ULB UNIVERSITÉ LIBRE DE BRUXELLES, UNIVERSITÉ D'EUROPE

La Matinée des Chercheurs

ACADÉMIE UNIVERSITAIRE WALLONIE-BRUXELLES

The source-tract separation of speech

Thomas Dubuisson, Thierry Dutoit

TCTS Lab, Faculté Polytechnique de Mons, Belgium

The Zeros of the Z-Transform Representation (*ZZT*) of speech

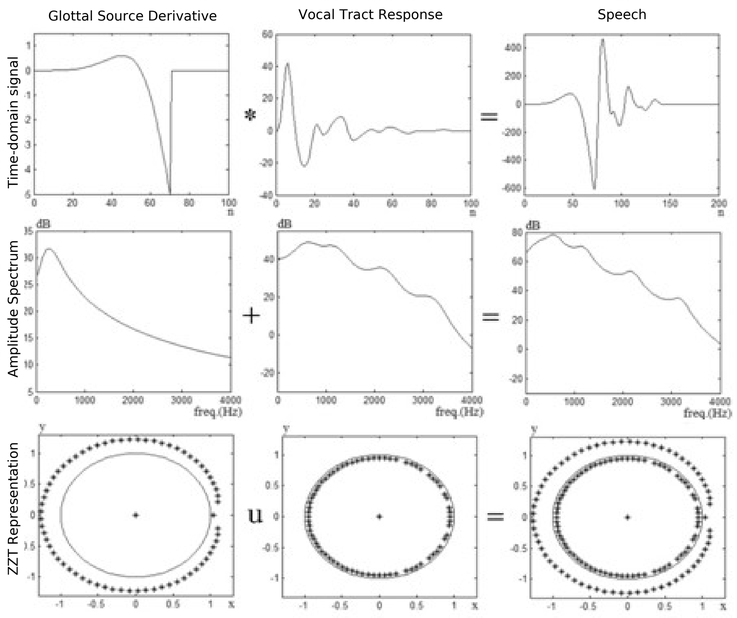
Definition

Numerical sequence x(n)

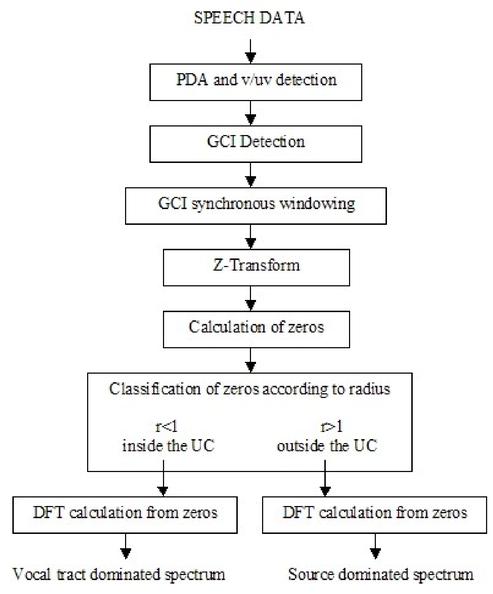
$$X(z) = \sum_{n=1}^{N-1} x(n) z^{-n} = x(0) z^{-N+1} \prod_{m=1}^{N-1} (z - Z_m)$$

$$ZZT = \{Z_1, Z_2, Z_3, ..., Z_{N-1}\}$$

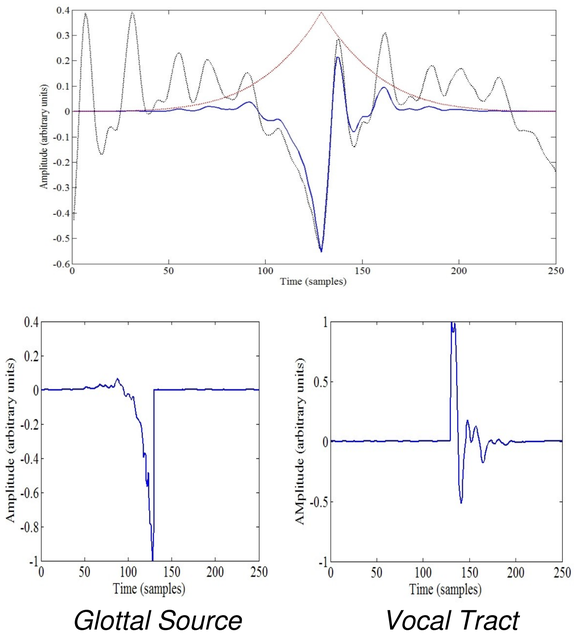
Mixed-phase model of speech



Decomposition algorithm



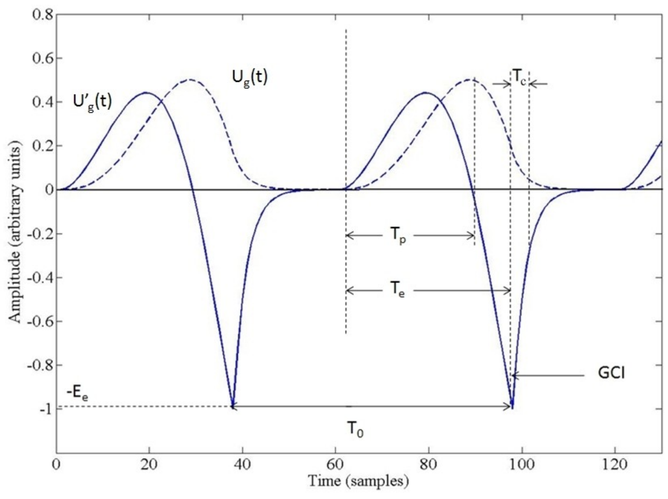
Example



Glottal Source *Vocal Tract*

Position of the problem

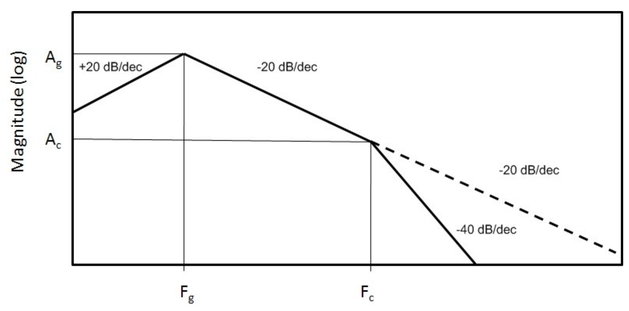
Glottal Source Model (Lijlencrants - Fant)



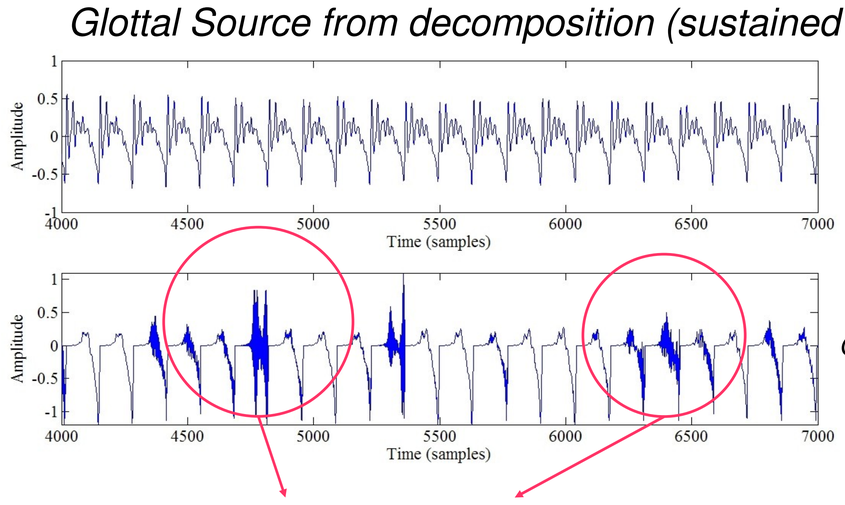
GCI = Glottal Closure Instant

$$\text{Open Quotient} = O_q = \frac{T_e}{T_0}$$

$$\text{Asymmetry Coefficient} = \text{Alpha}_m = \frac{T_p}{T_e}$$



Glottal Source from decomposition (sustained vowel /a/)



Speech

Glottal Source

Problems in the estimation of glottal source parameters like Open Quotient (Oq) or asymmetry coefficient (Alpham) !!

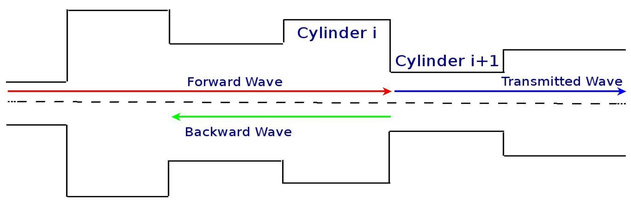
Improvement of the decomposition

Principles of the method

Observation: better glottal sources can be obtained by shifting the speech frame around each *GCI* and computing the *ZZT*-based decomposition of this frame

Idea: for each shift around each *GCI*

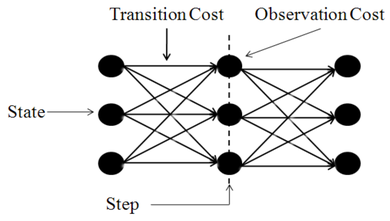
- Computation of the glottal source and the vocal tract
- Characterization of the glottal source by *FeatureGS* (defined as the energy ratio between the [0-2000] Hz band and the whole frequency band in the glottal source spectrum)
- Characterization of the vocal tract by *FeatureVT* (defined as the vector of radiuses from tube model)



Hypothesis: During the production of a sustained vowel, vocal tract geometry must remain as continuous as possible

Dynamic programming algorithm:

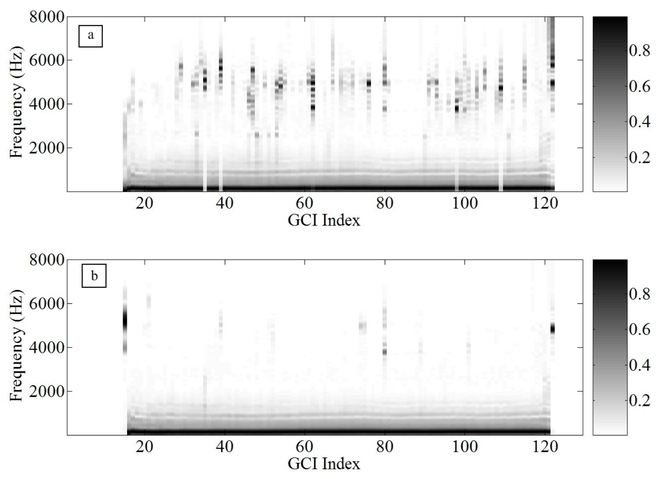
- Each « Step » corresponds to a given *GCI*
- Each « State » corresponds to a shift around a given *GCI*



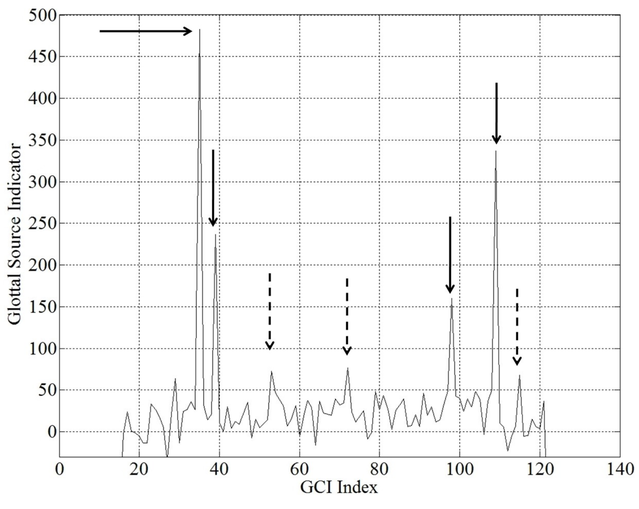
$$\text{Cost}(i, j) = \text{Cost}(i - 1, k) + \text{TransitionCost}_{kj}^{i-1/i} + \text{ObservationCost}(i, j)$$

Goal: Minimization of the Cost function along all the *GCIs detected* in the speech signal (sustained vowel)

Results on Glottal Source

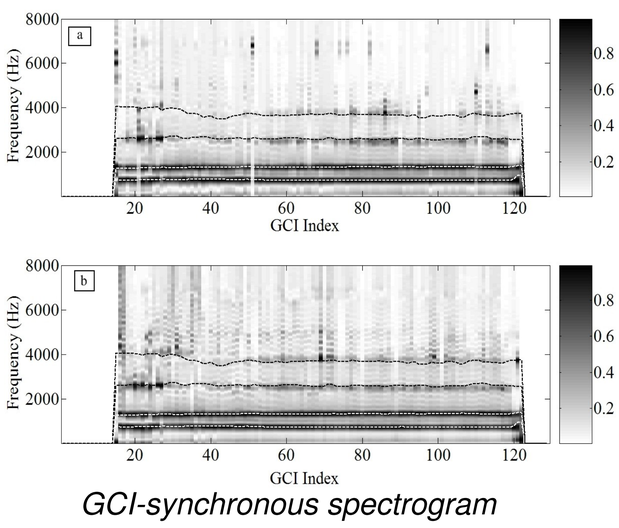


GCI-synchronous spectrogram

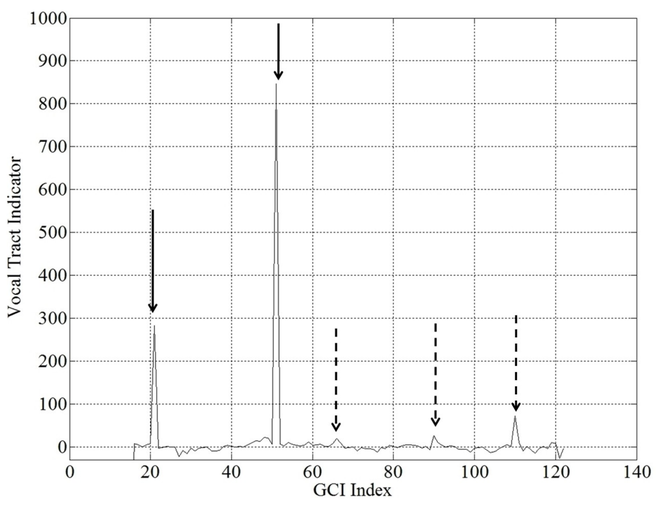


Glottal Source Indicator

Results on Vocal Tract



GCI-synchronous spectrogram



Vocal Tract Indicator

Work funded by the Walloon Region (ECLIPSE Project – Grant WALEO 2 516009)