

# Traitement Automatique des Langues

**Jean-Yves Antoine**

LI - Université Rabelais de Tours

Jean-Yves.Antoine@univ-tours.fr

# Traitement Automatique des Langues

## SYNTHESE DE LA PAROLE

# Rappels : analyse signal de parole

- **cordes vocales** : son voisé / non voisé
- **cavité nasale** : son nasal / non nasal
- **langue**
  - haut (fermé) / bas (ouvert)
  - avant / arrière
- **lieu d'articulation**
  - lèvres
  - dents
  - palais



**Fréquence fondamentale**  
 $f_0$  (*pitch*)

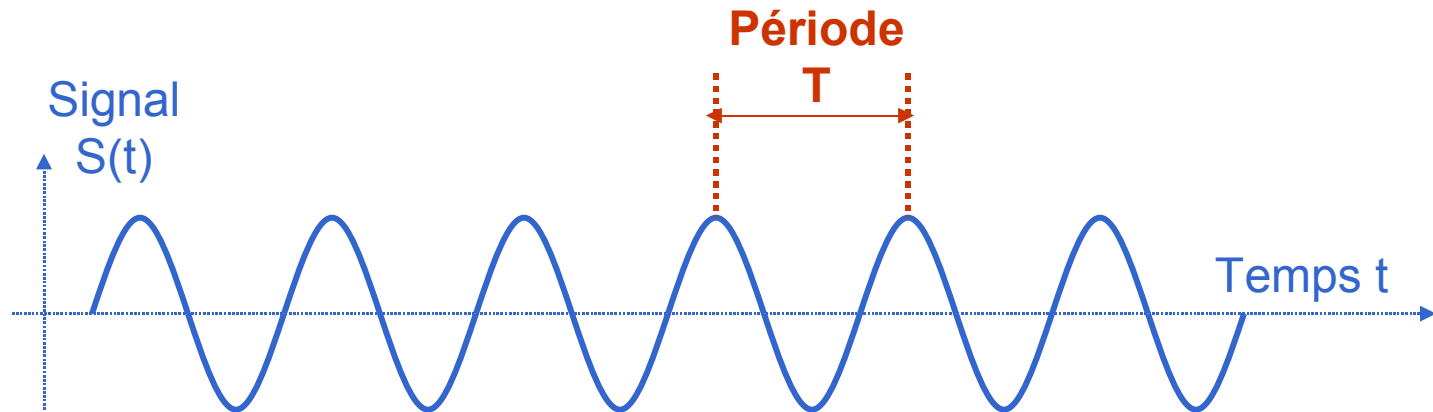


**Formants**  
Fréquences de résonance  
( $F_1, F_2, F_3 \dots$ )

# Rappels: analyse du signal

## Son périodique « pur » (sinusoïdal)

- Reproduction du même signal suivant un période **T** constante

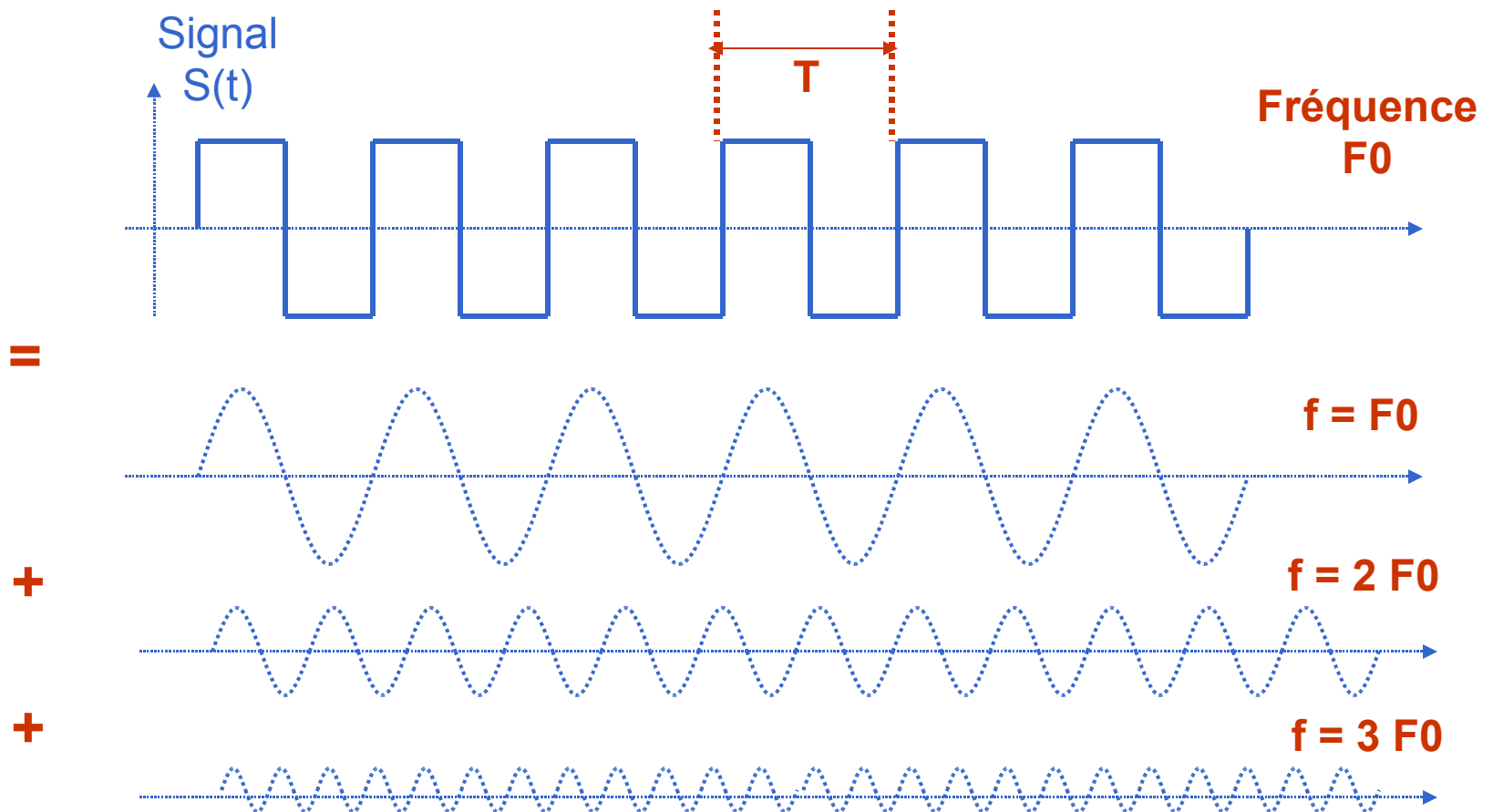


- Une seule composante fréquentielle : **fréquence fondamentale  $F_0$**
- **Fréquence** : nombre de périodes par seconde (Hertz)  $F = 1/T$

# Rappels: analyse du signal

## Son périodique quelconque

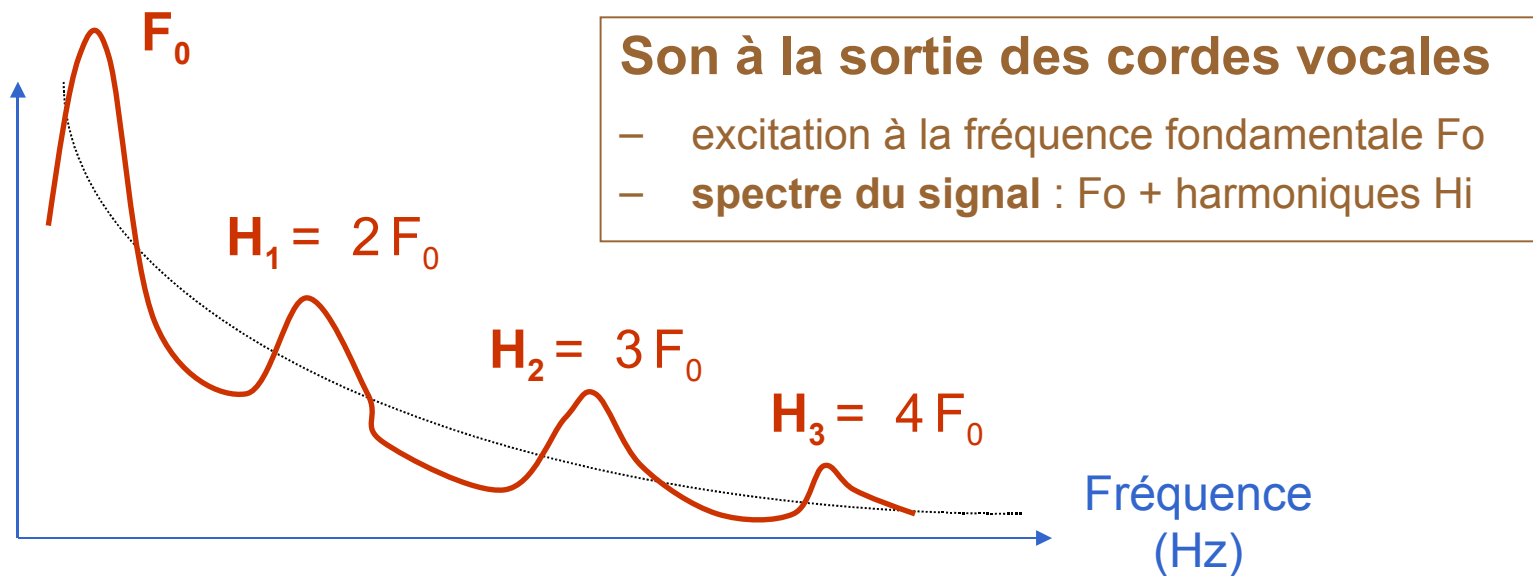
Correspond à la somme de signaux sinusoïdaux de fréquences multiples de la fréquence fondamentale et d'amplitude décroissante (séries de **Fourier**)



# Rappels: analyse du signal de parole

## Son périodique quelconque

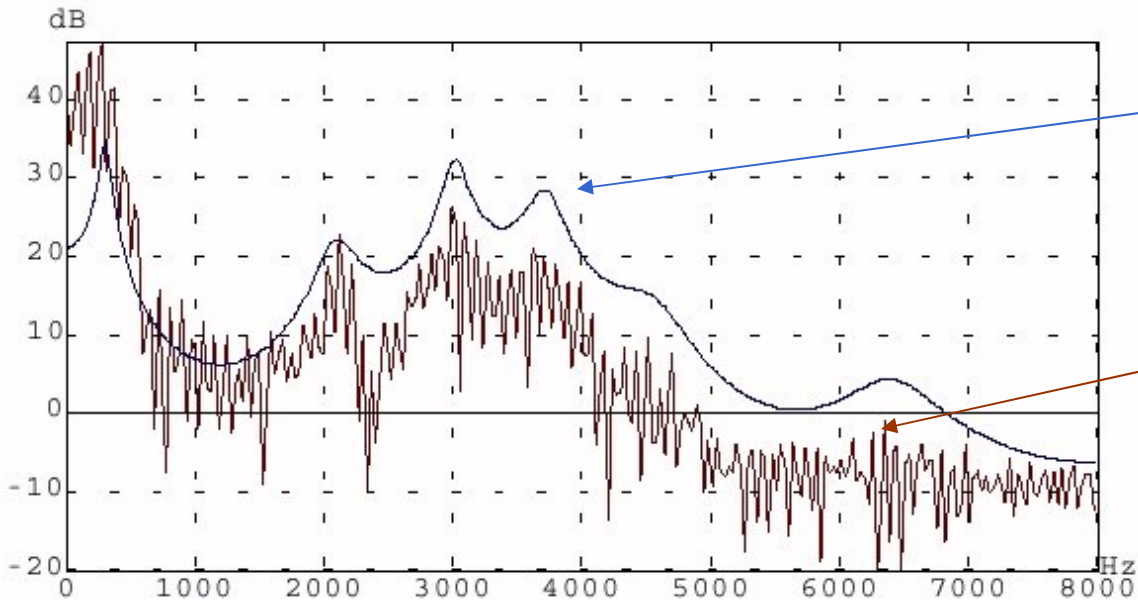
- Plusieurs composantes fréquentielles correspondant à la fréquence de chaque sinusoïde de l'analyse de Fourier
- Fréquences multiples de la fondamentale : **harmoniques**
- **Spectre** d'un signal à un instant donné : représentation de l'importance des différentes composantes fréquentielles



# Rappels: analyse du signal de parole

- **Articulation**

- lieux de constriction : cavités de résonance
- fréquences de résonance dues aux articulateurs : **formants**
- **spectre du signal** :  $F_0$  + harmoniques + formants  $F_i$

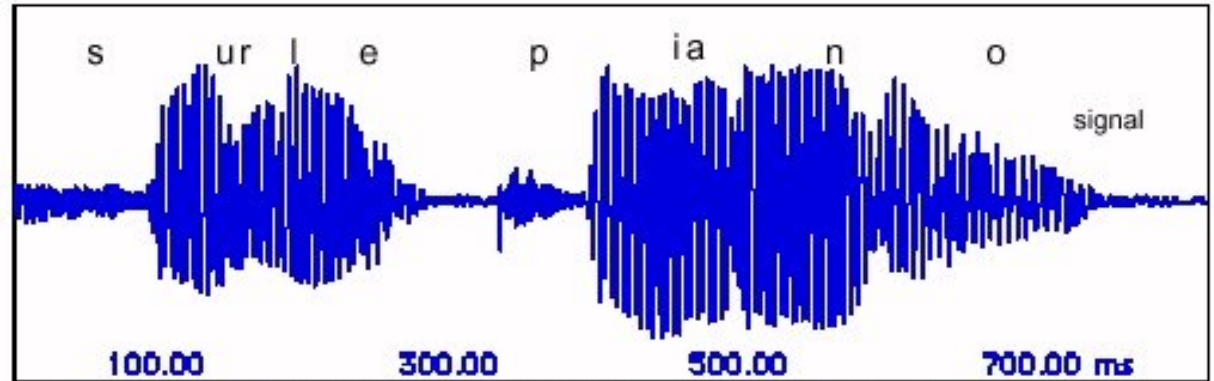


enveloppe  
spectrale :  
formants

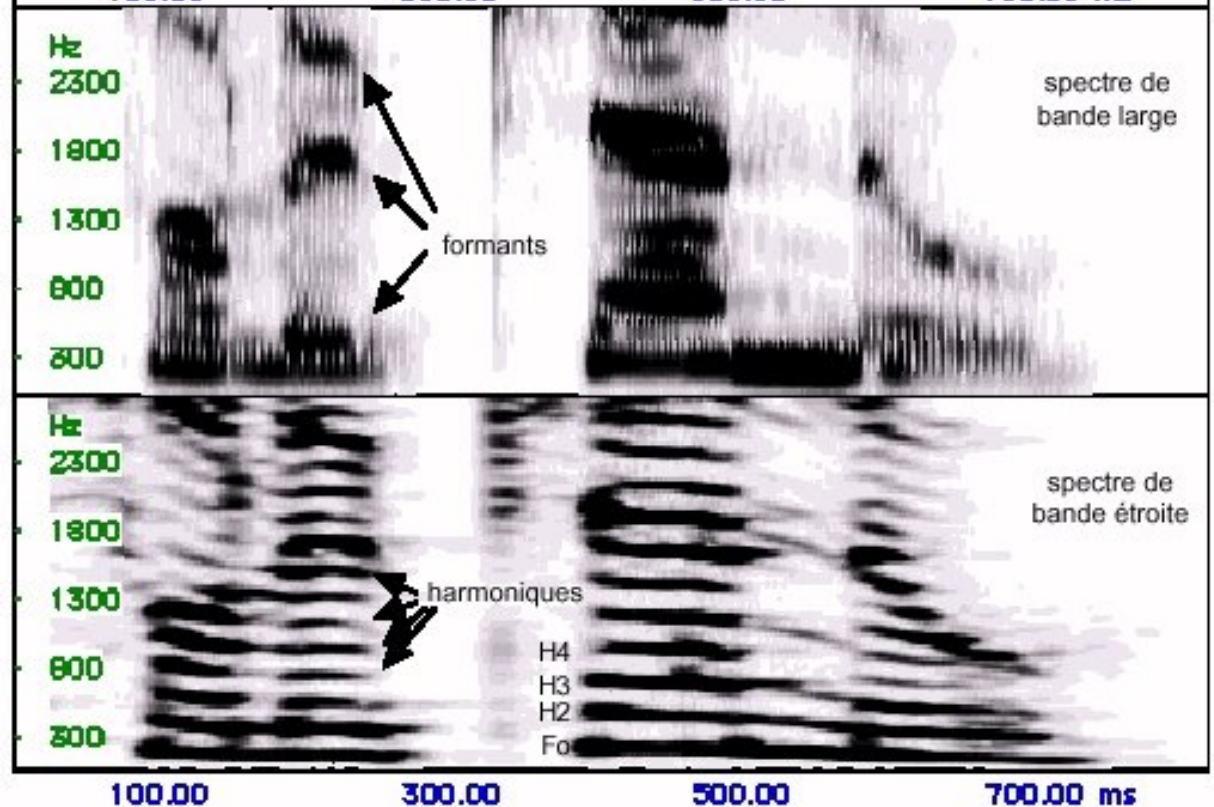
harmoniques

# Rappels: analyse du signal de parole

Énergie



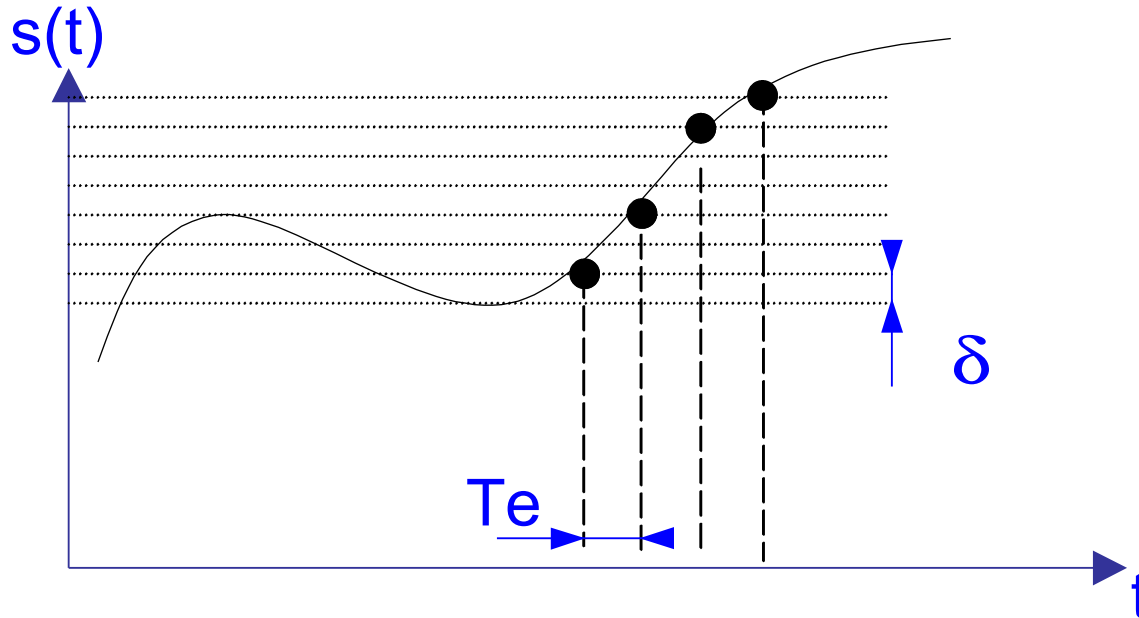
Spectrogramme



d'après [Keller 97]



# Rappels : son numérique



- **Fréquence d'échantillonnage**  $\Leftrightarrow f_e = 1 / T_e$
- **Numérisation sur b bits**  $\Leftrightarrow 2^b$  niveaux de **pas de quantification**  $\delta$
- **Théorème de Shannon:** l'échantillonnage n'induit aucune perte d'information si  $f_e$  est supérieure ou égale à 2 fois les plus hautes composantes fréquentielles du signal.

# Synthèse de la parole: techniques

- **Synthèse à partir du texte vs. synthèse par concaténation**

- ✓ TTS : *Text-To-Speech synthesis*

- **TTS: approches centrées connaissances**

- ✓ Premières approches historiques

- ✓ **Principe** : règles phonétiques générant le spectrogramme du signal

- ✓ Modèles explicatifs, mais qualité limitée



- **TTS: approches centrées données**

- ✓ **Principe**: ajustement (durée, intonation...) puis concaténation de signaux de parole déjà enregistrés.

- ✓ **Concaténation de diphones** – bonne intelligibilité (coarticulation) mais intonation perfectible PSOLA (CNET), MBROLA [Dutoit, Leich, 1993]

- ✓ **Sélection d'unités** – sélection d'unités de longueurs variables dans des bases de sons de plusieurs GOctet. Naturalité presque parfaite

[Hunt, Black, 1996]



# Synthèse par formants (source / filtre)

- **Synthèse par formants**

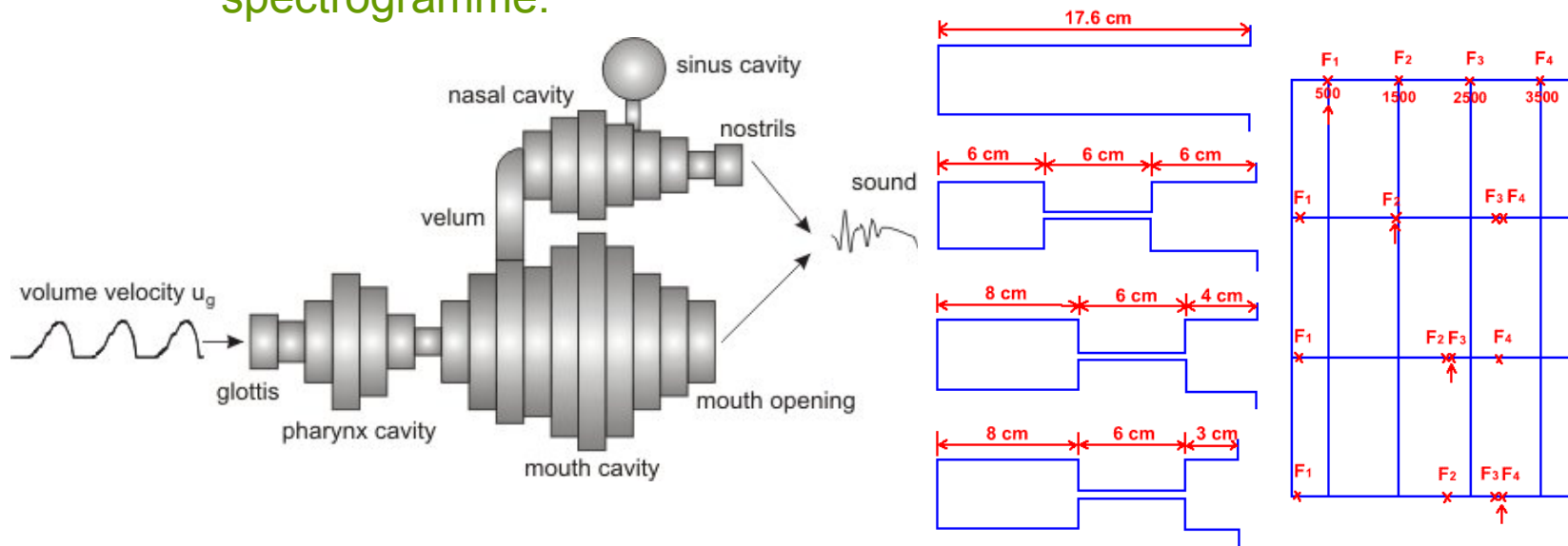
[Fant, 1960]

- ✓ Signal de parole = **source filtré** par les cavités du conduit vocal
- ✓ **Source** – voisement des cordes vocales (pitch  $F_0$ ) ou bruit chaotique dû à la constriction dans les fricatives + radiation au niveau des lèvres
- ✓ **Filtre** – fréquences de résonances (formants)

- **Application à la synthèse vocale**

[Klatt, 1990]

- ✓ Règles acoustiques décrivant le passage source + conduit => spectrogramme.



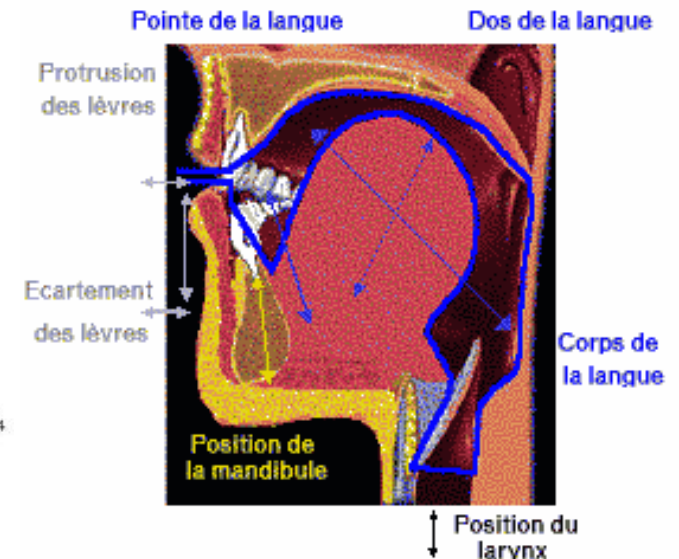
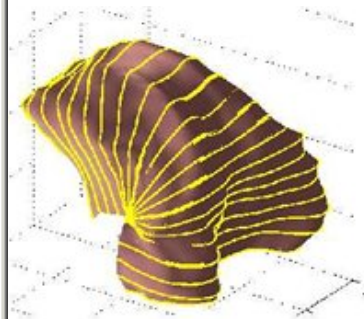
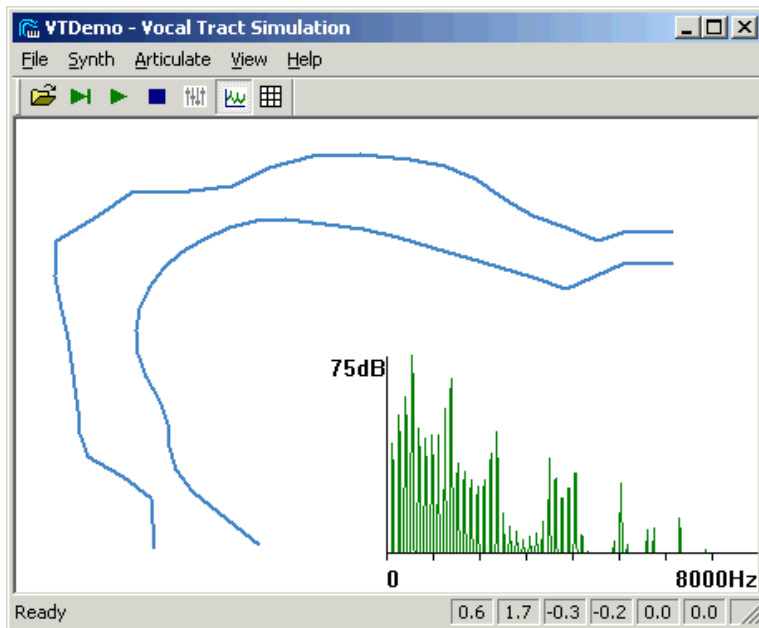
# Synthèse articulatoire

- **Modèle**

- ✓ Modèle physico-mathématique décrivant le signal généré par un appareil phonatoire décrit par certains paramètres

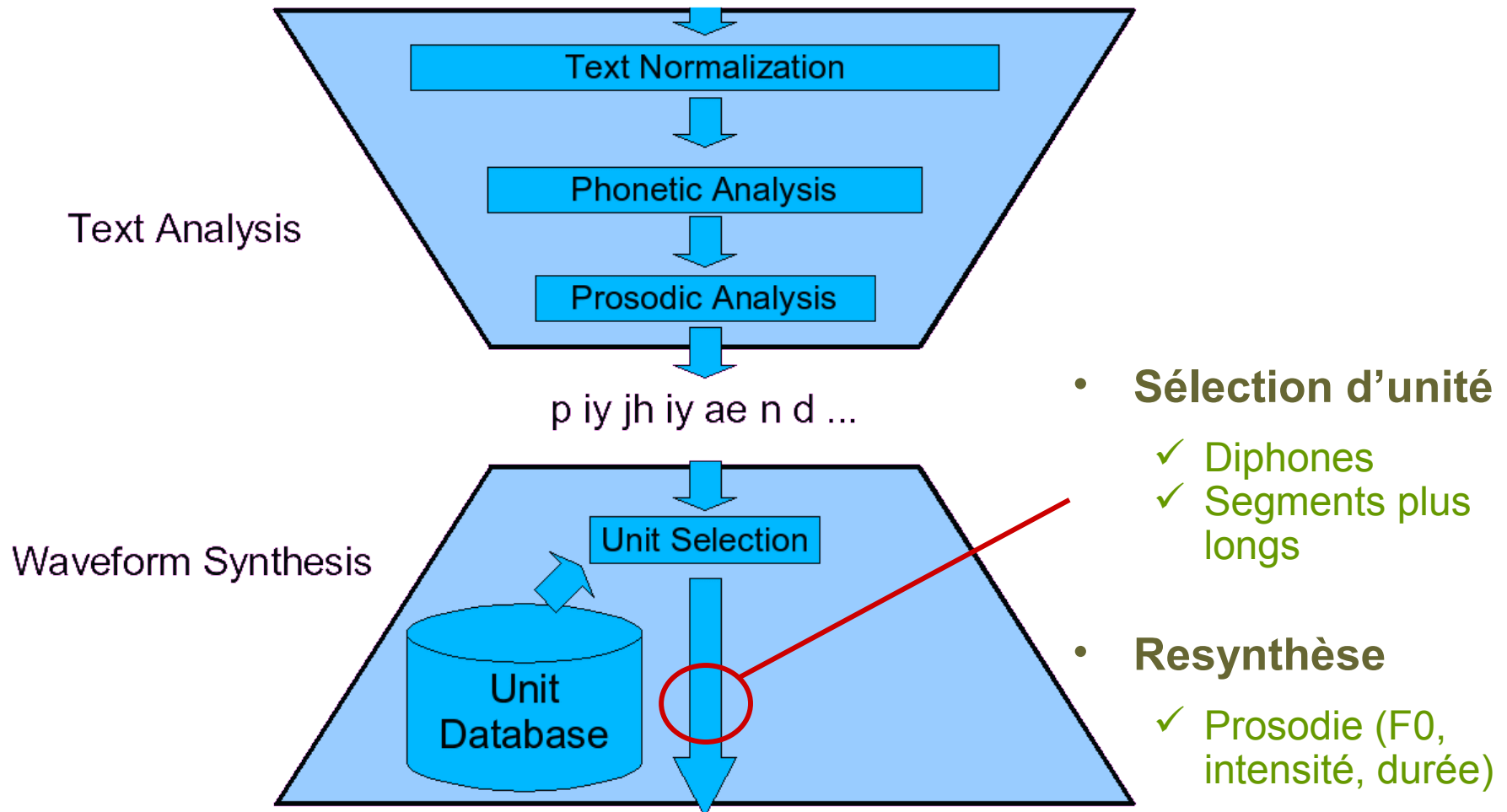
- **Application à la TTS**

- ✓ Synthèse articulatoire intégrée dans PRAAT, synthèse de l'ICP...



# Synthèse par sélection d'unités

- **Principe**



# Bibliographie

---

- **Dutoit T., Leich H.** (1993) MBR-PSOLA: Text-To-Speech Synthesis based on a MBE Re-Synthesis of the Segments Database. *Speech Communication*. 13 (3(4)). 435-440.
- **Fant G.** (1960) Acoustic Theory of Speech Production. Mouton, The Hague, The Netherlands.
- **Hunt A., Black A.** (1996) Unit selection in a concatenative speech synthesis system using a large speech database. *Proc. ICASSP'96*, Atlanta. Georgia. 373-376.
- **Klatt D.H. & Klatt L.C.** (1990) Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America*, 87, 820–856.