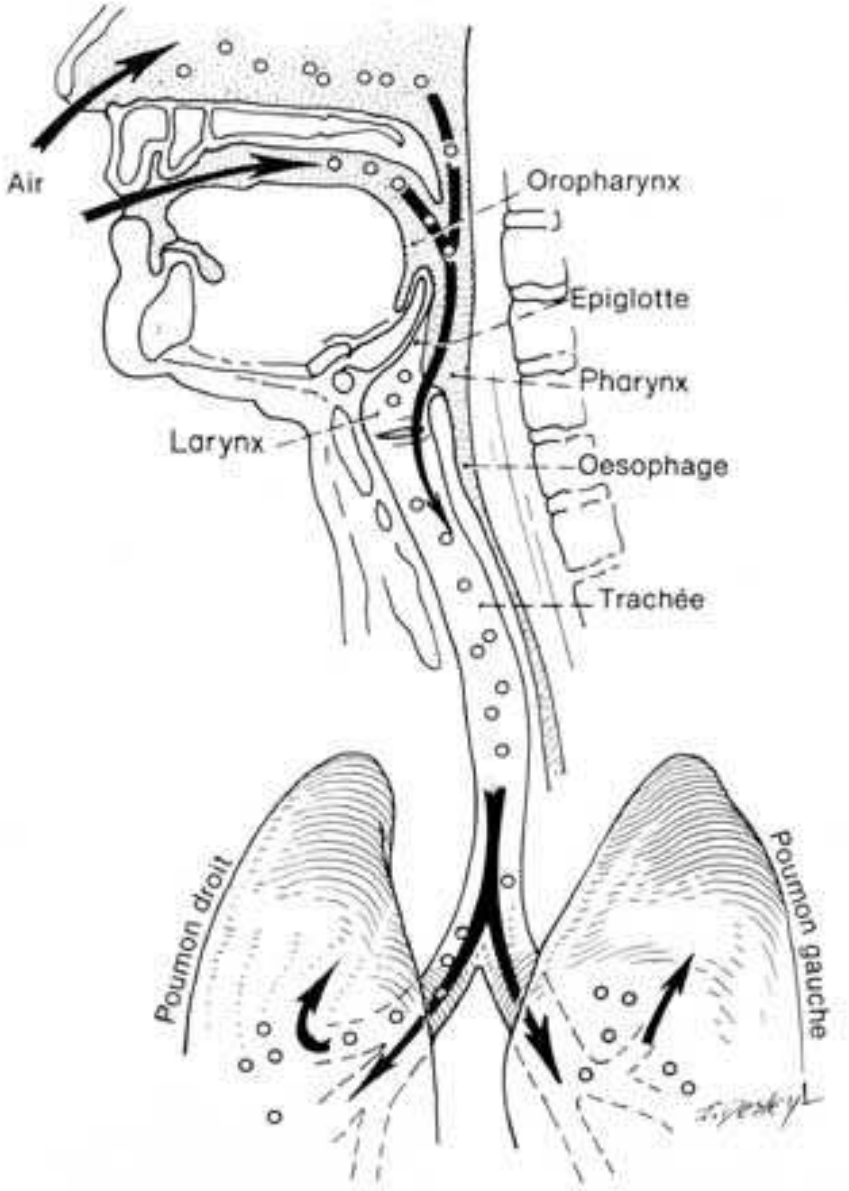
UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Traitement Automatique des Langues

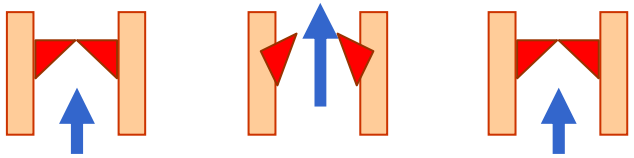
TRAITEMENT DE PAROLE

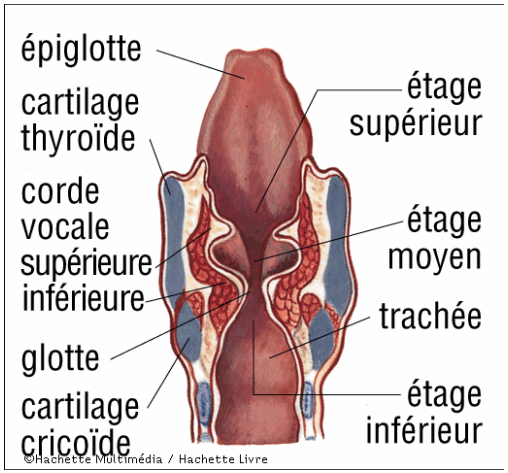
Production : appareil phonatoire

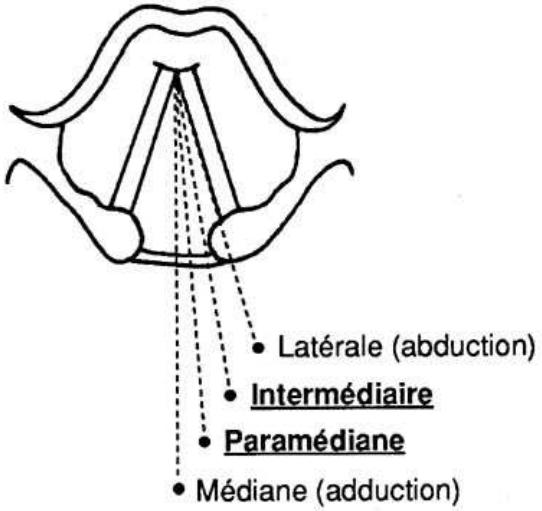


Production : appareil phonatoire

- **Cordes vocales**







- **Voisement** **production avec vibration des cordes vocales**
  - son très énergétique (son **sonores** *vs* son sourd)
  - **fréquence fondamentale** : fréquence de vibration des cordes vocales
  - **variabilité de la fréquence fondamentale**
    - hommes 60 - 100 Hz
    - enfants / femmes 200-300 Hz
    - au cours de la production : « programmation » prosodique

tu manges du poisson ➔ affirmation
tu manges du poisson ↗ question

Production : articulation

Cordes vocales

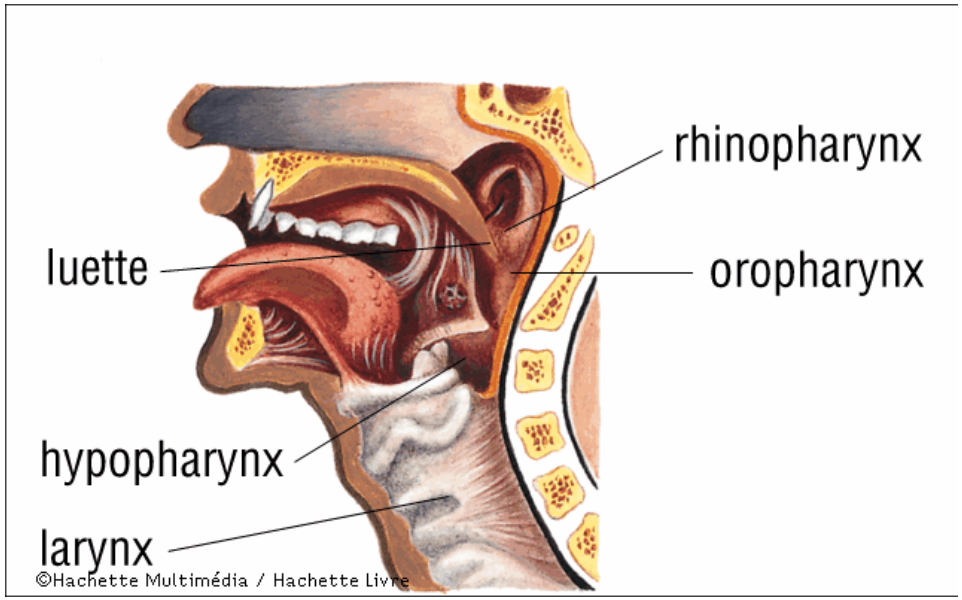
⇨ son voisé / non voisé

Cavité nasale

⇨ son nasal ou non

Langue

- haut (fermé) / bas (ouvert)
- avant / arrière



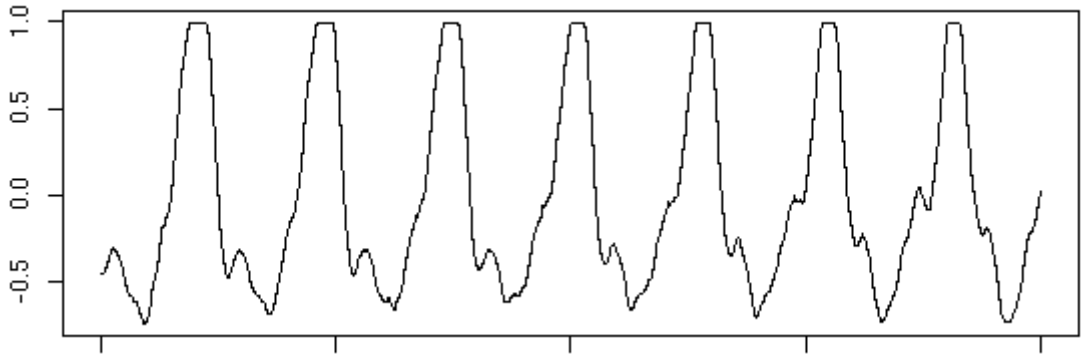
Dents / lèvres

Lieu d'articulation

- lèvres : son labial
- dents : son dental, labio-dental
- palais : son alvéolaire (avant), palatal, vélaire (arrière)

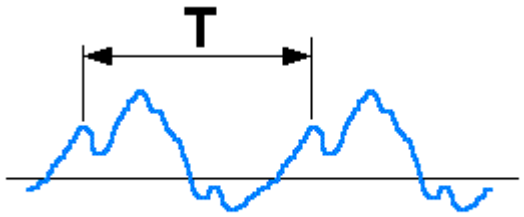
Analyse du signal de parole

Son : la variation de la pression acoustique au cours du temps peut-être décrite par une fonction : **signal acoustique**



Signal périodique : reproduction du même signal suivant un période **T**

- On dit qu'un signal f(t) est périodique si ∃T réel tq ∀t f(t) = f(t+T)
- La **période** d'un signal périodique est le plus petit réel strictement positif T tel que ∀t f(t) = f(t+T). La **fréquence** du signal est donnée par F = 1/T (Hertz)

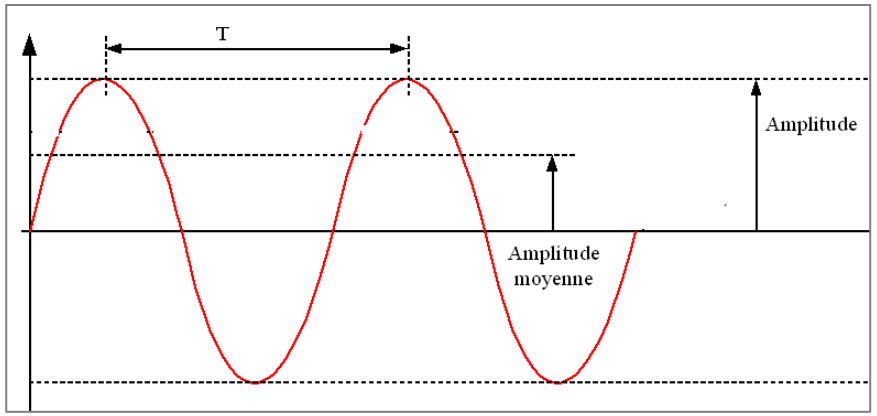


Analyse du signal de parole

Signal sinusoïdal

- $f(t) = A.\cos(2\pi\,\omega t + \varphi)$
- $g(t) = A.\sin(2\pi\,\omega t + \varphi)$

  - période : $1/\omega$
  - amplitude : A
  - phase : φ



- **Représentation complexe** $e^{i(2\pi\omega t + \varphi)} = A\,.\,(\sin(2\pi\omega t+\varphi) + i\,.\,\cos(2\pi\omega t+\varphi))$

Analyse en série de Fourier

Tout signal périodique de période $T = 1/\omega$ peut-être décomposé en une somme infinie de fonctions sinusoïdales (**série de Fourier**) de fréquences multiples $n.\omega$:

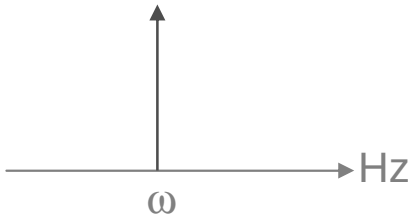
$$f(t) = \sum_{n=-\infty}^{n=+\infty} C_n(f).e^{i2\pi n\omega t}$$

Avec les **coefficients de Fourier** : $C_n(f) = \frac{1}{T}\int_{-T/2}^{T/2} f(t).e^{-2i\pi n\omega t}\,dt$

Analyse du signal de parole

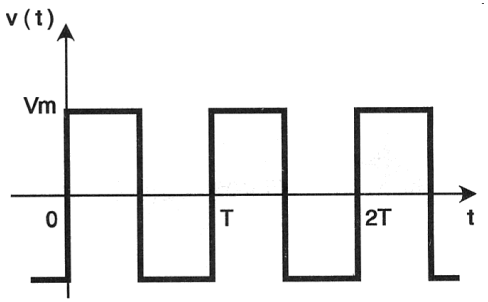
Signal sinusoïdal

- Un seul coefficient de Fourier non nul : C_0
- Signal pur : **spectre** à une seule composante fréquentielle



Signal périodique quelconque

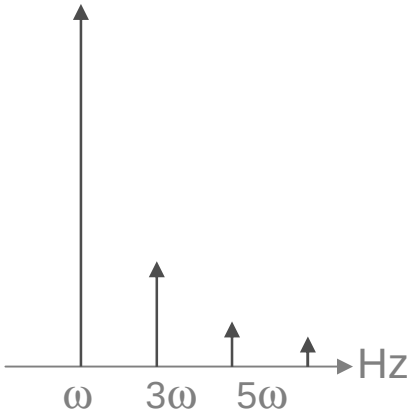
Exemple : fonction créneau



C_0 (f) = 1 **fréq. fondamentale** ω

C_n (f) = 1/π.n si n impair fréq. **Harmoniques**

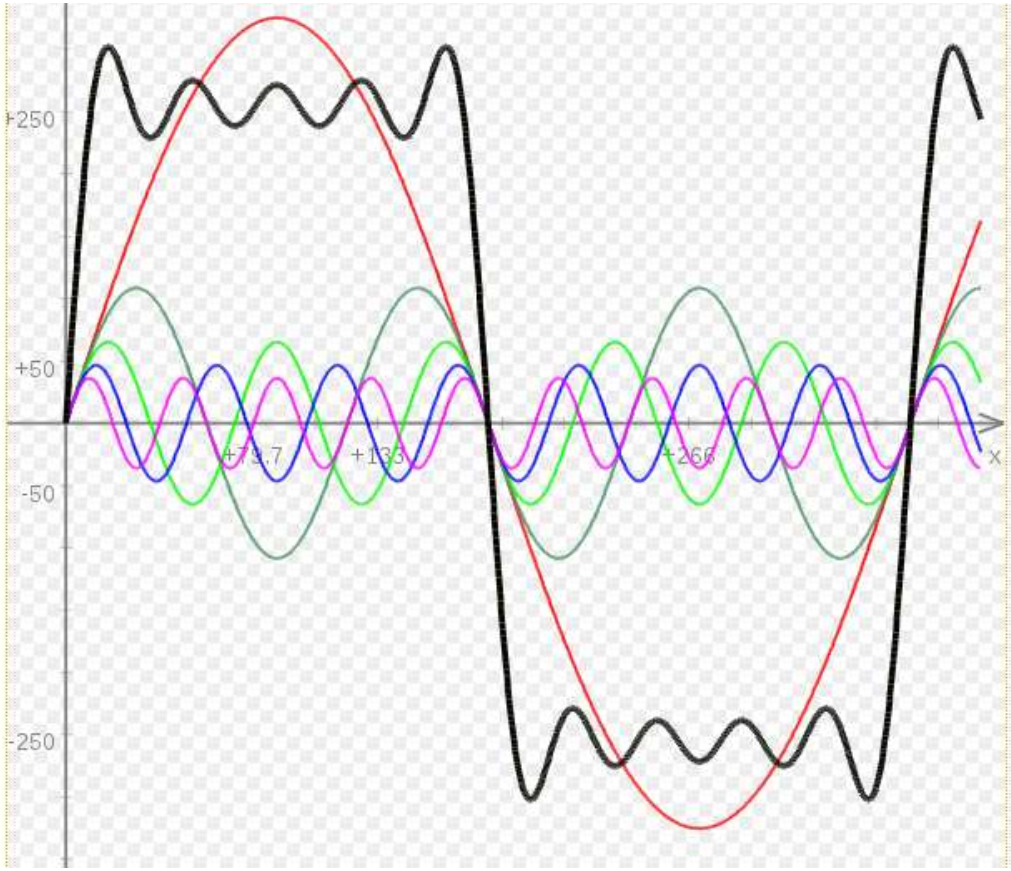
C_n (f) = 0 si n non nul pair



- **Signal multi-fréquentiel** : spectre à plusieurs bandes de fréquences d'amplitude décroissante

Analyse du signal de parole

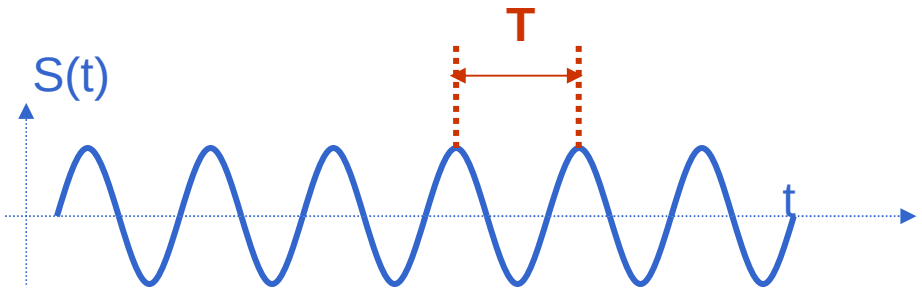
Exemple : somme des cinq premières fonction de Fourier d'une fonction créneau

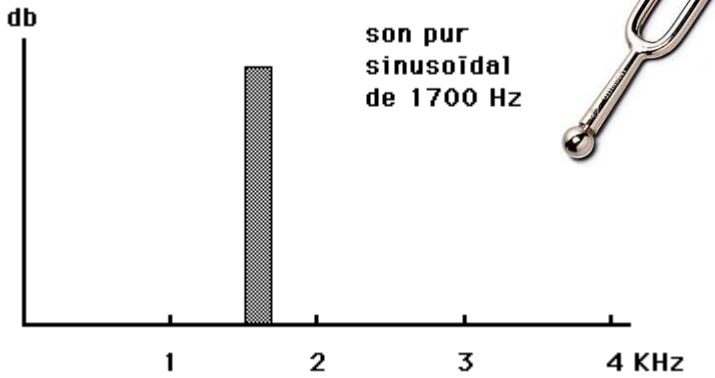


Analyse du signal de parole

Son pur

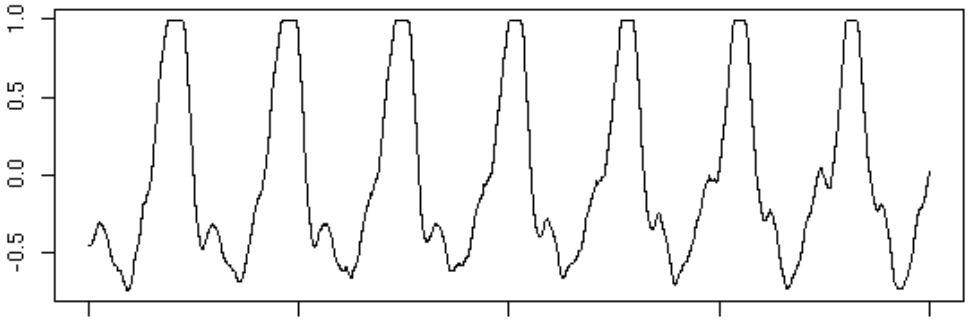
- Signal sinusoïdal monofréquentiel : **fréquence fondamentale** F0

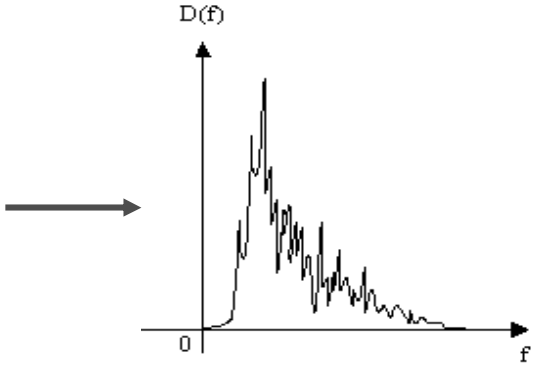




Son quelconque

- Signal complexe pour lequel l'hypothèse de périodicité ne tient que localement
- Adaptation de l'analyse de Fourier : **spectre continu**





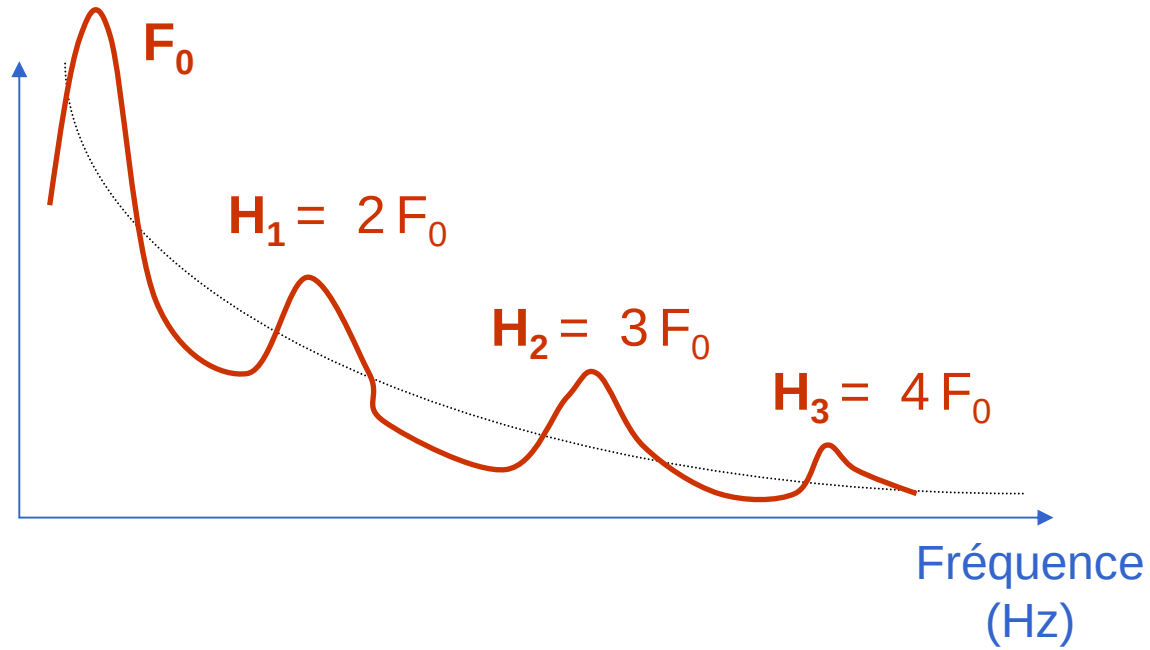
Analyse du signal de parole

Signal de parole

Son voisé à la sortie des cordes vocales

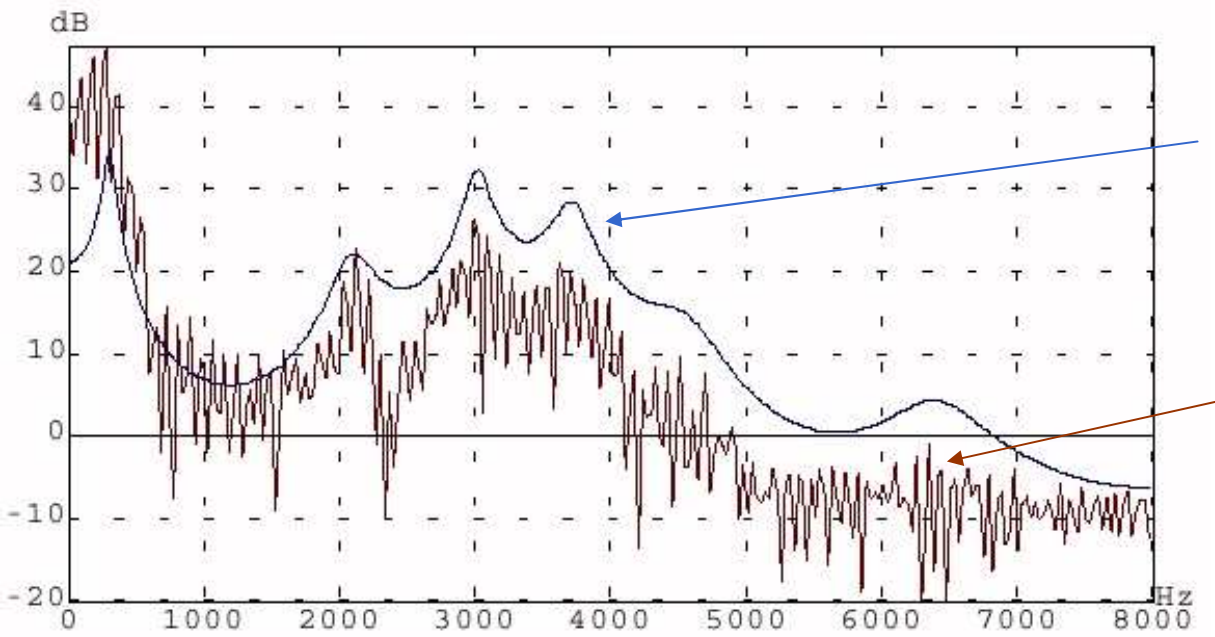
– **Fréquence fondamentale F0** : composante fréquentielle principale, égale à la fréquence de vibration des cordes vocales

– Autres maximum locaux du spectre : fréquences **harmoniques** multiples de la F0.



Analyse du signal de parole

- **Articulation**
  - lieux de constrictions : cavités de résonance
  - fréquences de résonance dues aux articulateurs : **formants**
  - **spectre du signal** : Fo + harmoniques + formantsFi

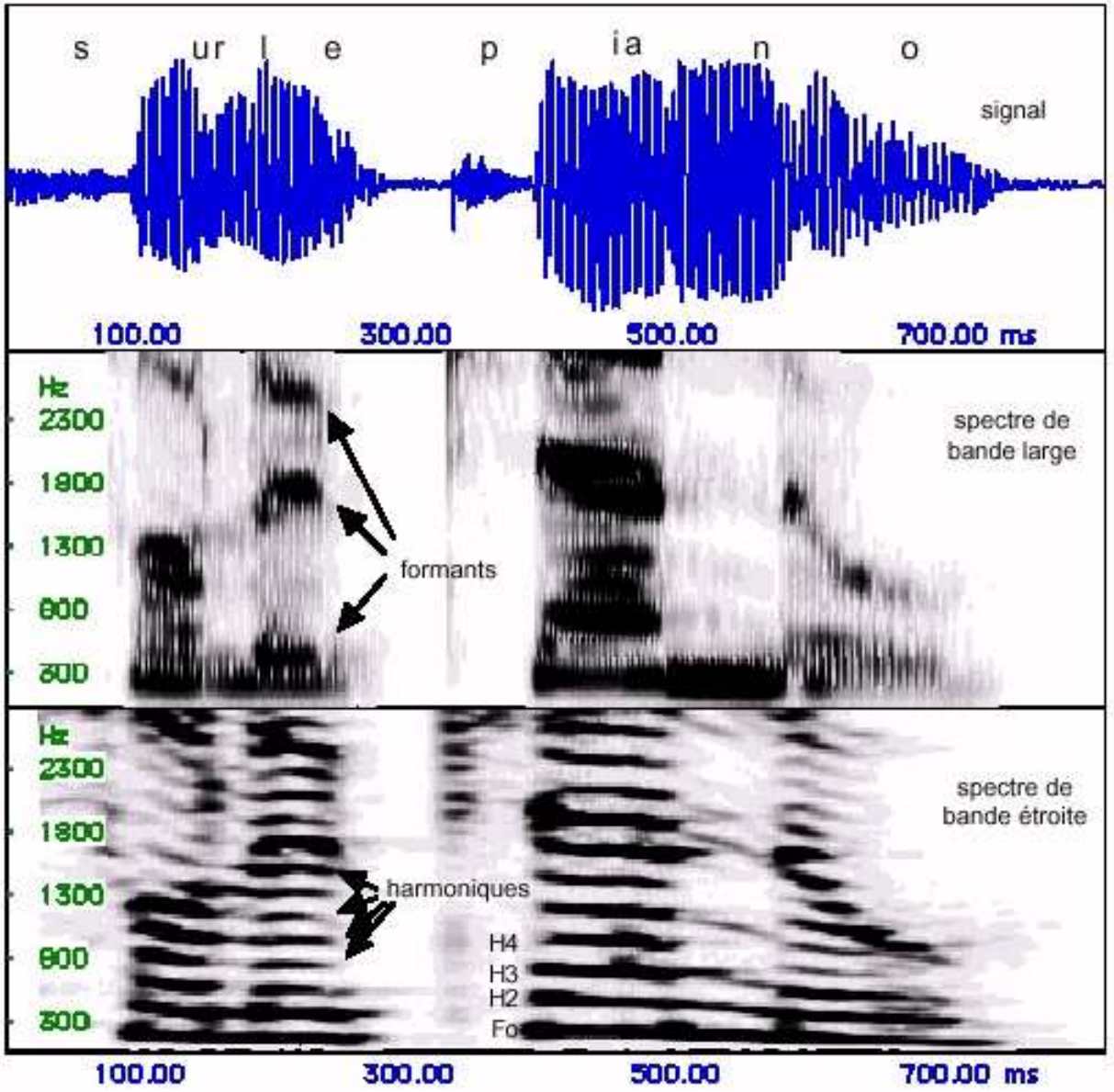


Structure formantique caractéristique de chaque son

Analyse du signal de parole

Énergie

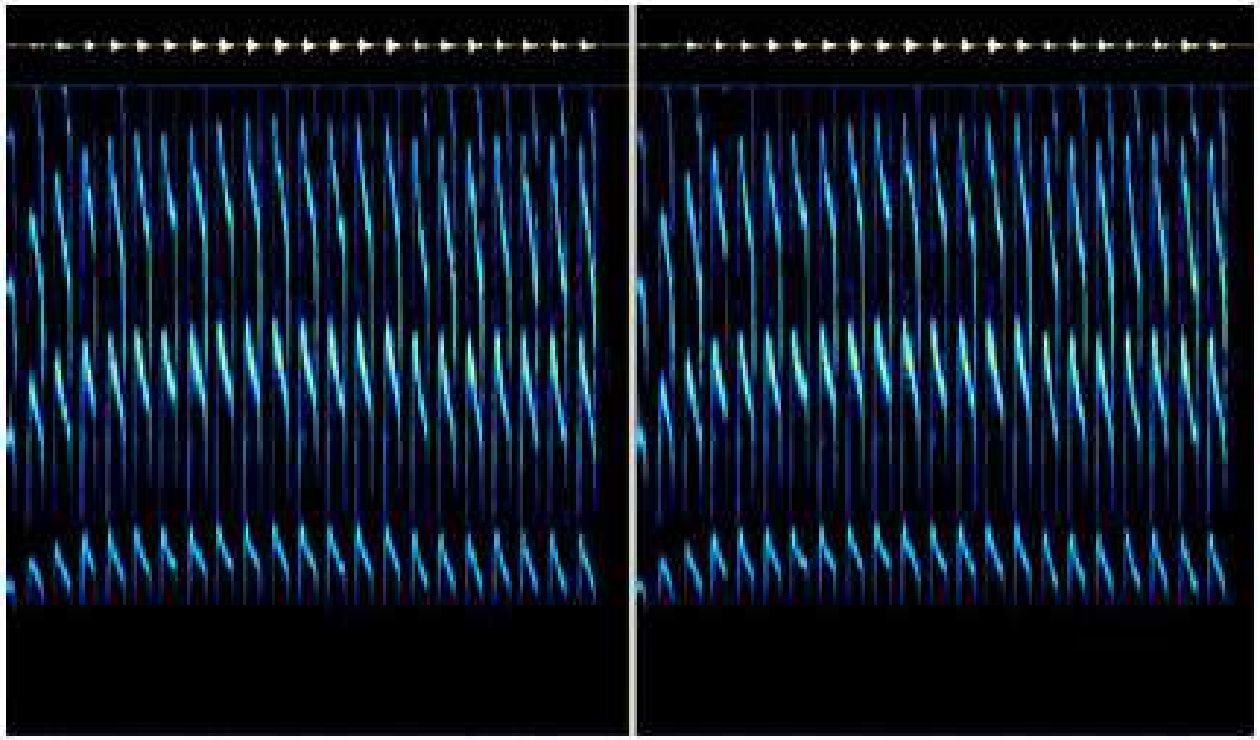
Spectrogramme



d'après [Keller 97]

Son non articulé

Chant du *Torcol fourmilier (Jynx Torquilla)*

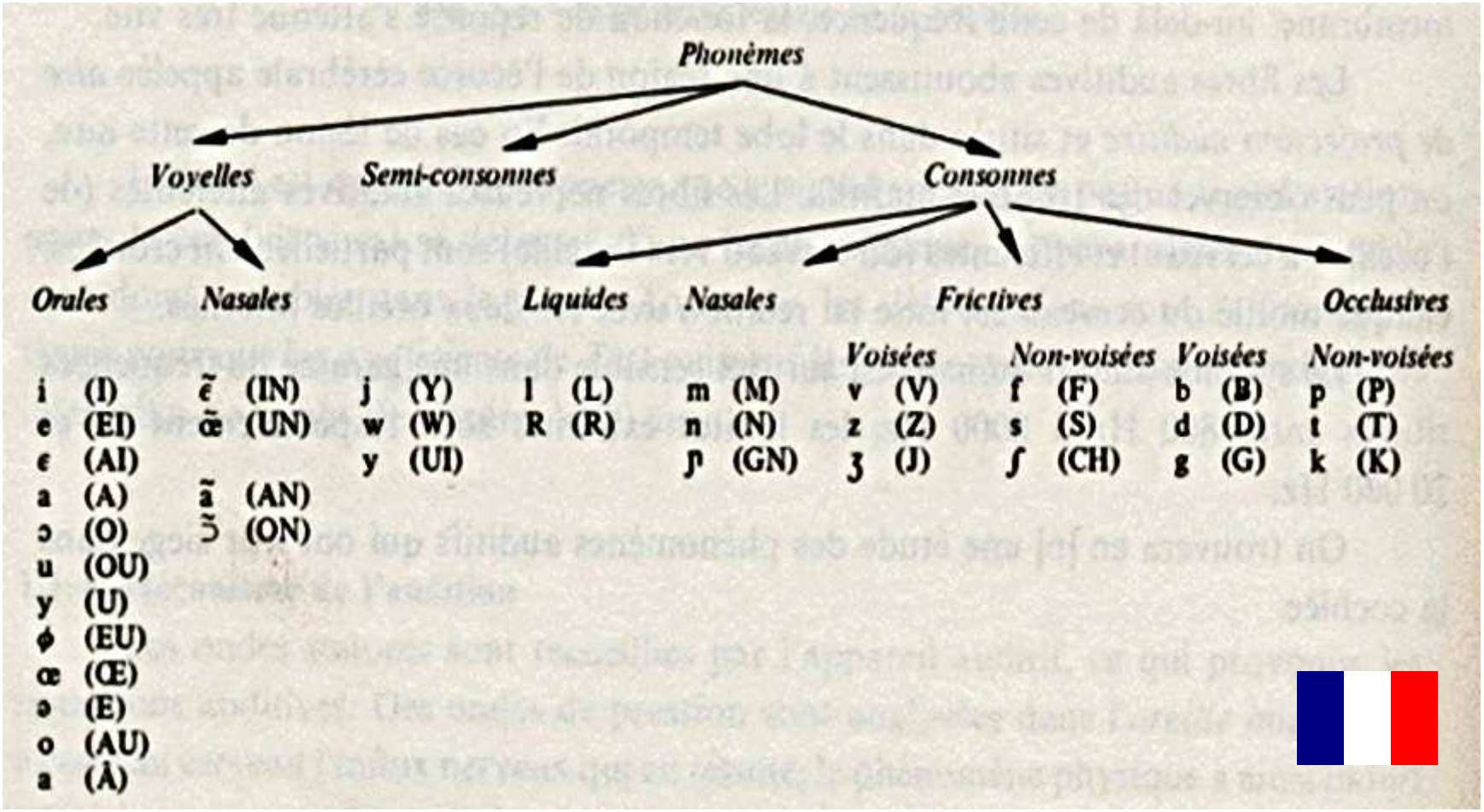


Parole articulée propre de l'homme

- Grands singes incapables de produire une parole articulée
- *Homo sapiens neanderthalis* pouvait articuler

Phonétique

Classification des sons de parole suivant leur mode de production



Phonétique

Spécificités idiomatiques

possibilités articulatoires communes où piochent les langues

- [x]

  *c'**h**oar , Cra**c'h***
  *na**ch**, do**ch***
  *ba**j**o, me**j**or, **J**orge*

- [ɦ]

  *di**lh**ad*
  *caba**ll**o*

- [ŋ]

  *so**ng***
  *kla**ng***
  *parki**ng***

- [ɐ̃]

  *d**e**n*

Phonétique : voyelles

Voyelles : voisement + conduit vocal ouvert (sons très énergétiques)

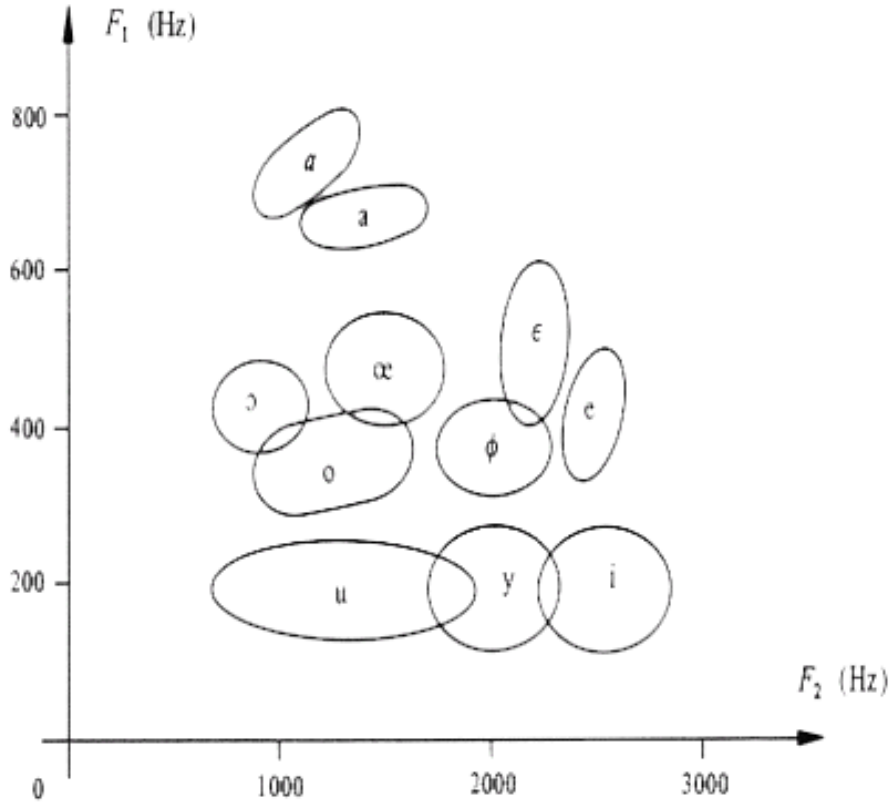
- **voyelles orales** français : [i e ɛ a ɑ ɔ o u y ∅ œ ɚ]
- **voyelles nasales** toute voyelle orale peut être nasalisée

  français : [ɛ̃ ɑ̃ ɔ̃ œ̃]

- **position de la langue**

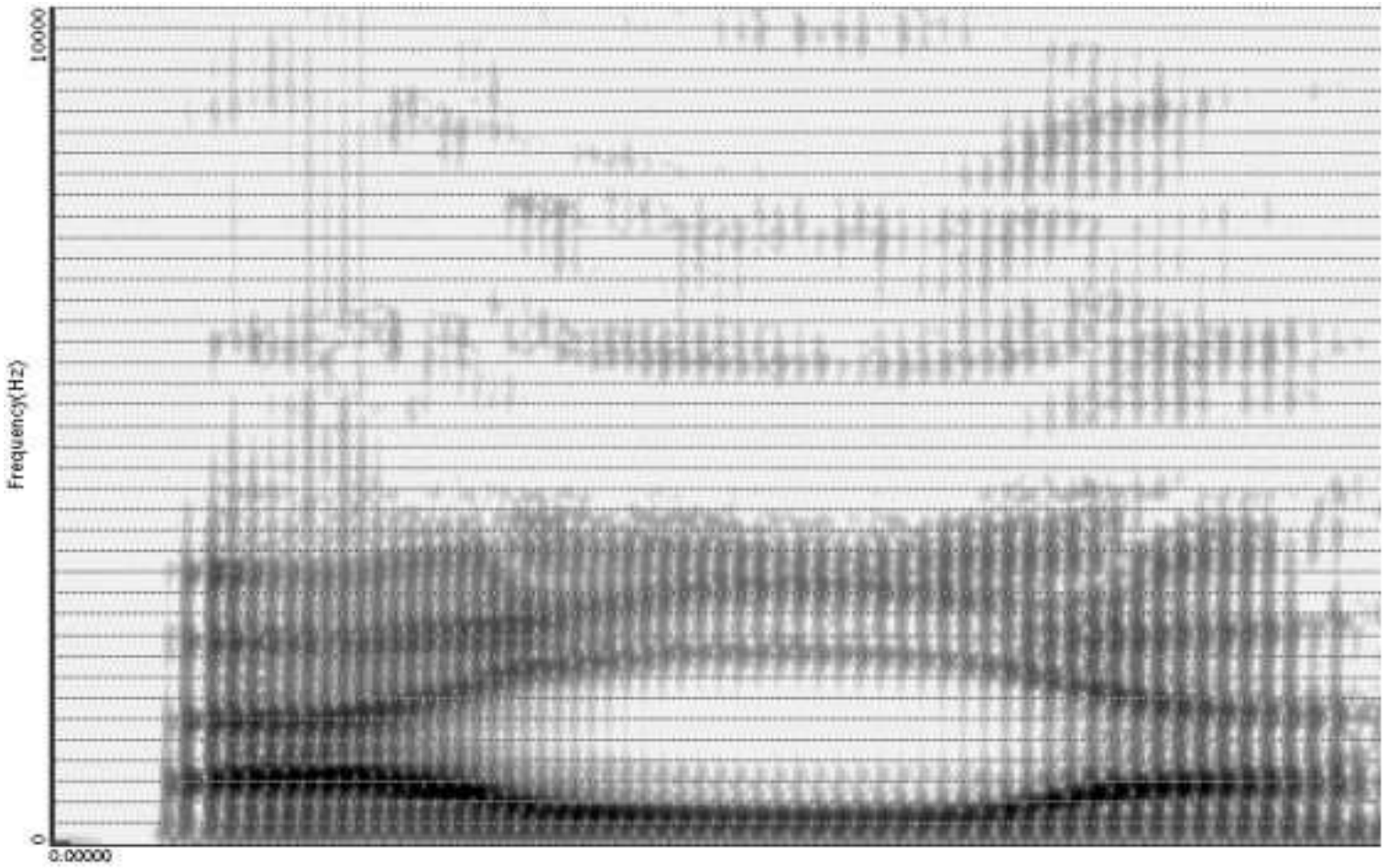
  antérieure / moyenne / postérieure
  aperture : fermée (haut) / ouverte (bas)

  ↳ **formants F1 et F2**



Phonétique : voyelles

Co-articulation



[a j a]

d'après [Keller 97]

Phonétique : consonnes

- **Constriction** lors de la production (son moins énergétique)
- **Occlusives** (ou plosives)
  - **son dynamique**: phase d'occlusion (tenue) suivi d 'un relâchement brutal
  - **voisées ou non voisées** : son basses fréquence émis pendant le tenue

    français : **[b d g]** plosives voisées, **[p t k]** non voisées
  - **lieu d'occlusion** palais [k g] , dents [t d] , lèvres [p b]
- **Fricatives**
  - **Turbulence** : forte constriction au lieu d'articulation ⇨ son non périodique
  - **Voisées ou non voisées** : excitation périodique + turbulence

    français : **[v z ʒ]** fricatives voisées, **[f s ʃ]** non voisées
  - **lieu d'occlusion** palais [ʃ ʒ] , dents [s z] , lèvres [f v]

Phonétique

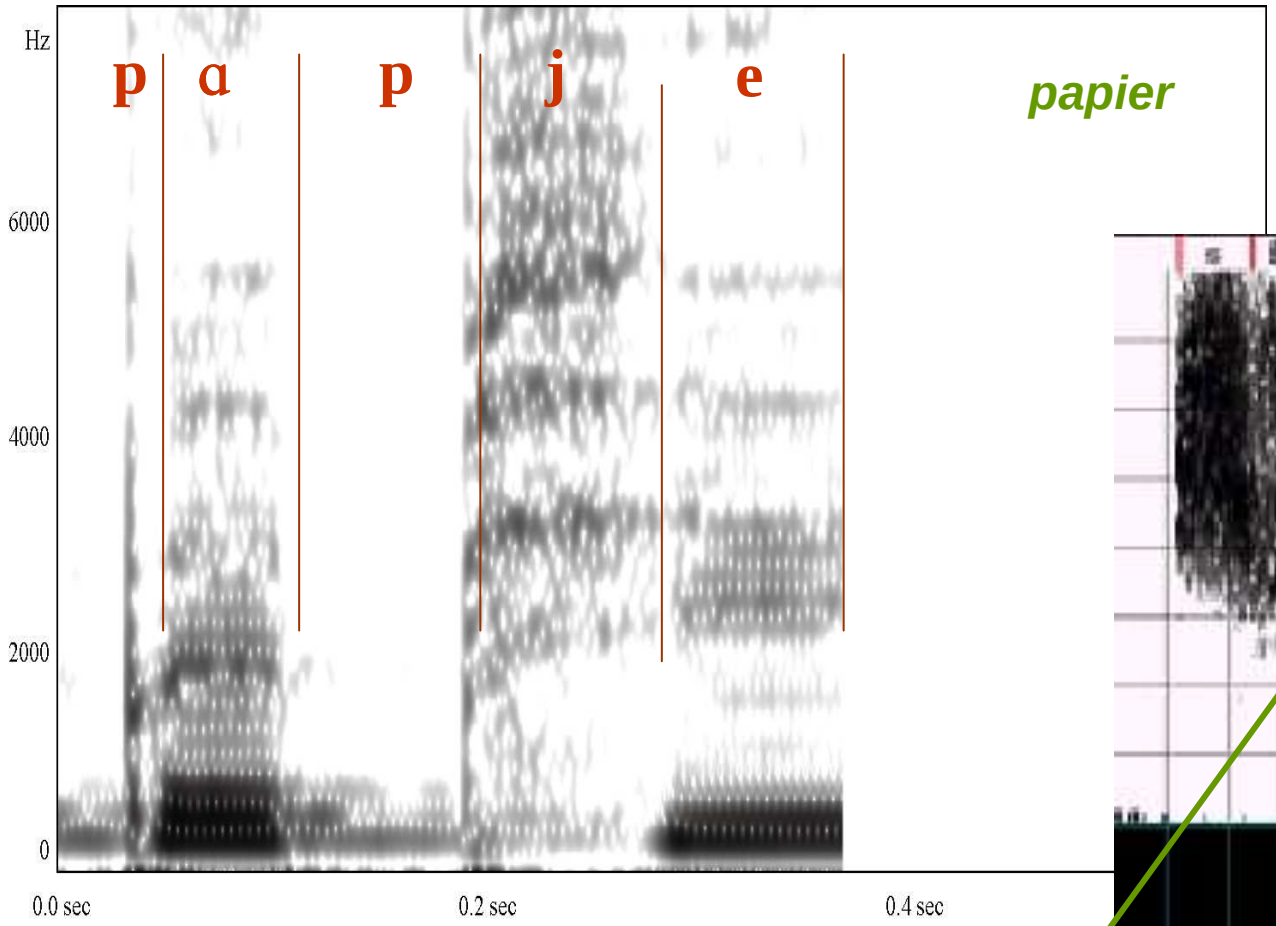
Autres classe phonétiques — **Voisement + constrictions non complète**

- **Semi-voyelles** ou … semi-consonnes (*glides*)
  français [j w ɥ]
- **Liquides** français [l ʀ]
- **Nasales** constriction assez forte mais dérivation nasale
  français [m n ɳ]

Lecture de spectrogramme

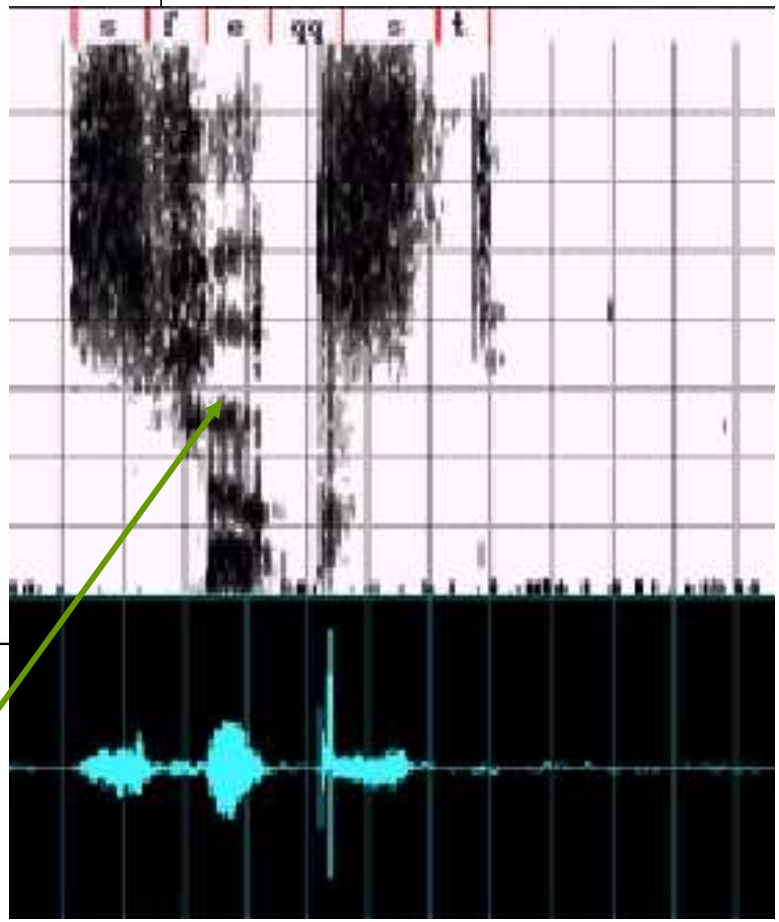
- **Occlusive** : tenue + barre de plosion
- **Fricative** : bruit (turbulence) haute fréquence)
- **Voyelle** : structure formantique énergétique
- **Semi-voyelle** : formants moins nets + bruit haute fréquence (constriction)

Lecture de spectrogramme



papier

sfqqst
(berbère)



schwa

d'après [Ridouane 02]

Prosodie

- **Rythme**
  - durée des phones et des silences
- **Mélodie**
  - variations de la fréquence fondamentale (*pitch*)

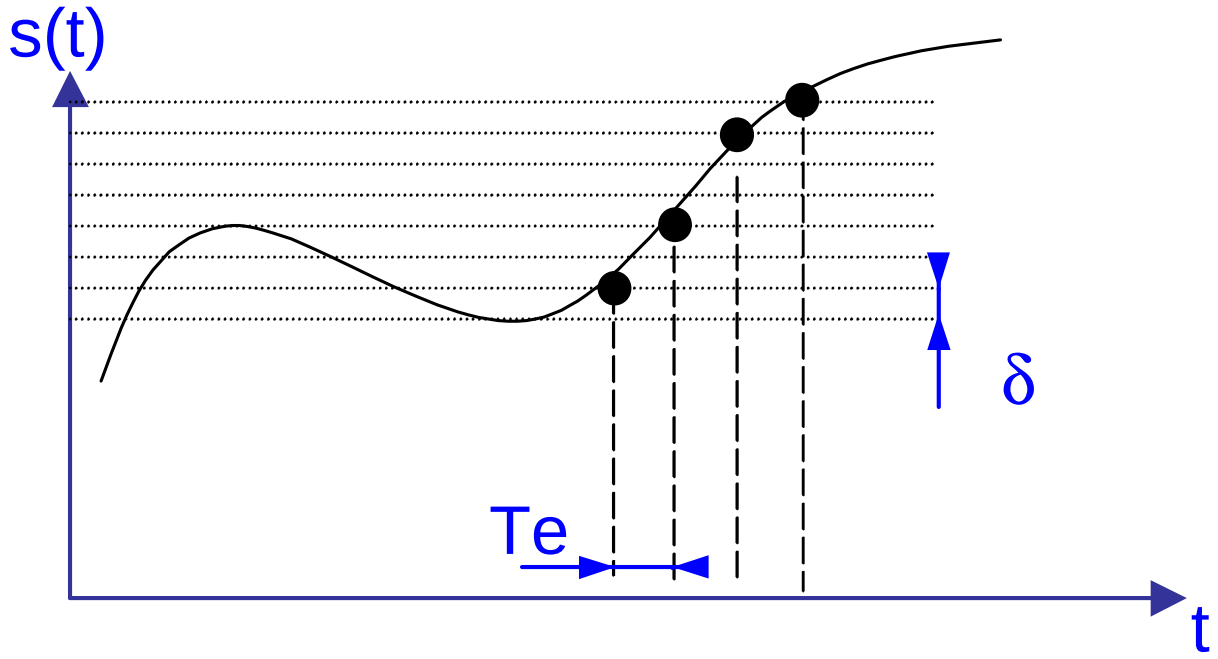
  exemples montée en fin de question intonative

  variations en fin de mot, de syntagmes
- **Intensité**
  - marques d ’accentuations (syllabes fortes ou faibles)
- **Quelle unité prosodique ?**

  Difficile à définir car plusieurs niveaux de significations (morpho-syntaxique, sémantique, pragmatique) : micro et macro-prosodie

Numérisation du son

- **Echantillonnage**
  - **fréquence d'échantillonnage** $f_e = 1 / Te$
  - **numérisation sur b bits** ⇨ 2^b niveaux de **pas de quantification** δ



- **Débit binaire**

  D = f_e x b **Exemple** : D = 8 Khz X 8 bits = 64 kb/s = 8 kO/s

Numérisation du son

- **Numérisation et détérioration du signal**
  - **théorème de Shannon**: l ’échantillonnage induit une perte d ’information pour toutes les composantes fréquentielles supérieures à la moitié de la fréquence d’échantillonnage (fréquence de coupure).
  - **bruit de quantification** dépendant de δ (rapport signal / bruit)

  ↳ **compromis entre limitation du débit binaire et minimisation de la détérioration du signal** (perte d’information)

Application	signal utile	f_e	b
	0-3400Hz	8 KHz	8 bits
	0-20 Khz	44.1 kHz	16 bits

Formats de fichiers sonores

Deux classes de formats

- **Formats bruts** (*raw*) — définissent une seule norme de codage : le fichier ne contient que le signal

  exemple .SND (1 canal sur 8 bits)

- **Formats auto-décrits** (*self describing*) — plusieurs normes : signal + entête décrivant ses caractéristiques (fe, type de codage, nombre de canaux…)

  exemples .AIFF, . WAV, .AU (compression optionnelle)
  .AIDC, .MP3 (compression)

Formats hybrides

- Multimédia : son + vidéo, son + MIDI

  exemple .MOV, .QT (Quicktime), .MPEG, .MP4

Compression

CODECs pilotes de COmpression / DECompression

- **PCM (Pulse Code Modulation)**
  - signal sur 8 ou 16 bits, f_e entre 8 et 44,1kHz
  - taux de compression 2,5:1
- **ADPCM / IMA (Adaptive Delta Pulse Code Modulation)**
  - non normalisé : Microsoft, Creative Labs, IMA
  - compression 16 bits vers 4 bits : taux de 4:1

  Exemple format Wave (.wav)

  format natif Microsoft Windows, jusqu'à 44,1 Khz en 16 bits

  compression optionnelle ADPCM/IMA
- **MP3**
  - norme MPEG (*Moving Picture Experts Group*)
  - excellent codage perceptif (psycho-acoustique) + suppression des signaux redondants (transformée en ondelettes) + extraction des fréquences peu audibles
  - taux de compression 8:1
  - **supporté par de nombreux éditeurs de signal**

Editeurs de signal de parole

- **SFSWin (Speech Filling System)**

  Gratuiciel (freeware)
  www.phon.ucl.ac.uk/resource/sfs

- **PRAAT**

  Gratuiciel (freeware) centré sur l'analyse phonétique
  Multiples modules + possibilité programmation (langage de script)
  www.praat.org

- **Winpitch**

  Gratuiciel (freeware) centré sur l'analyse mais aussi l'aide à l'enseignement de la prosodie [Martin 2005]
  http://www.winpitch.com/

- **GoldWave**

  Shareware (GoldWave Inc.) http://www.goldwave.com/
  Plus adapté au traitement de signal qu'à l'édition proprement dite.

Bibliographie

Ouvrages généraux

- **Huang X., Acero A., Hon H-W.** (2001) Spoken Language Processing : a guide to theory, algorithm and system development. Pretice Hall, Upper Saddle River, NJ. (chap. 2)

- **Boite R., Bourlard H., Dutoit T., Hancq J., Leich H.** (2000) Traitement de la parole. Coll. Electricité. Presses Polytechniques et Universitaires Romandes. Lausanne, Suisse (chap 1)

Travaux cités

- **Martin Ph.** (2005) WinPitch LTL, un logiciel multimédia d'enseignement de la prosodie, *ALSIC*, 8(2), pp. 95-108 http://alsic.revues.org/index332.html

Traitement Automatique des Langues

TRAITEMENT DE PAROLE

Compléments hors programme … mais néanmoins dignes d'intérêt !

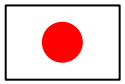
Phonétique : voyelles

- **Petite typologie des langues** [N. Vallée, ICP, Grenoble]
  - étude sur 317 langues
  - **nombre de voyelles**
    - 24 voyelles 1 langue
    - 2 voyelles 2 langues
    - 5 voyelles 23 %
    - 6 voyelles 13 %
  - **voyelles orales**
    - [i] 99% des langues
    - [a] 98 % des langues
    - [u] 94 % des langues
    - [o] 44 % des langues
    - [e] 40 % des langues
  - **voyelles nasales** : langues (22%) utilisant plus de 9 voyelles
    - [ã] 20 % des langues
  - **français** variable suivant la région: *brun* vs. *brin*

Phonétique

Autres mécanismes spécifiques à certains idiomes

- **Durée du phonème** (20 % des langues)



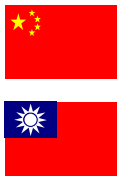
kado	*coin*
kaado	*carte*

- **Consonne roulée**

pero	*mais*
perro	*chien*

↬ Distinction porteuse de sens en espagnol, contrairement au roulement du [ʀ] dans le français provençal.

- **Variation du pitch** (langues tonales : Asie, Afrique)



ma	ton élevé	*maman*
	forte montée	*imbécile*
	faible montée	*cheval*
	forte descente	*réprimander*

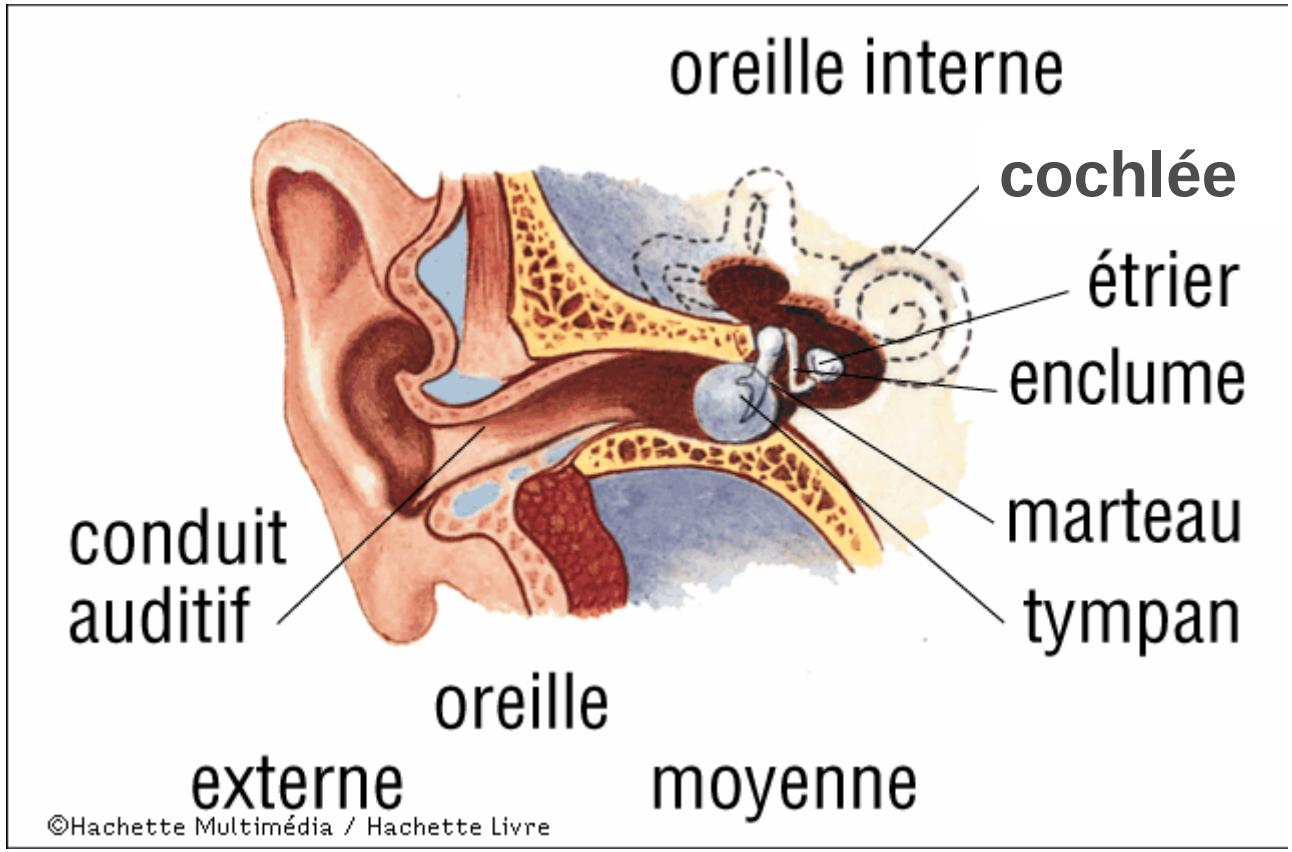
Phonologie : phonèmes

- **Interface entre phonétique et linguistique**
- **Phonèmes**
  - plus petite unité phonique fonctionnelle *i.e.* distinctive d 'un point de vue sémantique
  - un phonème = plusieurs réalisations articulatoires (**allophones**)
  - notation API entre barres inclinées / f ɔ n ɛ m /
- **Allophones**
  - roulement du [ ] dans le français provençal
  - phénomènes de **co-articulation**, assimilation, réduction

médecin

il y a /i l i j a/ ⇨ [i l j a] [j a]

Perception


oreille interne
cochlée
étrier
enclume
marteau
tympan
conduit
auditif
oreille
externe
moyenne
©Hachette Multimédia / Hachette Livre

Perception

- **Psycho-acoustique : signal perçu *vs*. signal physique**
  - **intensité** : variable suivant le fréquence (sensibilité maximale sur la plage de fréquence [500Hz - 10 kHz])
  - **hauteur** : *pitch* relativement lié à la fréquence fondamentale
  - **timbre** : relation non linéaire entre fréquence physique et fréquence perçue, due à la répartition des fréquences de résonance des cellules ciliées de la cochlée.

    Échelle de Bark $b(f) = 13. \text{Arctan}(0{,}00076\, f) + 3{,}5 * \arctan((f/7500)^2)$

    Échelle de Mel $m(f) = 1125 \ln(1 + f/700)$

- **Approche perceptive**
  - synthèse de la parole
  - télécommunications : réduction du débit binaire non perceptible (mp3)