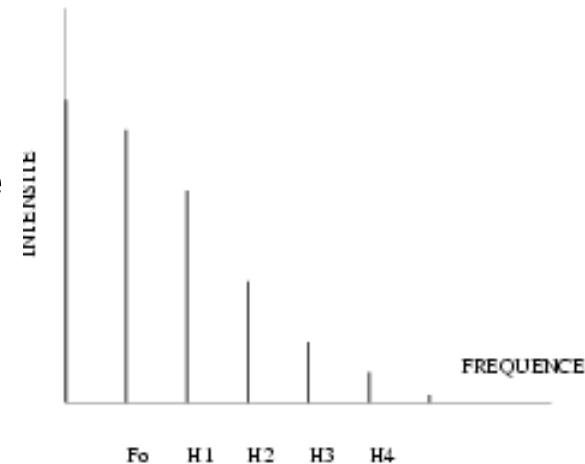
# F<sub>0</sub> et harmoniques

## ■ Fréquence fondamentale

- ◆ la fréquence de vibration des cordes vocales des sons voisés est appelée **fréquence fondamentale** ou F<sub>0</sub>
- ◆ elle correspond à la hauteur « musicale » du son

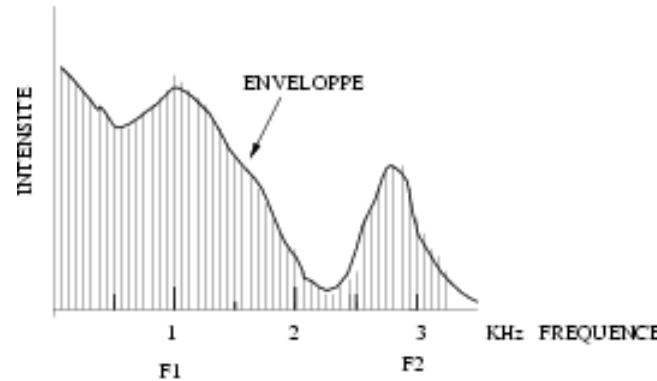
## ■ Le spectre d'un signal (quasi-)périodique est un spectre de raies aux multiples entiers de la fréquence fondamentale ou **harmoniques** (H<sub>1</sub> = 2 F<sub>0</sub>, H<sub>2</sub> = 3 F<sub>0</sub>...)

- F<sub>0</sub> évolue lentement dans le temps et n'est pas spécifique d'un phonème
- Elle est plus élevée pour les femmes que les hommes (voix plus aigues)
- 70 à 250 Hz pour les hommes
- 150 à 400 Hz pour les femmes
- 200 à 600 Hz pour les enfants



# Enveloppe spectrale et formants

- En reliant les maxima des raies par une ligne continue on obtient l'enveloppe du spectre.
- Les fréquences où les maxima de cette enveloppe ont lieu sont les formants du signal  $F_1, F_2\dots$
- Les formants sont spécifiques de chaque phonème
  - ◆ exemple de spectre de son /a/

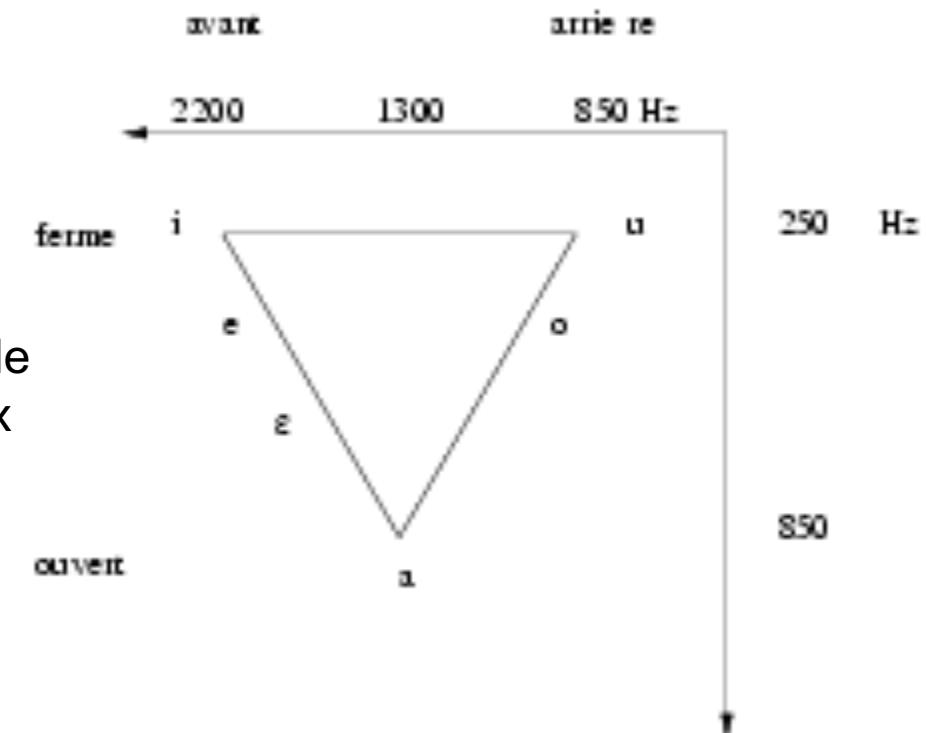


# Triangle vocalique

- Statistiquement, les 2 premiers formants

d

- /u/ : F1 = 250 Hz et F2 = 800 Hz
- /a/ : F1 = 850 Hz et F2 = 1300 Hz
- /i/ : F1 = 250 Hz et F2 = 2200 Hz



- Il est possible de représenter le triangle vocalique en superposant sur les deux axes avant-arrière et fermé-ouvert, deux autres axes : premier formant F1 et deuxième formant F2.

# Reconnaissance de mots isolés

- Problème de reconnaissance des formes
  - ◆ une ou plusieurs formes acoustiques de référence  $R_m$  pour chaque mot  $m$  du vocabulaire
  - ◆ une forme de test  $O$
  - ◆ une distance  $D$  entre formes acoustiques

$$\tilde{m} = \operatorname{argmin}_m D(O, R_m)$$

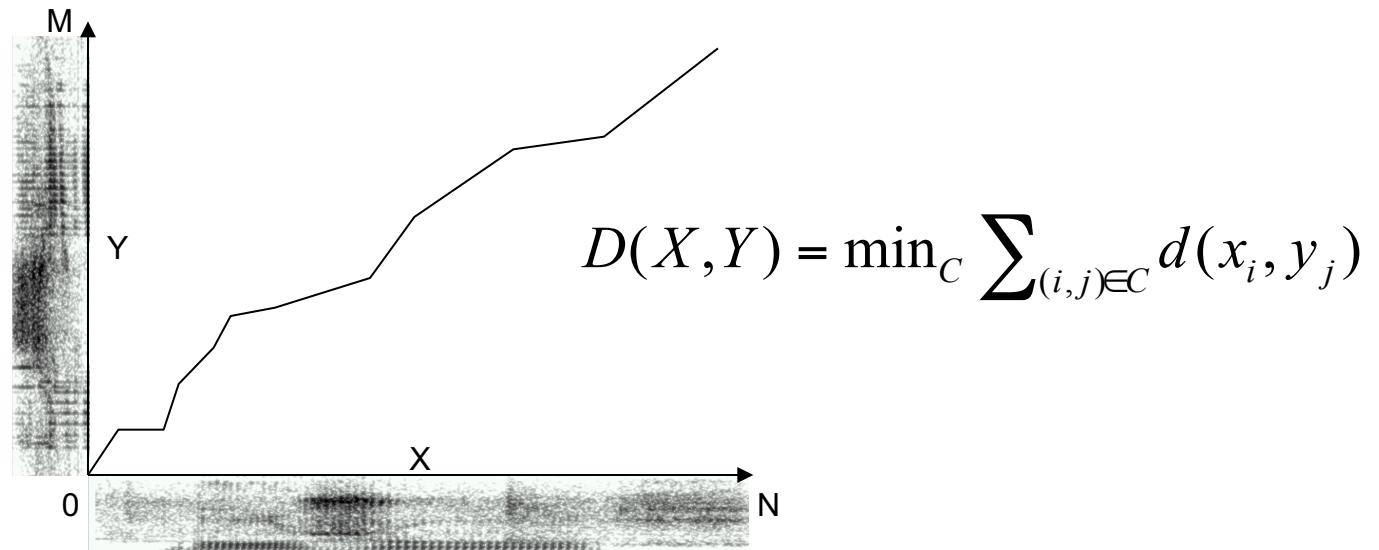
- Forme acoustique
  - ◆ une suite de vecteurs acoustiques
- Distance entre formes acoustiques  $D(X, Y)$  ?
  - ◆ à partir d'une distance acoustique locale  $d(x_i, y_j)$
  - ◆ déformation temporelle linéaire?

$$D(X, Y) = \sum_{i=1, j=\frac{M}{N}}^N d(x_i, y_j)$$

- ◆ insuffisant! il peut y avoir des déformations temporelles non linéaires

# Problème d'alignement

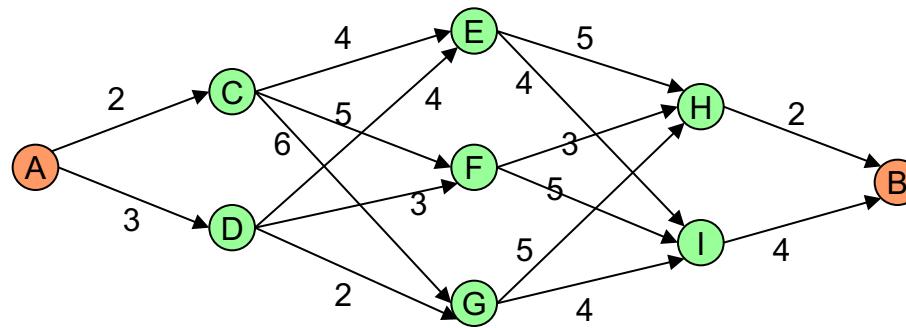
- Distance acoustique globale  $D(X, Y)$  ?
  - ◆ cumul des distances locales sur le « meilleur » chemin



- Problème
  - ◆ explosion combinatoire du nombre de chemins!

# Programmation dynamique

- Recherche du plus court chemin dans un graphe de A à B

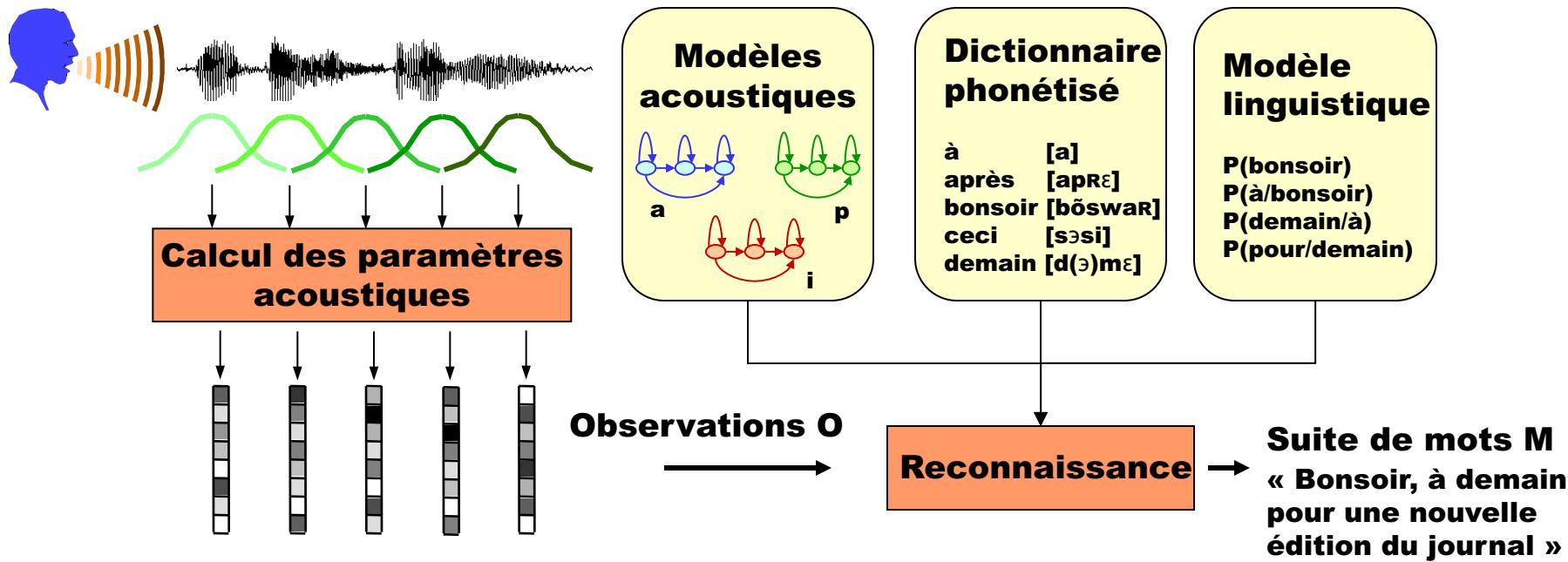


- Principe d'optimalité de Bellman (1957) :  
tout sous-chemin d'un chemin optimal est optimal
- Application
  - ◆ si le chemin optimal A-B passe par E, le chemin A-E est optimal
  - ◆ construction du chemin A-B par recurrence, par extension des sous-chemins optimaux et en ignorant les autres

# Alignment temporel dynamique

- Déformation temporelle dynamique (DTW, Dynamic Time Warping)
  - ◆ principe de la programmation dynamique (Vintsyuk, 1968)
  - ◆ détermination du « meilleur » chemin
  - ◆ calcul par récurrence
- Reconnaissance de mots isolés, à petit vocabulaire, en mono-locuteur
  - ◆ paramètres acoustiques
    - paramètres dynamiques, normalisation spectrale
  - ◆ détection de début et fin de mots
- Extensions
  - ◆ ajout d'un mot nouveau au vocabulaire
    - enregistrement du mot par le locuteur
  - ◆ multi-locuteurs
    - plusieurs références par mot
  - ◆ mots enchaînés
    - level-building, one-pass
- Applications
  - ◆ commande vocale simple (par ex. téléphonie)

# Transcription automatique de parole continue



$$\tilde{M} = \arg \max_{M \in E_M} P(M|O)$$

# Règle de Bayes

- Quelle est la suite de mots M la plus probable étant donnée l'observation O ?

$$\begin{aligned}
 \tilde{M} &= \arg \max_{M \in E_M} P(M|O) \\
 &= \arg \max_{M \in E_M} \frac{P(O/M) \cdot P(M)}{P(O)} \\
 &= \arg \max_{M \in E_M} P(O/M) \cdot P(M)
 \end{aligned}$$



Modélisation acoustique

# Modèle linguistique

## Probabilité d'une suite de mots

$$P(m_1 \dots m_K) = P(m_1) \prod_{i=2}^K P(m_i | m_{i-1} \dots m_1)$$

## Hypothèse n-gram : chaîne markovienne d'ordre n-1

$$P(m_i | m_{i-1} \dots m_1) \approx P(m_i | m_{i-1} \dots m_{i-n+1})$$

## Estimation de bigrammes, trigrammes, quadrigrammes

$$\tilde{P}(m_j | m_i) = \frac{f(m_i, m_j)}{f(m_i)}$$

Mesdames messieurs bonsoir voici les  
titres de l' actualité de ce lundi

# Estimation du modèle linguistique

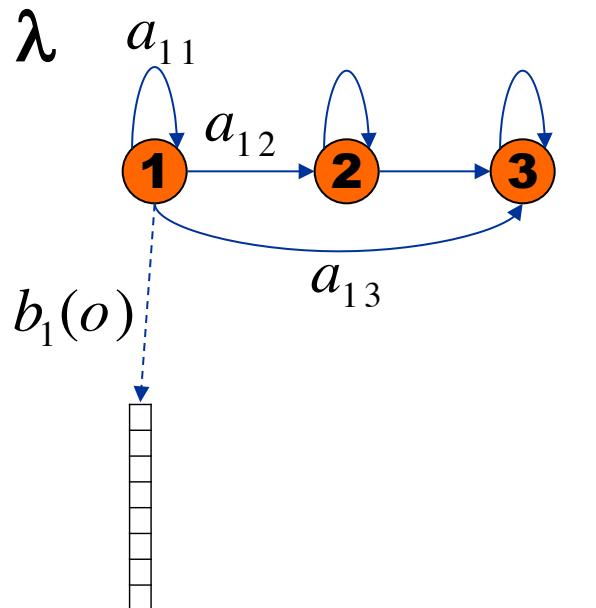
- Les probabilités a priori de suites de mots sont estimées à partir de corpus de textes
  - ◆ typiquement : plusieurs centaines de millions de mots provenant des archives de journaux (Le Monde, etc.)
- Etapes:
  - ◆ Collecte de corpus
  - ◆ Prétraitements : formatage, normalisations
  - ◆ Définition du vocabulaire
  - ◆ Comptes de mots, suites de mots
  - ◆ Estimation des N-grammes
  - ◆ Lissage : probabilité d'observation de suites non observées

# Modélisation acoustique

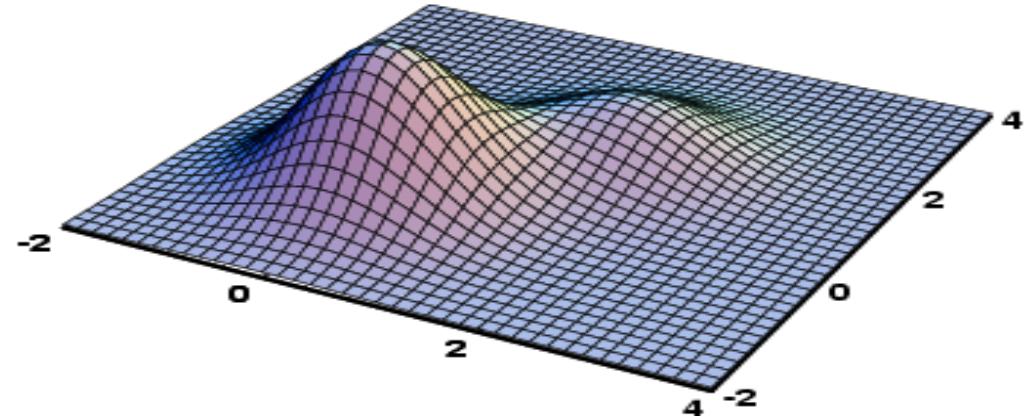
## Modèles de Markov cachés (HMM) :

Automate probabiliste  $\lambda = (\text{états}, a_{ij}, b_j)$

- Transitions entre états « cachés » qui correspondent aux configurations de réalisation des phones
- Emission d'observations : distribution statistique dans l'espace des paramètres acoustiques (MFCC)



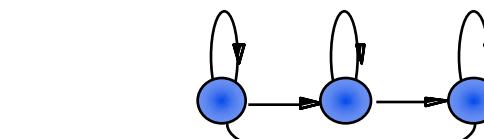
**en suivant une densité de probabilité multi-gaussienne ou un réseau neuronal**



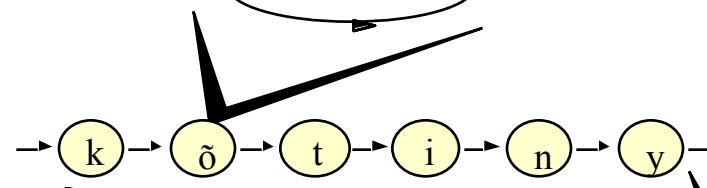
# Décodage en parole continue

**Recherche du chemin optimal  
dans le graphe des phrases possibles  
en utilisant le dictionnaire de prononciation  
et la grammaire probabiliste**

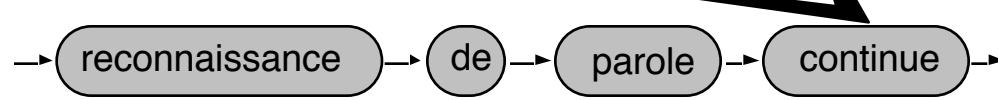
Modèle de phonème



Modèle de mot



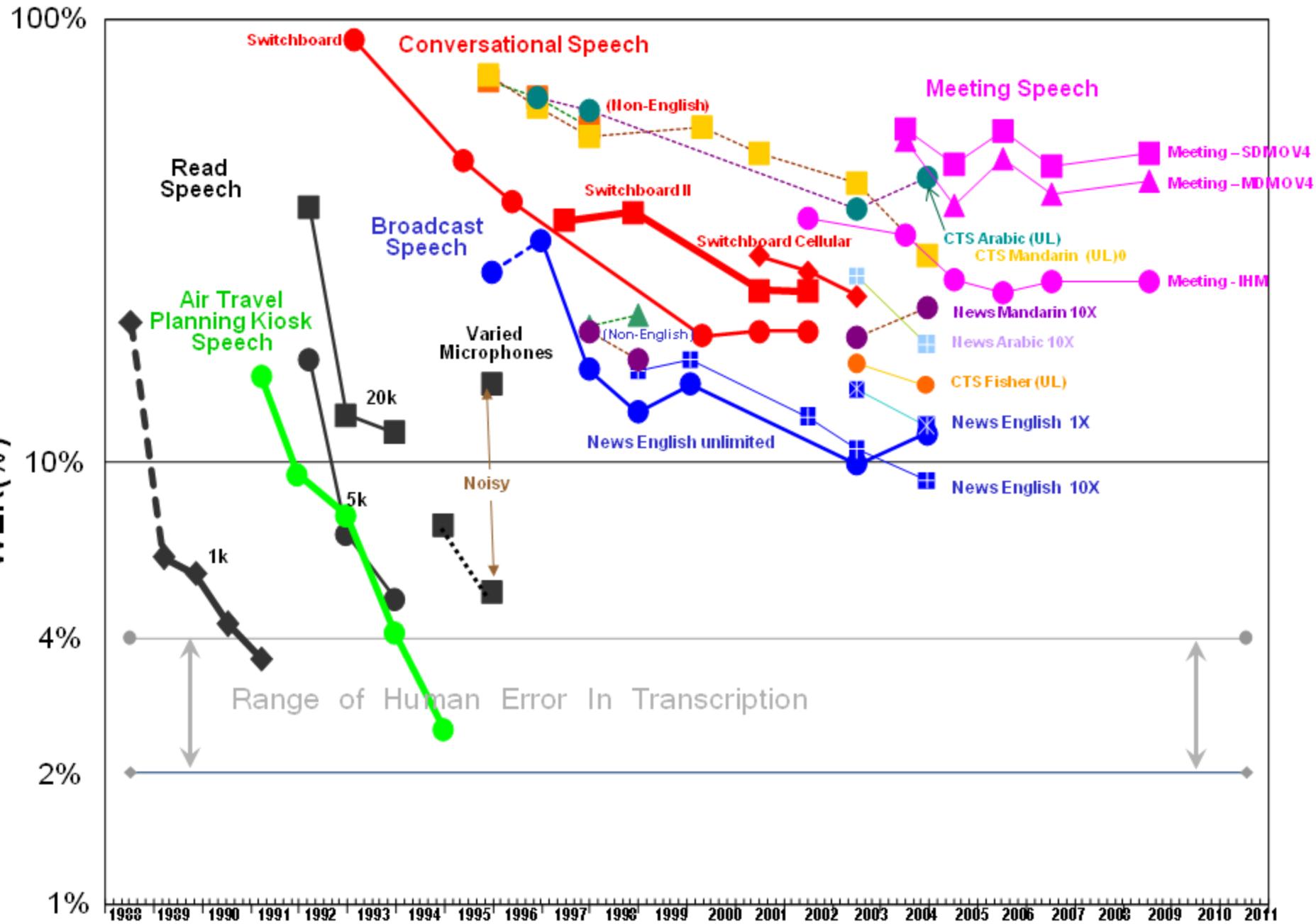
Modèle de phrase



# Les données d'apprentissage

- Grandes quantités de parole transcrive et de textes
  - ◆ > 100 h d'audio avec transcription manuelle fine et synchronisées (> 1 M mots)  
(inspirations, hésitations, conditions acoustiques, tours de parole avec identification des locuteurs...)
  - ◆ > 100 M mots de transcriptions type "revue de presse"
  - ◆ > 1 G mots de textes de journaux
- Production des données: organismes et outils dédiés
  - ◆ LDC aux États-Unis, ELRA en Europe
- Impulsion des campagnes d'évaluation
  - ◆ évaluations organisées par DARPA et NIST  
(d'abord sur l'anglais américain, ensuite le mandarin et l'arabe)
  - ◆ évaluations francophones (ESTER, ETAPE...)

# NIST STT Benchmark Test History – May. '09



# Perspectives

- Des progrès considérables, une technologie déployée dans des applications grand public avec les assistants vocaux interactifs
- ...mais encore loin des performances humaines:
  - ◆ changement de conditions acoustiques, bruit
  - ◆ accents régionaux et étrangers
  - ◆ parole superposée
- défis toujours actuels
  - ◆ traduction automatique de l'oral
  - ◆ interaction multimodale
  - ◆ affective computing
  - ◆ recherche d'information multilingue
  - ◆ compte-rendu/résumé automatique de réunions

# Ressources scientifiques

## Sociétés savantes, conférences et journaux spécialisés

- ◆ ISCA (International Speech Communication Association) - [www.isca-speech.org](http://www.isca-speech.org)
  - revue **Speech Communication**
  - congrès **Interspeech** tous les ans + divers workshop (ITRW)
- ◆ IEEE (Institute of Electrical & Electronics Engineers) / Signal Processing Society
- ◆ ACM (Association for Computing Machinery)
  - revue **IEEE/ACM Transactions on Audio, Speech and Language Processing**
  - congrès **IEEE ICASSP** tous les ans
- ◆ AFCP (Association Francophone de la communication parlée) - [www.afcp-parole.org](http://www.afcp-parole.org)
  - congrès des **JEP** (journées d'étude sur la parole) tous les 2 ans
- ◆ Autres
  - revue **Computer, Speech and Language**
  - revue **JASA** éditée par l'ASA (Acoustical Society of America)
  - nombreuses conférences liées au traitement des langues (TALN, ATALA...), à la linguistique, la phonétique, la traduction automatique...

# Ressources scientifiques (2)

- Laboratoires de recherche
  - ◆ large dispersion en France: région parisienne, Avignon, Le Mans, Grenoble, Nancy, Rennes, Toulouse... (cf. sites de l'ISCA et de l'AFCP)
- Programmes de recherche
  - ◆ programmes nationaux (projets ANR), projets européens, évaluations américaines (DARPA/NIST)
- Production et diffusion de ressources linguistiques
  - ◆ **LDC** (Linguistic Data Consortium)
  - ◆ **ELRA** (European Language Resources Association)
- Bibliographie générale en français
  - ◆ J-S. Liénard, «Les processus de la communication parlée », Masson, 1977.
  - ◆ Calliope, « La parole et son traitement automatique », Masson, 1989.
  - ◆ R. Boite et al., « Traitement de la parole », Presses Polytechniques et Universitaires Romandes, 2000.
  - ◆ J. Mariani et al., « Traitement automatique du langage parlé » (vol. 1 et 2), Hermès Science, 2002