Speech analysis

Yves Laprie

Lorraine University

Overview

- Spectral analysis
  - ❶ Spectrogram (Fourier transform)
  - ❷ Other spectral analyses (LPC, cepstra)
  - ❸ Spectral description of speech sounds
- ❹ Determining the fundamental frequency
- ❺ Modifying the fundamental frequency – PSOLA
- ❻ Using resonators to synthesize speech: formant synthesis
- ❼ Qualitative acoustics of the vocal tract

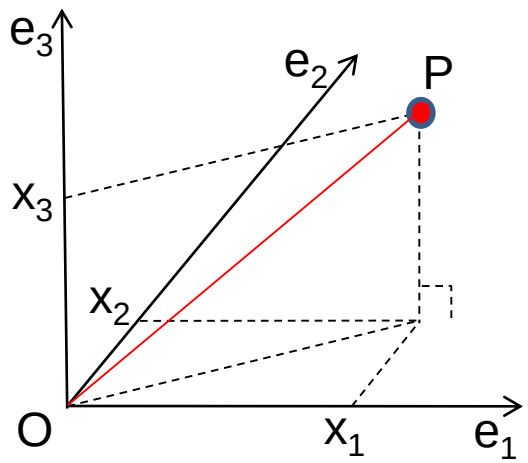
Loria

Spectral analysis

- Objective:
  - Displaying the energy distribution of speech along time and frequency
  - Studying the acoustic properties of the speech sounds

❶ Spectrogram

- The mathematic tool to study the distribution of energy along frequency is the Fourier transform.
- Analogy with a geometrical system coordinate:
  - The coordinates of P are given by the inner products of the base vectors with the OP vector.



Here, the point P is represented by the three coordinates x_1, x_2 and x_3.

  - Using an appropriate base to decompose a signal.
    This base is a base of functions since the object is not a scalar.

Spectrogram

- Discrete Fourier Transform

$$X(k) = \sum_{n=0}^{N-1} s(n) e^{-j\frac{2\pi}{N}kn}$$ time domain → frequency domain

s(n) is the speech signal, $b_k(n) = e^{-j\frac{2\pi}{N}kn}$ is the kth base function, and the dot product the inner product.
X(k) is thus the kth coordinate of the speech signal.

- Remarks:
  - this is a discrete definition (the speech signal has been sampled beforehand)
  - *N* has to be chosen relevantly and this choice amounts to “cut” abruptly a speech signal
  - is the base appropriate? The signal should be periodic, a window is thus applied before.

Spectrogram

- The signal is windowed before applying the Fourier transform

$$F_w(\omega) = \sum_{n=-N/2}^{N/2-1} w(n)s(n)e^{-j\omega n}$$

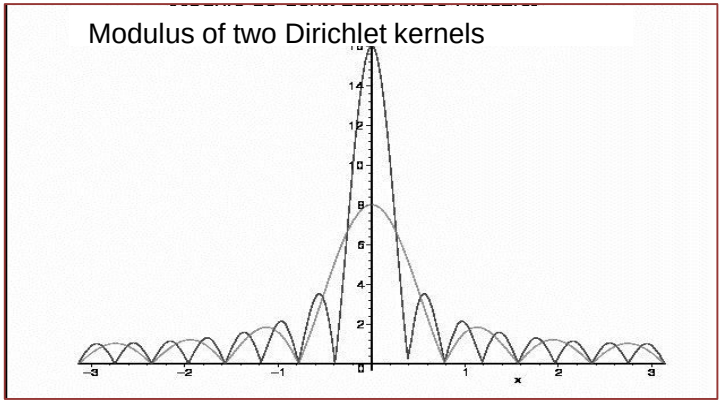
- where ω is the frequency and w the window $w(n) = 0$ when $|n| > N/2$. w is usually an even window.
- Impact of the window multiplication
  - A convolution in the frequency domain:

$$x(n) * w(n) = \sum_{k=-\infty}^{\infty} x(n-k)w(k)$$

  - From a practical point of view: since $F_w(\omega)$ is observed instead of $F(\omega)$ windows as neutral as possible are used.

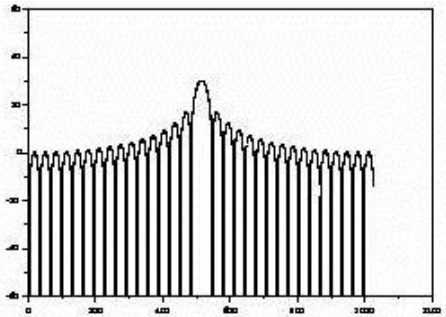
Windows used to compute spectrograms

- Two Dirichlet kernels (the effect of the rectangular window) one with 16 points and the second with 8 points.

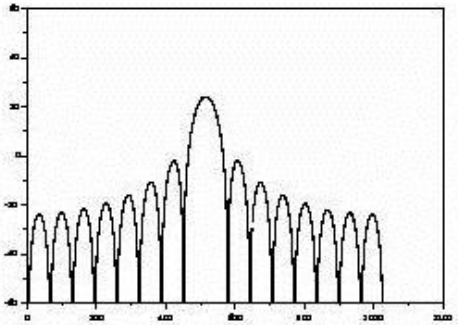


1. The higher the main peak the smaller the effect of convolution.
2. The sharper the main peak the smaller the effect of convolution.

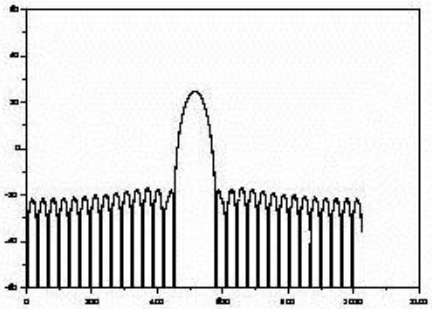
- Some classical windows (1024 samples in these examples, spectrum in dB):



Rectangular



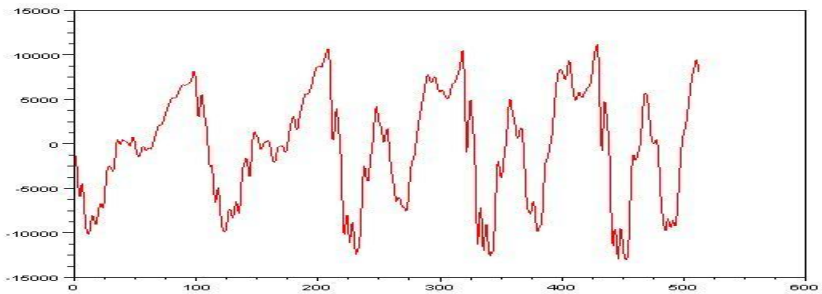
Triangular



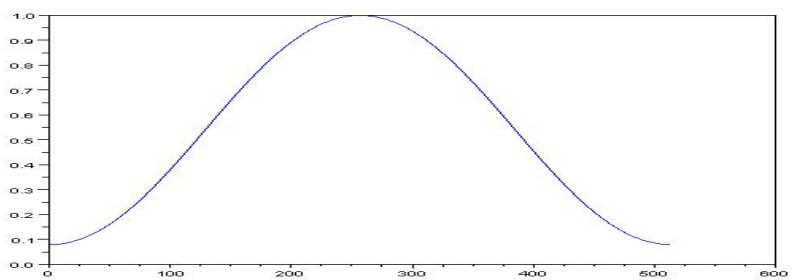
Hamming

- The longer the window the sharper the first peak (or equivalently,the smaller the effect of convolution)

The Hamming window



- The original speech signal.

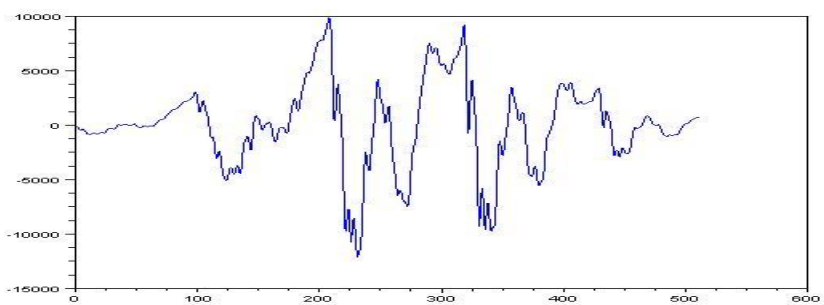


- The Hamming window

$$w(n) = 0.54 - 0.46\cos(2\pi \frac{n}{N}) \quad \text{with} \quad 0 \le n \le N$$

Or Hanning window

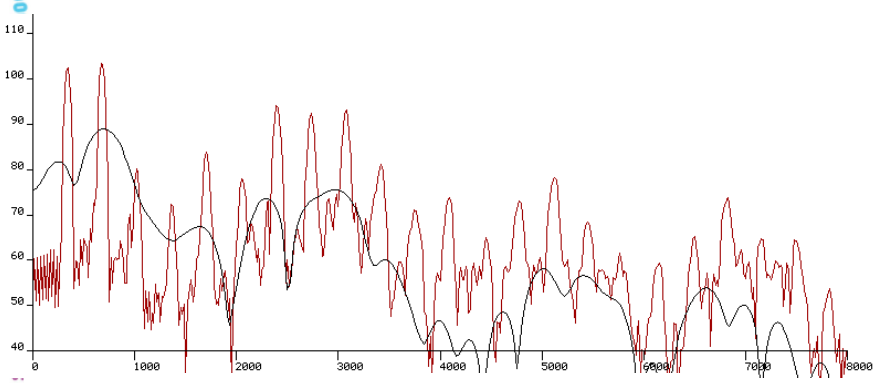
$$w(n) = 0.5(1 - \cos(2\pi \frac{n}{N-1}))$$



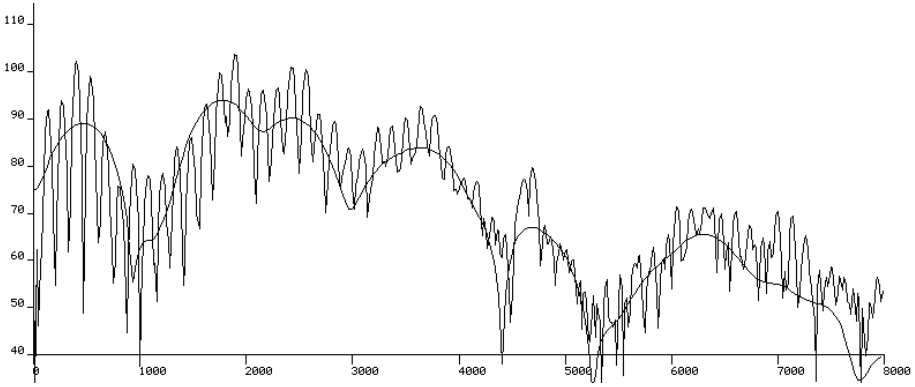
- The windowed signal

Practical implementation

- Sampling frequency between 10 and 22 kHz
- Running windows between 4 and 32 ms, shifted by half the duration of windows:
  - Small windows → wide band spectrograms
  - Long windows → narrow band spectrograms
- Often, the signal is pre emphasized $s'(n) = s(n) - \alpha\, s(n-1)$ so as to raise the contribution of high frequencies.
- Use of fast algorithms: Fast Fourier Transform (FFT)
- The log spectrum is displayed ($20 \log10|X(k)|$)

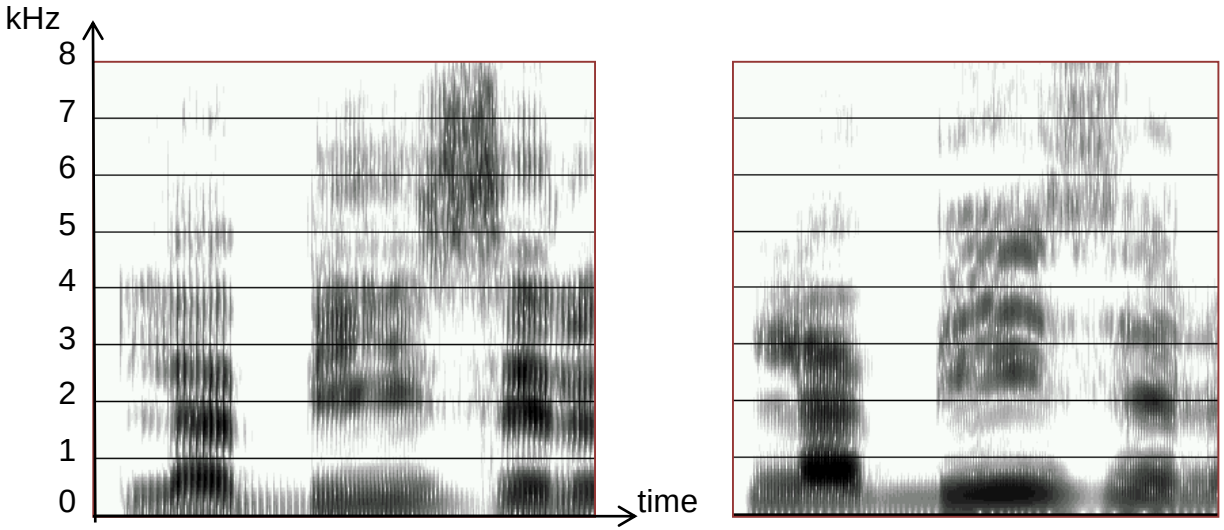


Female speaker, narrow and wide band spectra.

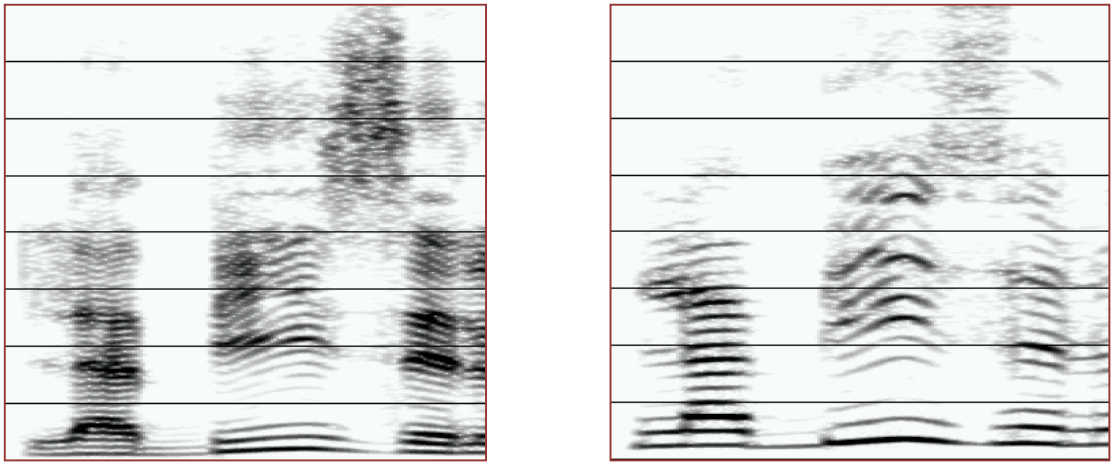


Male speaker, narrow and wide band spectra.

Examples



Wide band spectrograms (left: male, right: female)



Narrow band spectrograms (left: male, right: female)

❷ Other spectral analyses

- One difficulty faced with the Fourier transform:
  - Both the source (the vibration of vocal folds) and the vocal tract contributions are taken into account.
  - A simple source – vocal tract model:

$$s(n) = h(n) * e(n)$$

    the source $e(n)$ and the vocal tract $h(n)$ are convolved. The vocal tract behaves as a filter applied onto the source signal.
  - How can these two contributions be separated?
    - Cepstral filtering: a transform which isolates both contributions
    - Linear prediction: a filter model fitted on the speech signal

    Despite their interest none of these methods is completely satisfactory!

Cepstral smoothing

$$x(n) = x_1(n) * x_2(n)$$

1. Fourier transform to change from a convolution to a product

$$X(k) = X_1(k) \times X_2(k)$$

2. Logarithm (from a product to a sum)

$$\ln|X(k)| = \ln|X_1(k)| + \ln|X_2(k)|$$

3. Inverse Fourier transform (remains a sum but in the pseudo time domain)

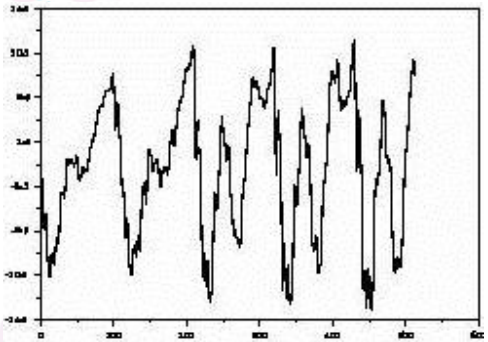
$$\hat{x}(n) = \hat{x}_1(n) + \hat{x}_2(n)$$

4. Linear processing (removing the source contribution)

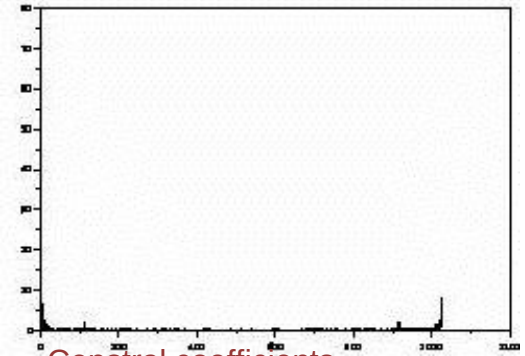
$$\hat{x}(n) = \hat{x}_1(n)$$

The signal contains (should contain) no more source contribution. Then it is possible to come back to a smooth spectrum by applying a Fourier transform.

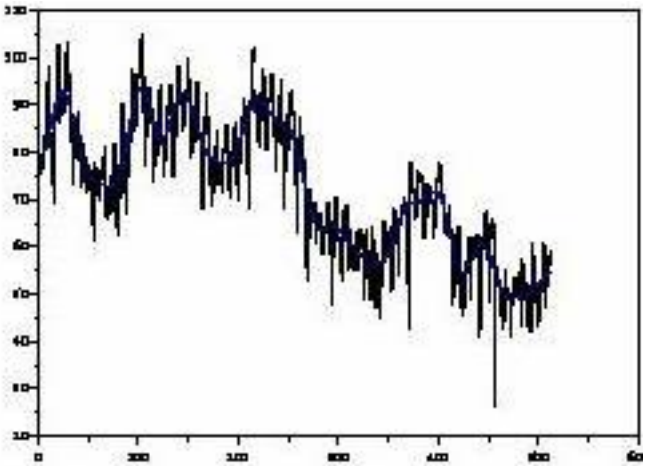
Example



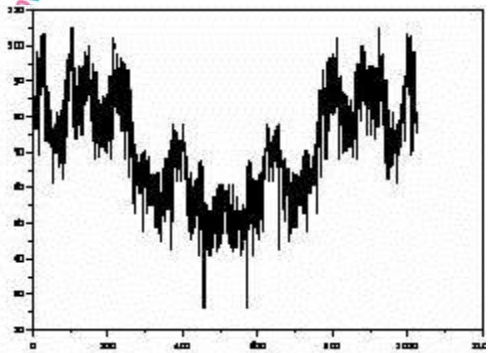
Speech signal



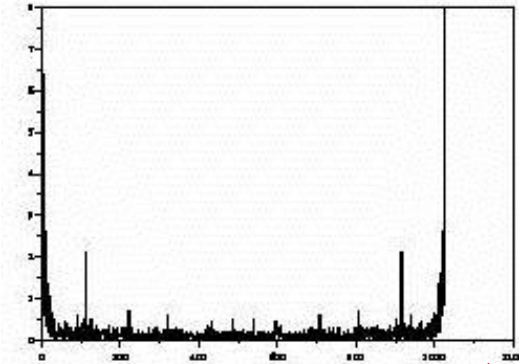
Cepstral coefficients



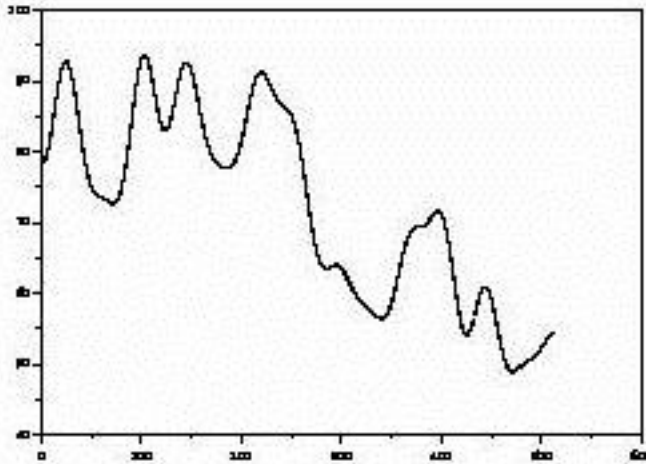
Spectrum (0 to π) and
cepstrally smoothed spectrum



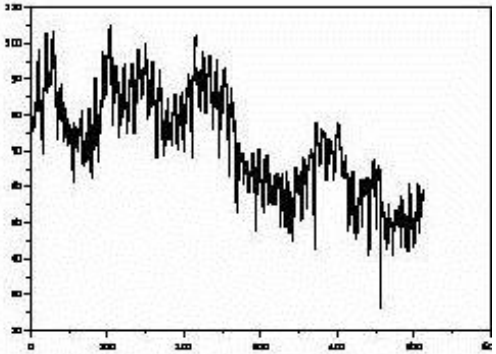
Spectrum (0 to 2π)



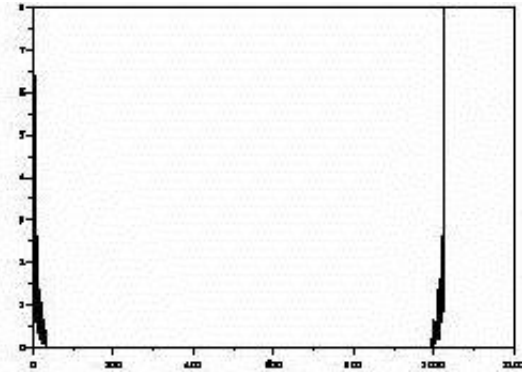
Cepstral coefficients (all but the 1st)



Cepstrally smoothed spectrum



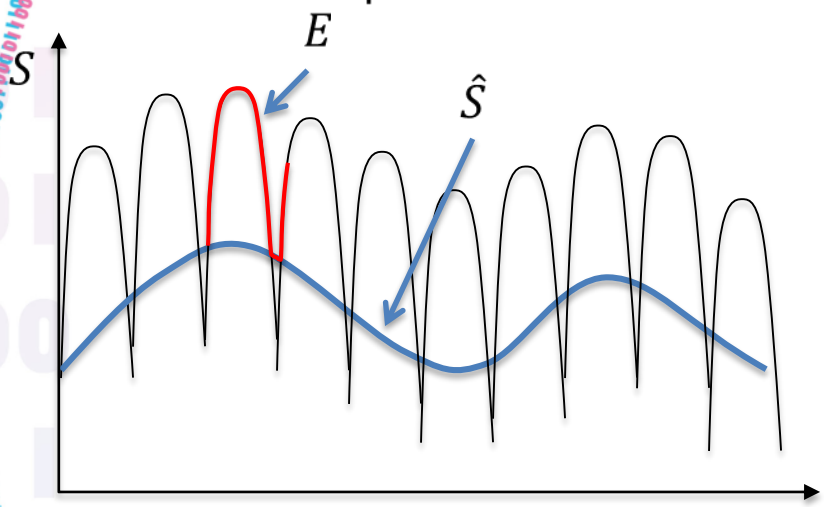
Spectrum (0 to π)

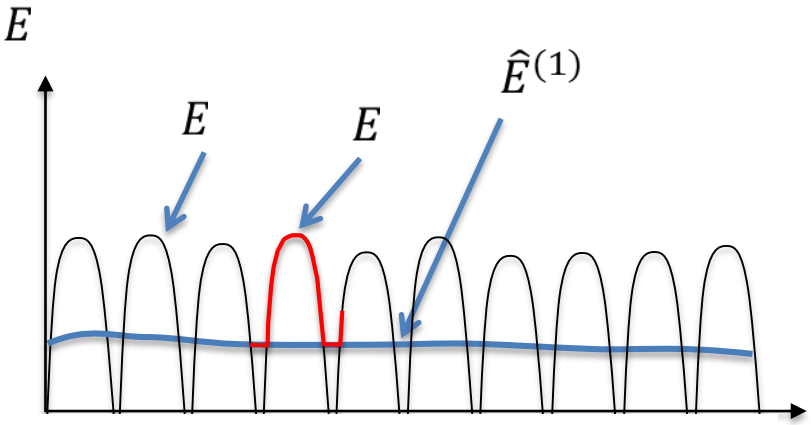


Cepstral coefficients after liftering

True envelope

- Objective: to get a smooth spectrum through the harmonics (perceived by the ear).
- Idea: start from the cepstral smoothing and correct it iteratively by discarding the spectral values below the smoothed spectrum.





- S spectrum
- $V^{(1)} = \hat{S}$ (cepstral smoothing)
- $E^{(1)} = g(S - \hat{S})$ where $g(y) = \max(y, 0)$

$E^{(1)}$ represents the spectrum above the cepstral smoothing

$\hat{E}^{(1)}$ represents the cepstral smoothing of the overrun.

True envelope algorithm

1. initial solution

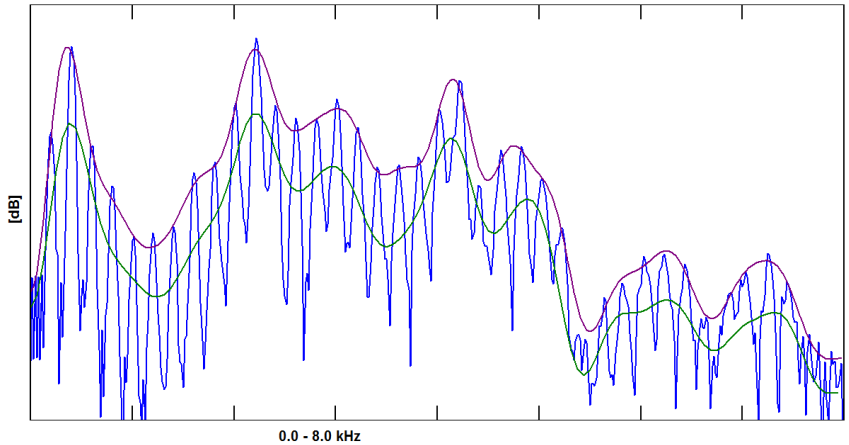
   $\hat{E}^{(1)} = \sum_{m=0}^{(N-1)} e_m^{(1)} h_m \cos\left(\frac{2}{N} mk\right)$ where $e^{(1)} = IDFT\left(E^{(1)}\right)$ and h_m is the liftering window

1. Iteration $i+1$

   Let $V^{(i)}$the envelope obtained at the previous step, $E^{(i)}$and $\hat{E}^{(i)}$ the overrun and smoothing of the overrun.

   - $V^{(i+1)} = V^{(i)} + \hat{E}^{(i)}$
   - $E^{(i+1)} = g(E^{(i)} - (1+\alpha)\hat{E}^{(i)}$ where α is a acceleration coefficient
   - $\hat{E}^{(i+1)} = DFT(h\left(IDFT\left(E^{(i+1)}\right)\right))$

3. End or new iteration



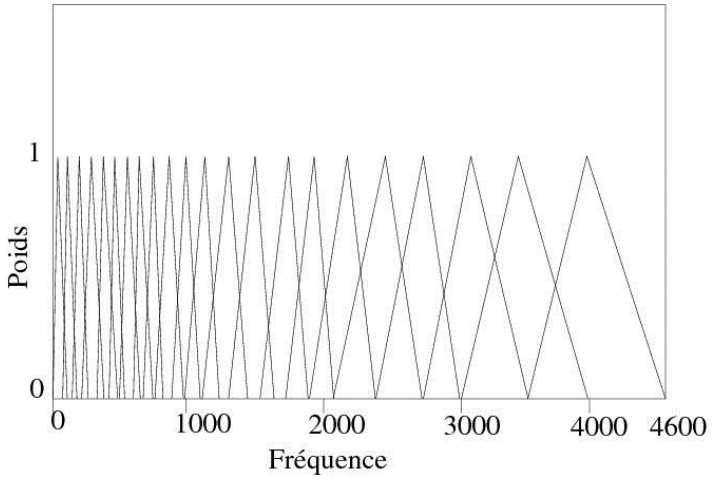
Narrow band spectrum, cepstral smoothing and true envelope.

Advantages:

1. No more energy variation due to the window position relative to the pitch period.
2. The spectrum fits harmonics.

Mel Frequency Cepstral Coefficients (MFCC)

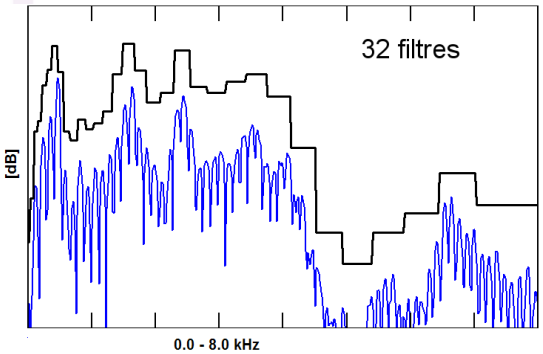
- The same principle but applied to a “perceptive” spectrum.
- The perceptive spectrum is obtained via filtering the magnitude spectrum with Mel filters

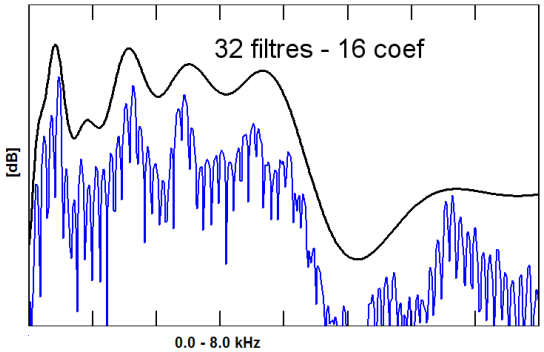


- The inverse Fourier transform is replaced by a discrete cosine transform (DCT).
- The MFCC are used in most of the automatic speech recognition systems.

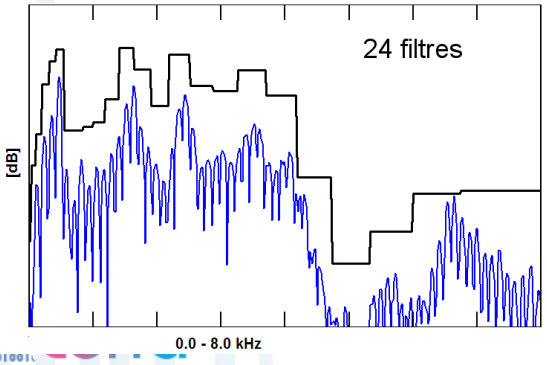
What do MFCC?

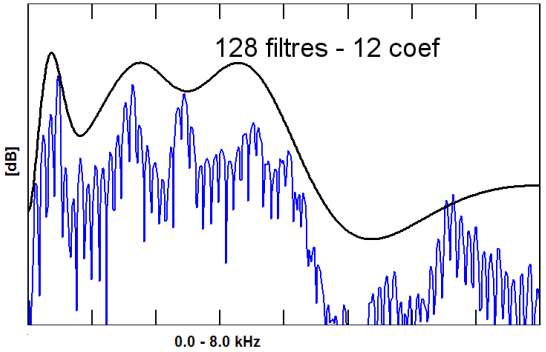
- Usually only MFCC coefficients are used without visualizing the corresponding smoothed spectrum.
- Here the smoothing is displayed.





32 filters, 16 coefs, filter output (left) smoothed spectrum (right)





24 filters, 12 coefs, filter output (left) smoothed spectrum (right)

High frequency integration: the higher the frequency, the stronger the smoothing.

Phonetic details can be deleted (F3 is replaced by spectral minimum).

Linear prediction

- Origin: speech signal is not a random signal, successive samples are correlated. Can this correlation be used to reduce the amount of data?
- Principle: $s(n)$ is represented as the sum of a linear combination of previous samples and an error.

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k)$$

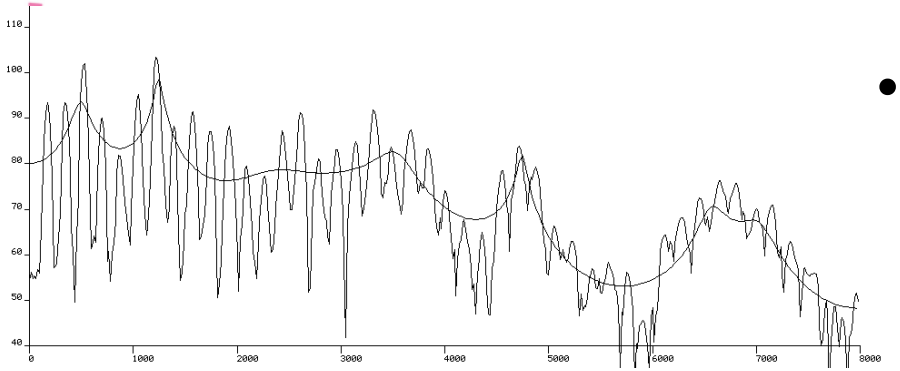
Coefficients are found by minimizing the error with the original signal.

$$E = \sum_m (s(m) - \hat{s}(m))^2$$

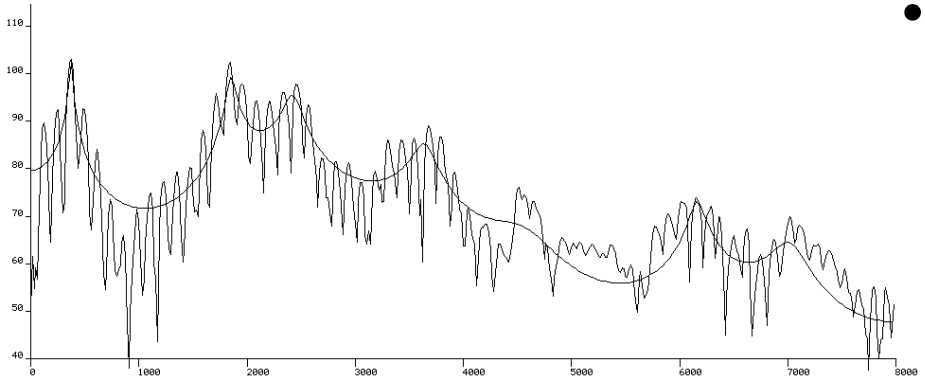
- From a spectral point of view the approximation corresponds to

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

Example



One example which fails /ɛ̃/



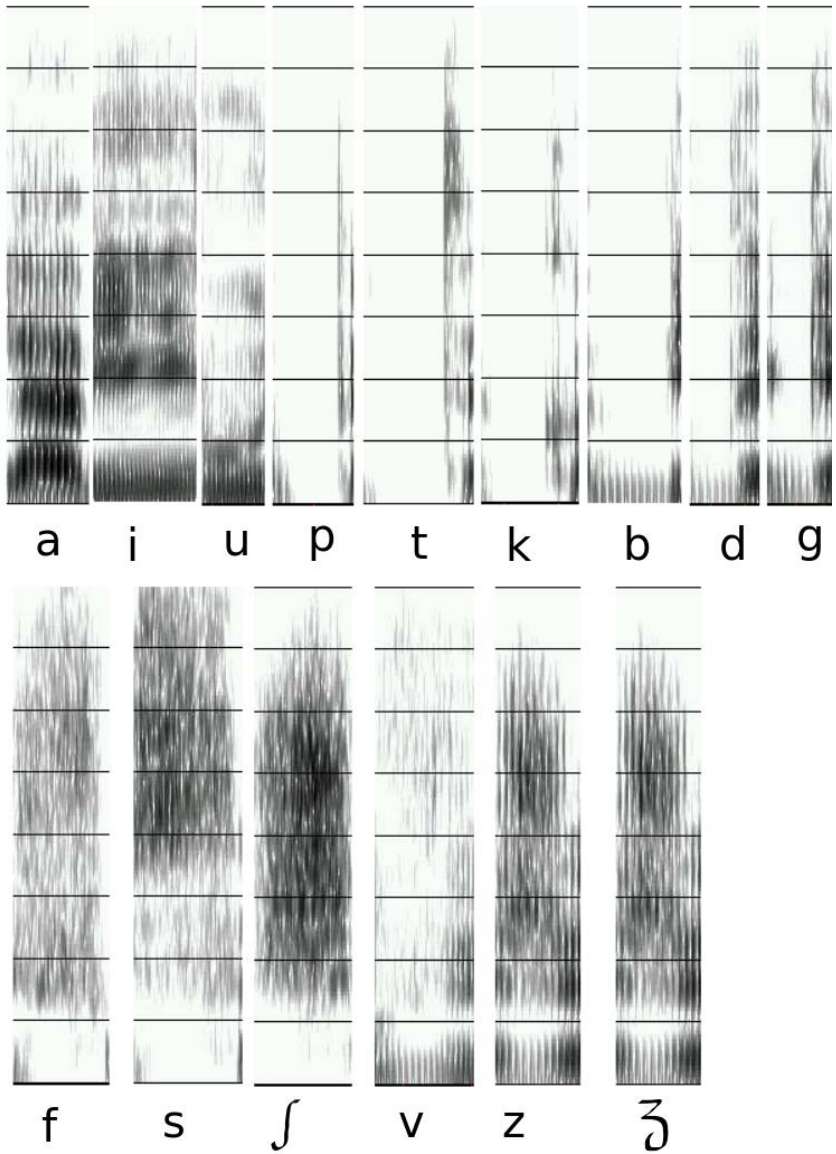
One example which succeeds /ɛ/

- Linear prediction corresponds an implicit physical model → all sounds which do not fit the hypothesis cannot be approximated correctly:
  - nasal vowels and all the nasalized sounds
  - Consonants
- Other variants exist:
  - Selective LPC (on a special region)
  - Perceptive LP (PLP) to mimic the peripheral auditory system.

❸ Spectral description of speech sounds

- Articulation modes
  - **vocalic** vibration of the vocal folds (voicing) and not too strong a constriction
  - **fricative** strong narrowing somewhere in the vocal tract creating a frication noise
  - **occlusive** partial or complete closure of the vocal tract, increase of the pressure behind the constriction and then brutal release which produces an explosion noise (burst).
- Place of articulation = location of the main constriction of the vocal tract: pharynx, palate /k/, teeth /t/, lips /p/

Cardinal vowels and consonants of French



vowels

F2

front

back

close

open

F1

/i/

/y/

/u/

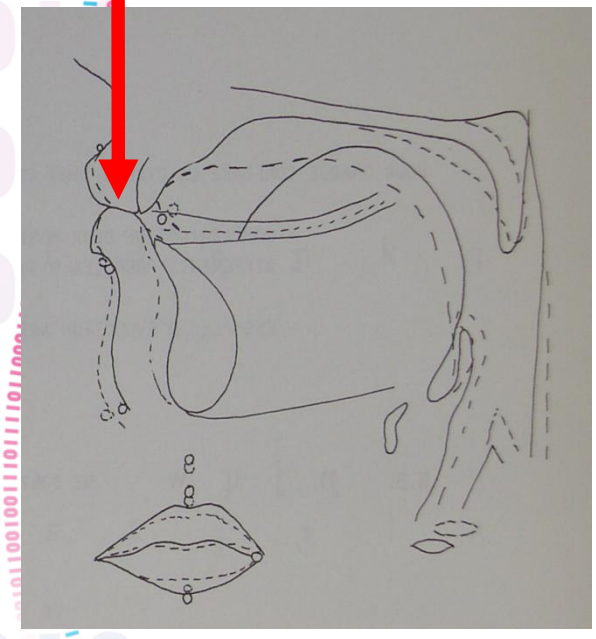
/a/

Effect of stops(CV)

- Labialization lengthens the vocal tract and thus tends to lower formant frequency at the consonant.
- In general, bursts of /p/ are shorter than those of /t/.
- In general, bursts of /k/are long.
- F2 et F3 of central vowels get closer in the context of /k,g/
- For back vowels there is often a peak in front of F2 for /k,g/.
- /t,d/ present a locus (called dental locus) for F2 between 1500 and 2000 Hz. → strong transition for F2 in case of a back vowel.

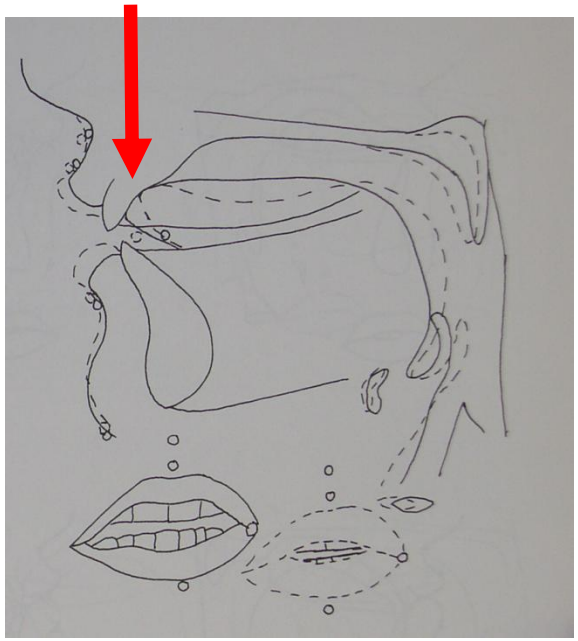
Places of articulation of French stops

/p/



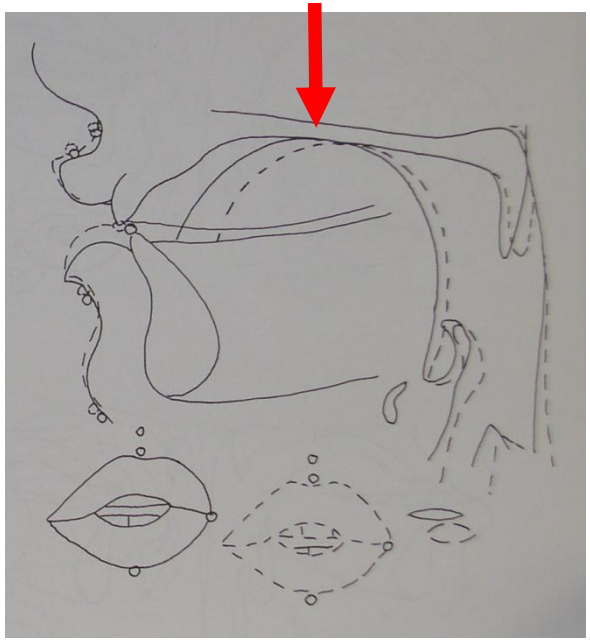
/pət,pRɛ/
dotted, solid

/t/



/pat,tab/
dotted, solid

/k/



/ky,ku/

❹ Pitch determination

- Remark: pitch is not F0 (fundamental frequency):
  - Set F0 at 50 Hz and select the 13th , 25th and 29th harmonics → gives a pitch at 334 Hz or 650 Hz.
  - Probably some perceptual adjustments at voicing onset when vocal folds do not vibrate at the target F0.
- Language learning → pitch since the objective is use perception.

Pitch determination algorithms

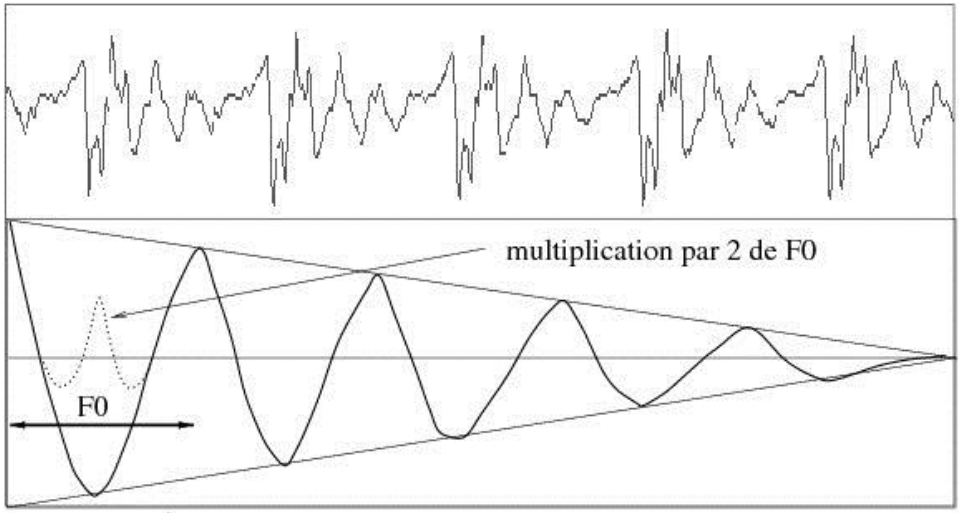
- General idea:
  - Combine several F0 determination algorithms
  - Provide results together with a confidence measure
- Available F0 determination algorithms:
  - spectral comb (Martin) – spectral
  - Yin (Kawahara & De Cheveigné) – time
  - Swipe (Camacho & Harris) – spectral

Autocorrelation method

- Autocorrelation method (temporal domain)
  - Calculation of the autocorrelation function

$$\phi(k) = \sum_{i=0}^{N} x(i)x(k+i)$$

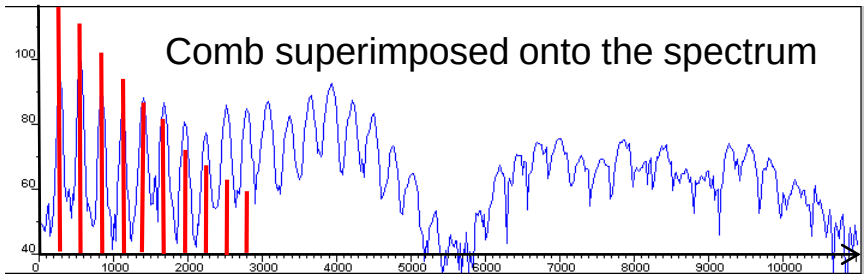
k is a shift, Φ is maximal when k is the F0 period



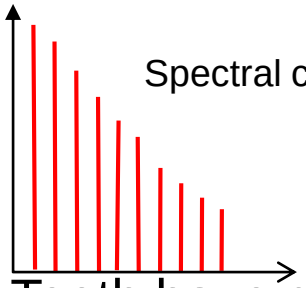
  - Many technical problems (pitch doubling or halving, voicing decision…) require elaborated correction algorithms.

A spectral method (Martin)

- Intercorrelation between a narrowband spectrum and a spectral comb



$$I(\omega_c) = \int_0^{F_s/2} P(\omega_c, \omega) |F(\omega)| d\omega$$

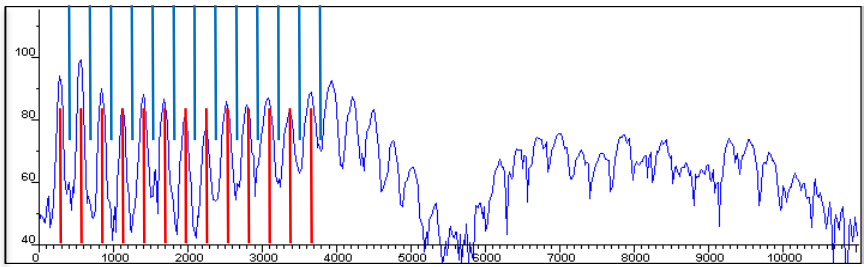


Teeth have decreasing height to avoid finding F0/2 instead of F0.

- A correction algorithm is required.

SWIPE (Camacho & Harris)

- Maximizing the difference between harmonics and valleys



Criterion to be maximized:

$$D_n(f) = \frac{1}{n}\left(\frac{1}{2}\left|X\left(\frac{f}{2}\right)\right| - \frac{1}{2}\left|X\left(\left(n+\frac{1}{2}\right)f\right)\right| + \sum_{k=1}^{n} |X(kf)| - \left|X\left(\left(k-\frac{1}{2}\right)f\right)\right|\right)$$

- Similarly to the spectral comb teeth have not the same amplitude but $\frac{1}{k^p}$
- And other improvements (blurring the harmonics…)

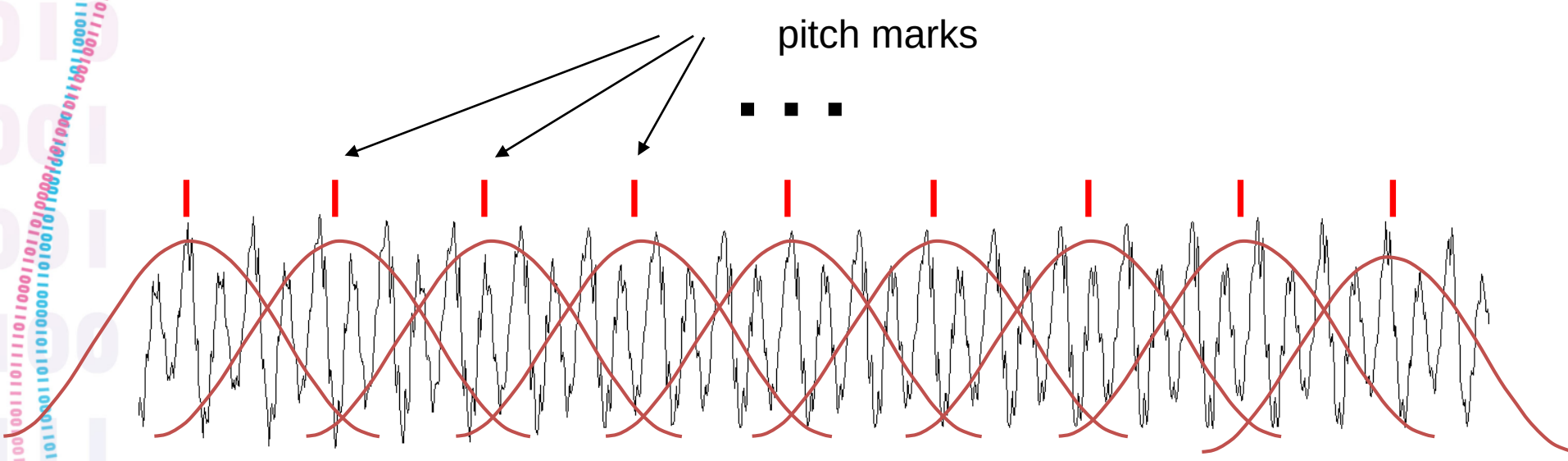
Combining several pitch determination algorithms

- F0 from the algorithms presented above and with several parameter setups in order to get all relevant candidates
- Additional information to get the voicing determination:
  - Energy
  - Mel cepstral coefficients
- Annotated speech corpora in terms of F0 + corrupted versions of these corpora to learn (DNN ?) F0 together with a confidence measure.

❺ PSOLA- modifying the fundamental frequency

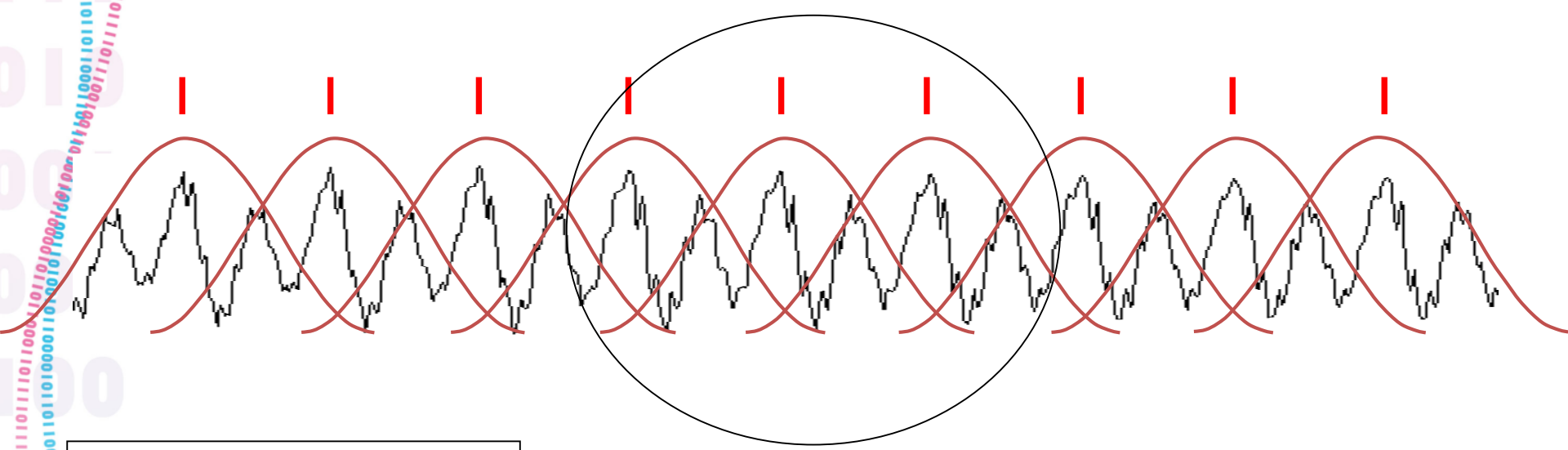
- Pitch Synchronous Overlap and Add :
  - Proposed by Charpentier et al. in 1987
  - Decomposition of the speech signal into overlapping windows synchronized with F0
  - Very simple from an algorithmic point of view (only a sum and a division for every sample synthesized.
  - Requires a speech database whose pitch marks are known (detected automatically or manually).

Decomposing a speech signal into overlapping windows

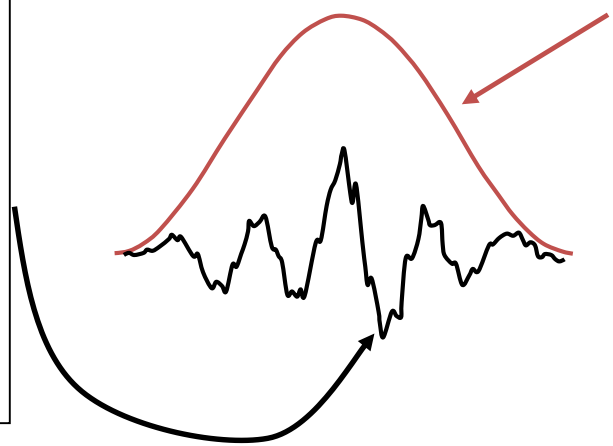


Requires a pitch marking algorithm.

Decomposing the signal into pitch synchronous signal windows



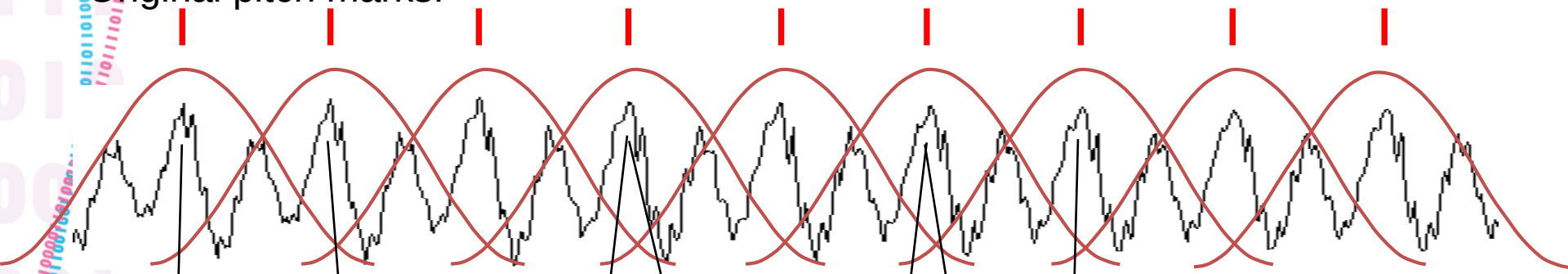
Windowed signal. The signal can be reconstructed by summing windowed signals. Each window has the same spectral properties as the original signal.



A hamming window whose size is twice that of a fundamental window.

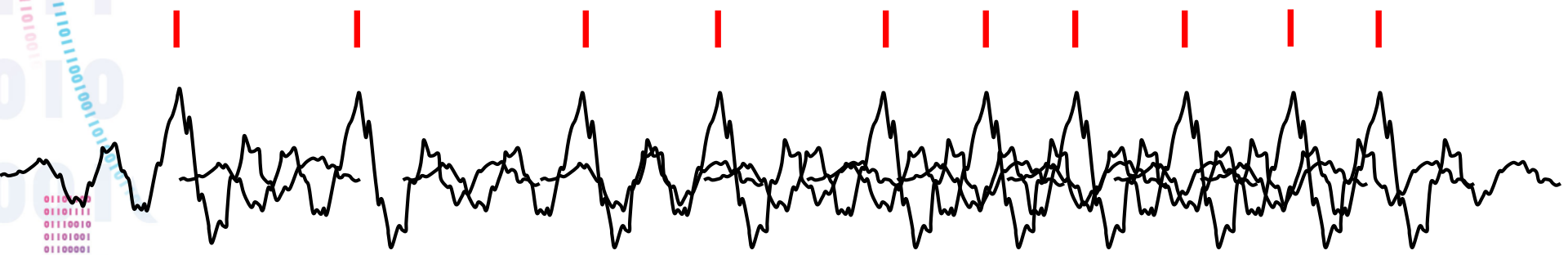
Modification of the fundamental frequency

Original pitch marks.



The signal window selected is the closest from the target pitch mark.

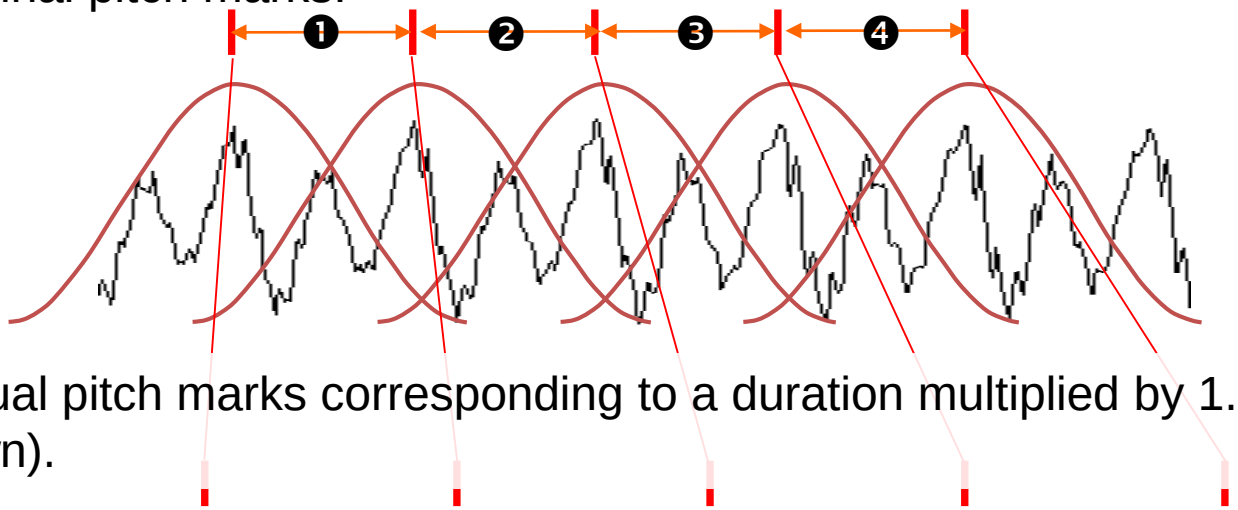
Target pitch marks.



And finally: summing and weighting.

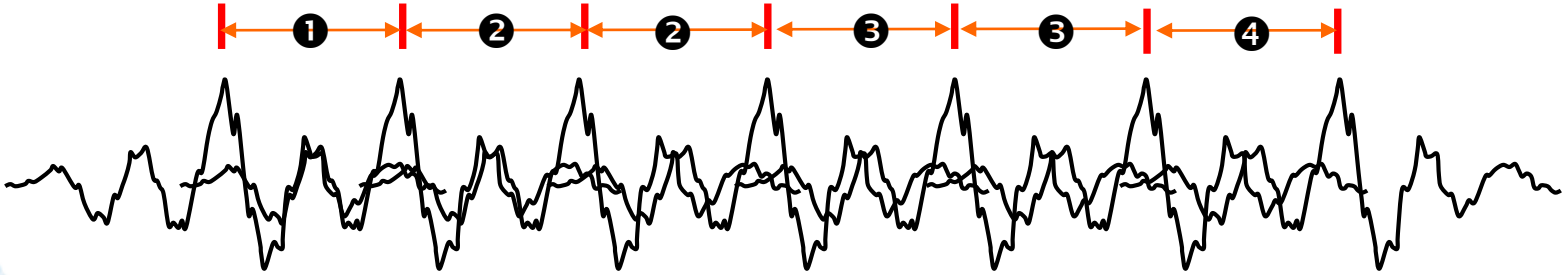
Modifying the speech signal duration (slowing down or speeding up)

Original pitch marks.



Virtual pitch marks corresponding to a duration multiplied by 1.4 (slowing down).

Copying the window whose virtual pitch mark is the closest from the synthetic pitch mark.



And finally: summing and weighting.

Summing and weighting

- Unlike the classical OMA method weighing by Hamming windows has to be taken into account explicitly since windows are not spaced from a quarter of window size.

$$s(n) = \frac{S(n)}{WeightingSum(n)}$$

s(n) is the new signal, *S(n)* is the sum of windowed signals and *WeightingSum(n)* is simply obtained by summing all the windows contributing to the sample *n*.

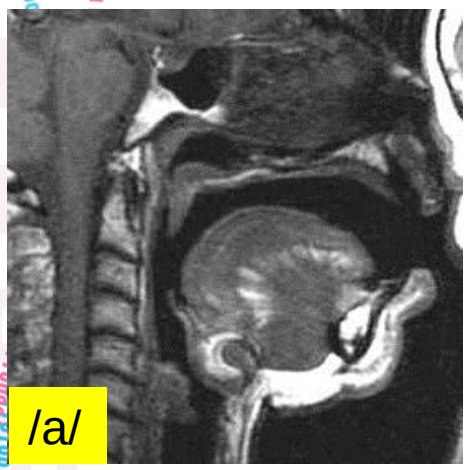
- Caution, it is not possible to space windows too much otherwise the signal is not defined everywhere.

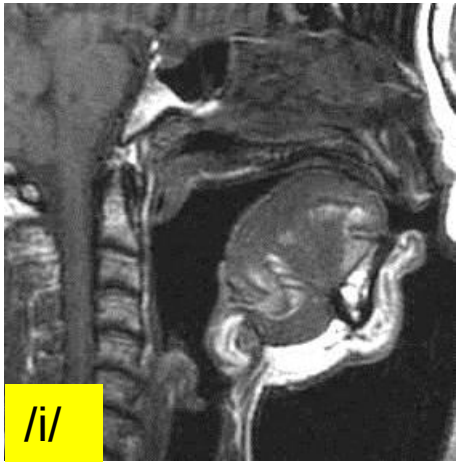
❻ Using resonators to synthesize speech: formant synthesis

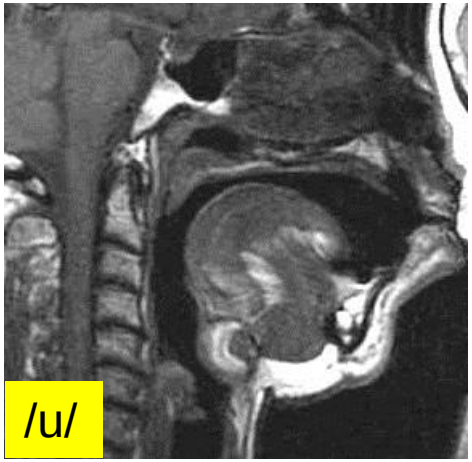
- Idea: represent spectral maxima (formants) by second order resonators.
- Specify the source parameters:
  - Voiced source → vowels and other voiced sounds
  - Noise source → unvoiced stops and fricatives
- Noise to be done to synthesize a speech
  - Specify temporal evolution rules for these parameters and for all phonetic contexts.
  - Parameters should represent speech faithfully. ..
- This approach of synthesis is not used anymore but:
  - this is a good example of speech analysis
  - This is useful to generate speech stimuli and to analyze pathological voices

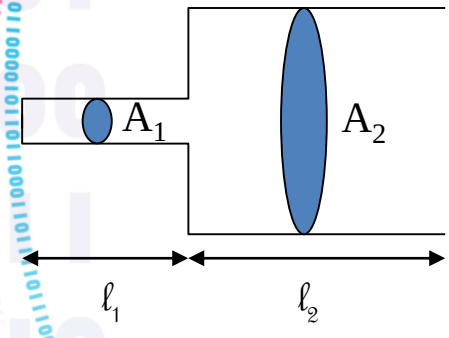
Loria

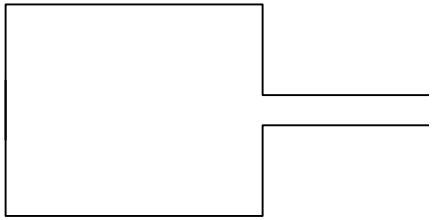
Resonance frequencies of vowels

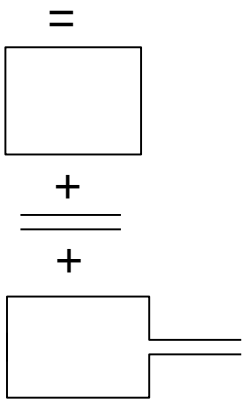


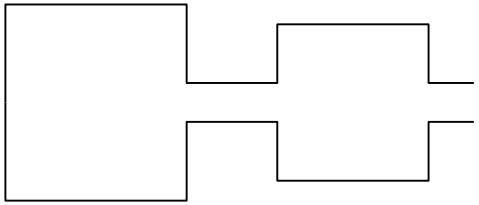












Resonator

- A resonator

$$y(n) = As(n) + By(n-1) + Cy(n-2)$$

where s(n) is the source signal and y(n) the synthetic signal.

- Its transfer function:

$$H(z) = \frac{A}{1 - Bz^{-1} - Cz^{-2}}$$

where A, B et C are defined from the characteristics of formants (resonance frequencies).

$$\begin{cases} B = 2e^{-\pi B_w T}\cos(2\pi F_i T) \\ C = -e^{-2\pi B_w T} \\ A = 1 - B - C \end{cases}$$

F is the frequency and B_w then formant bandwidth.

Two useful reminders

Transform of a signal shifted in time

- Z transform

$$G(w) = \sum_{-\infty}^{\infty} s(n-k) z^{-n} = \sum_{-\infty}^{\infty} s(m) z^{-(m+k)} = z^{-k} \sum_{-\infty}^{\infty} s(m) z^{-m}$$

$$G(w) = z^{-k} F(w)$$

- With the Fourier transform, $z = e^{i\omega}$

$$TF(s(n-k), \omega) = e^{-i\omega k} TF(s(n), \omega)$$

$$X_{s(n-k)}(e^{j\omega}) = e^{-i\omega k} X_{s(n)}(e^{j\omega})$$

- And phase and its derivative with respect to frequency: $\arg(X_{s(n-k)}(e^{j\omega})) = \omega k + \arg(X_{s(n)}(e^{j\omega}))$

$$\frac{d \arg(X_{s(n-k)}(e^{j\omega}))}{d\omega} = k + \frac{d \arg(X_{s(n)}(e^{j\omega}))}{d\omega}$$

Transfer function of a resonator

$$y(n) = As(n) + By(n-1) + Cy(n-2)$$

$$Y(z) = AS(z) + Bz^{-1}Y(z) + Cz^{-2}Y(z)$$

$$Y(z)(1 - Bz^{-1} - Cz^{-2}) = AS(z)$$

$$Y(z) = \frac{A}{1 - Bz^{-1} - Cz^{-2}} S(z)$$

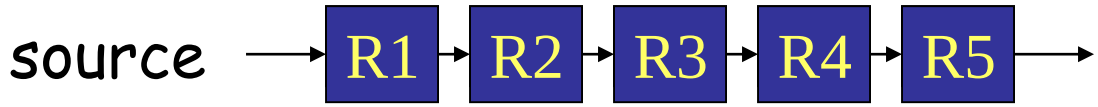
- And more generally:

$$y(n) = \sum_{k=1}^{N} a_k y(n-k) + \sum_{k=0}^{M-1} b_k s(n-k)$$

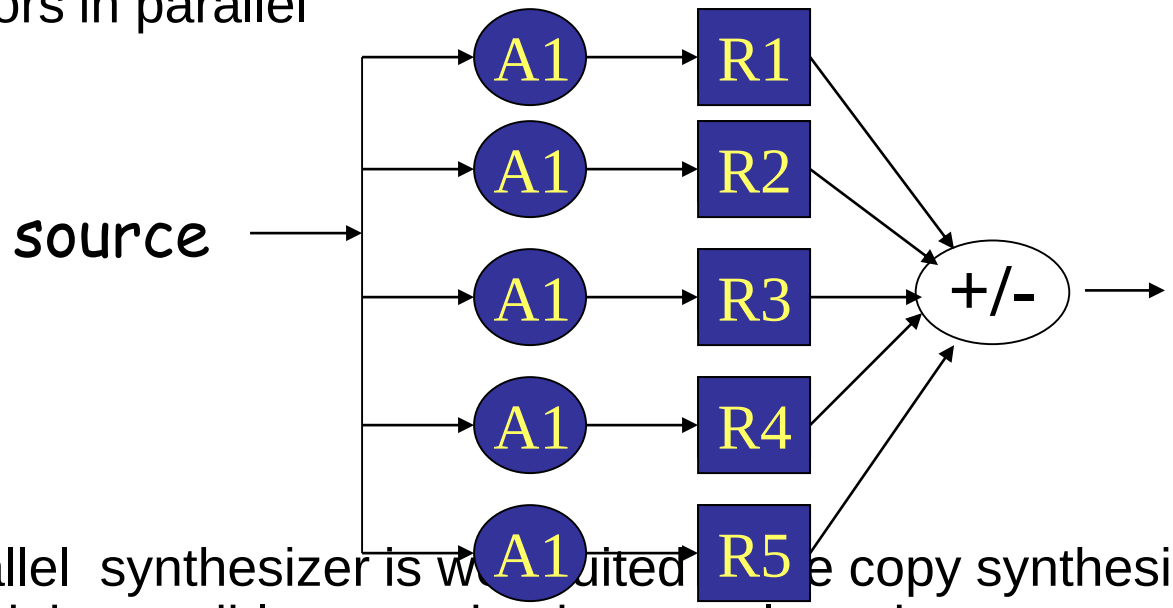
$$H(z) = \frac{Y(z)}{S(z)} = \frac{\sum_{k=0}^{M-1} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}$$

The Klatt formant synthesizer

- Resonators in cascade



- Resonators in parallel



- The parallel synthesizer is w uited e copy synthesis because it is possible to synthesize vowels and consonants as well.

Transfer function of the synthesizer

- For the parallel branch:

$$H_p(z) = -D_1 \times H_1(z) + D_2 \times H_2(z) - D_3 \times H_3(z) + \ldots$$

with $$H_1(z) = \frac{1 - B_1 - C_1}{(1 - B_1 z^{-1} - C_1 z^{-2})(1 - z^{-1})}$$

- For the cascade branch:

$$H_c(z) = H_1(z) \times H_2(z) \times H_3(z) \times \ldots$$

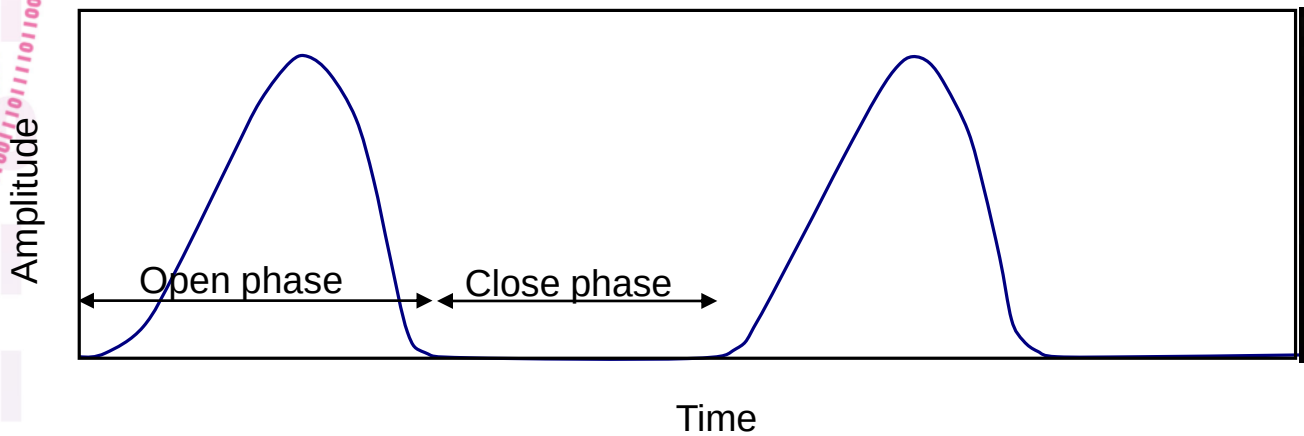
- For the whole:

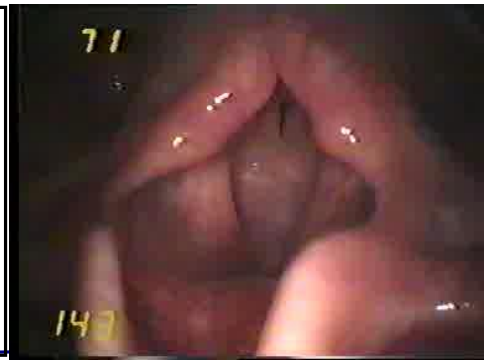
$$P(z) = (H_c(z) + H_p(z)) \times S(z)$$

where *S(z)* is the spectrum of the source plus the lip radiation.

And the source?

- Periodic signal → voiced source





- Create an artificial source signal (Rosenberg (1971) source, used by Klatt and called KLglott88)

$$\begin{cases} U_g(t) = at^2 - bt^3 & \text{for} \quad 0 \le t < O_q T_0 \\ U_g(t) = 0 & \text{for} \quad O_q T_0 \le t < T_0 \end{cases} \qquad a = \frac{27 A_v}{4 T_0 O_q^2} \quad \text{and} \quad b = \frac{27 A_v}{4 T_0^2 O_q^3}$$

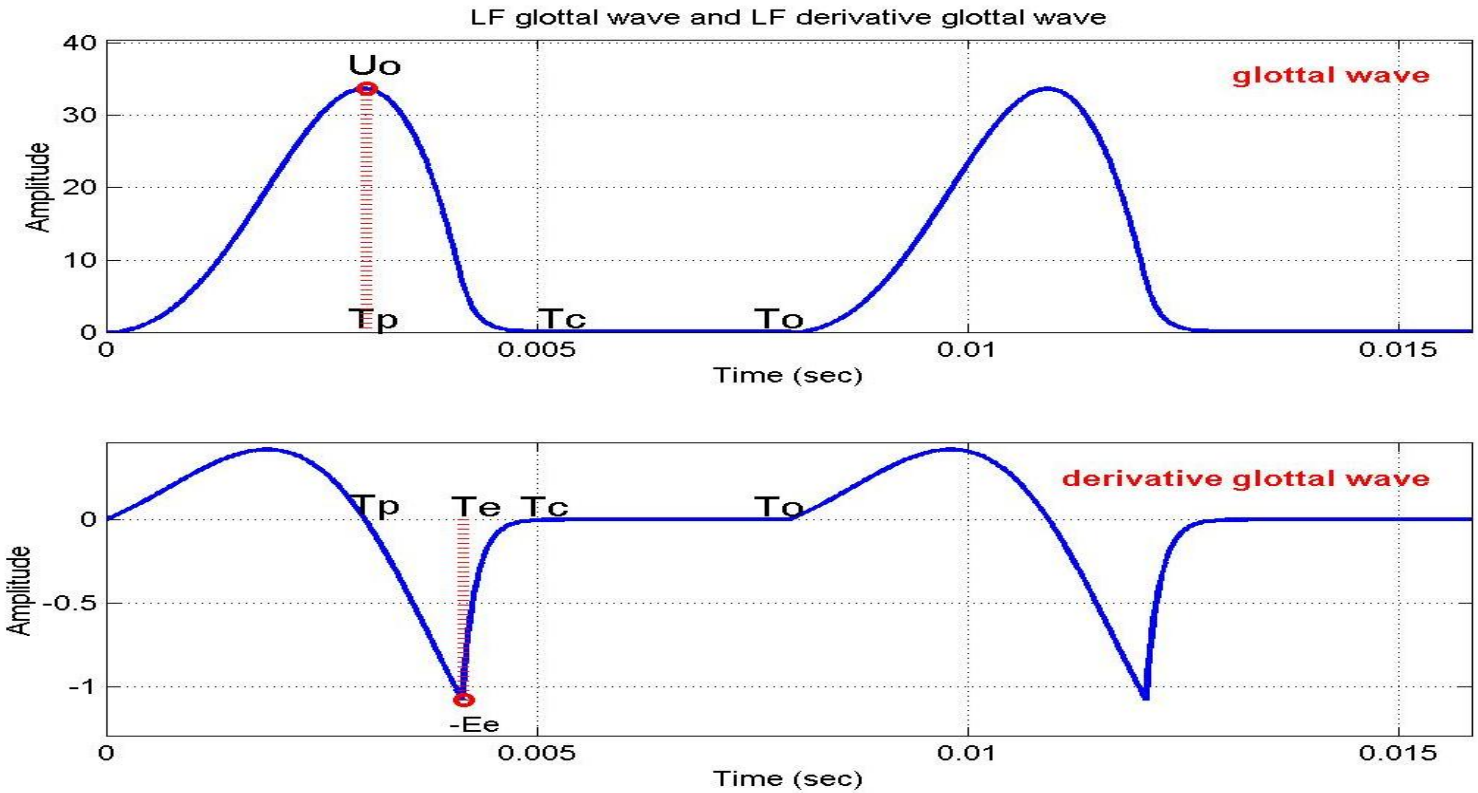
O_q is the open quotient and A_v the amplitude of voicing.

- Noise → fricatives, bursts, noise in high frequency

Another famous source: Liljencrants-Fant (LF) model

$$g(t) = E_0 e^{\alpha t} \sin(\omega_g t) \qquad 0 \le t \le t_e$$

$$= -\frac{E_e}{\varepsilon t_a}\left[e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}\right] \qquad t_e \le t \le t_c \le t_0$$



Determination of the LF parameters (1/2)

Parameter	Description
E_e	Maximum of the negative derivative of the flow
R_a	Ratio of t_a over t_c-t_e
R_k	Ratio of t_a over t_c-t_e
R_g	The ratio of half-period of F0 over t_p

- α, ε and ω_g have to be determined from R_a, R_k and R_g.

1. At time t_e g equals E_e and thus: $\varepsilon t_a = 1 - e^{-\varepsilon(t_c - t_e)}$
2. No increase of the air flow during a period.

Determination of the LF parameters (2/2)

1. Newton by setting ε at $1 / t_a$
   after a glance on the function with Matlab

$$\varepsilon t_a - 1 - e^{-\varepsilon(t_c - t_e)}$$

2. Is equivalent to $\int_0^{t_0} E(t)dt = 0$ or:

$$\int_0^{t_e} E(t)dt = -\int_{t_e}^{t_0} E(t)dt$$

qui est résolu une fois de plus avec Newton en partant de $\alpha = 0$ (en faisant attention que cette seconde solution peut ne pas avoir de solution).

Speech analysis - Conclusions

- Spectral analysis
  - Results are obtained via a computation.
  - Results are exact
  - Results are relevant provided that relevant parameters have be chosen correctly.
- Extraction of speech parameters (formants, F0…)
  - Results are obtained via an algorithm
  - Results may be erroneous depending on the reliability of the algorithm and the quality of the speech signal
  - Inspect data before further processing, determine parameters by hand in some cases to get a first evaluation.

❼ Acoustics of the vocal tract

Equations of acoustics

Assumptions:

- One dimensional plane propagation

→ the vocal tract may be unfolded without changing solutions

Acoustic variables:

- Particule velocity $\boldsymbol{v(t,x)}$
- Volume velocity $\boldsymbol{V(t,x)}$ $(V = vA)$
- Sound pressure variation $\boldsymbol{p(x,t)}$ $(P = P_0 + p)$
- Density of air $\boldsymbol{\rho}$
- Velocity of sound $\boldsymbol{c}$

Geometry given by $\boldsymbol{A(x,t)}$

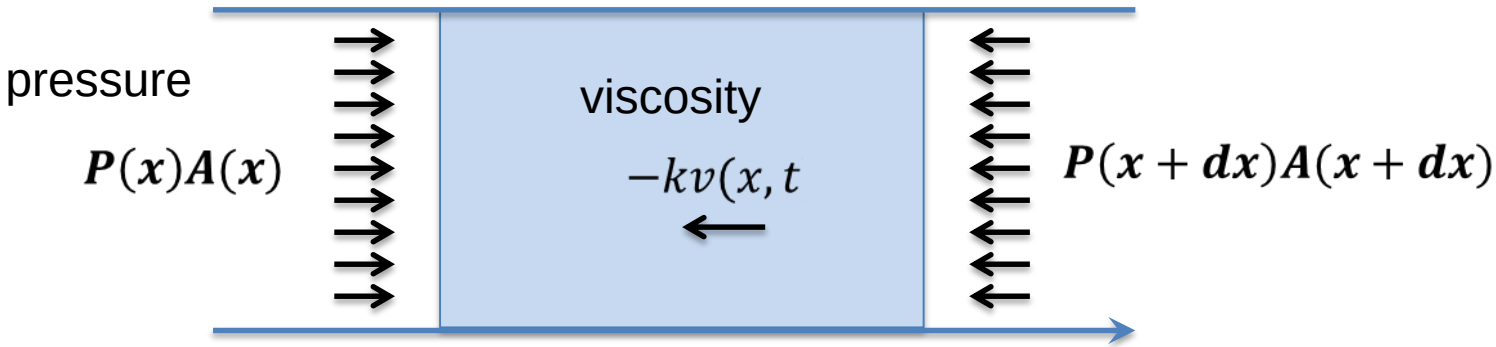
Air is considered as an ideal gaz. This means that p and ρ are linked by:

$$\boldsymbol{\rho = \rho_0 + \frac{1}{c^2} p}$$

and their derivatives by:

$$\boldsymbol{\frac{\partial \rho}{\partial t} = \frac{1}{c^2} \frac{\partial p}{\partial t}}$$

Euler equation



- Sum of forces applied to a small volume of air at one time point:

$$F = -\frac{\partial}{\partial x}(AP)dx - kv(t,x)$$

- Derivative of the momentum:

$$\mathrm{m}\frac{d}{dt}(v) = A(x,t)\rho(x,t)\frac{dv}{dt}dx$$

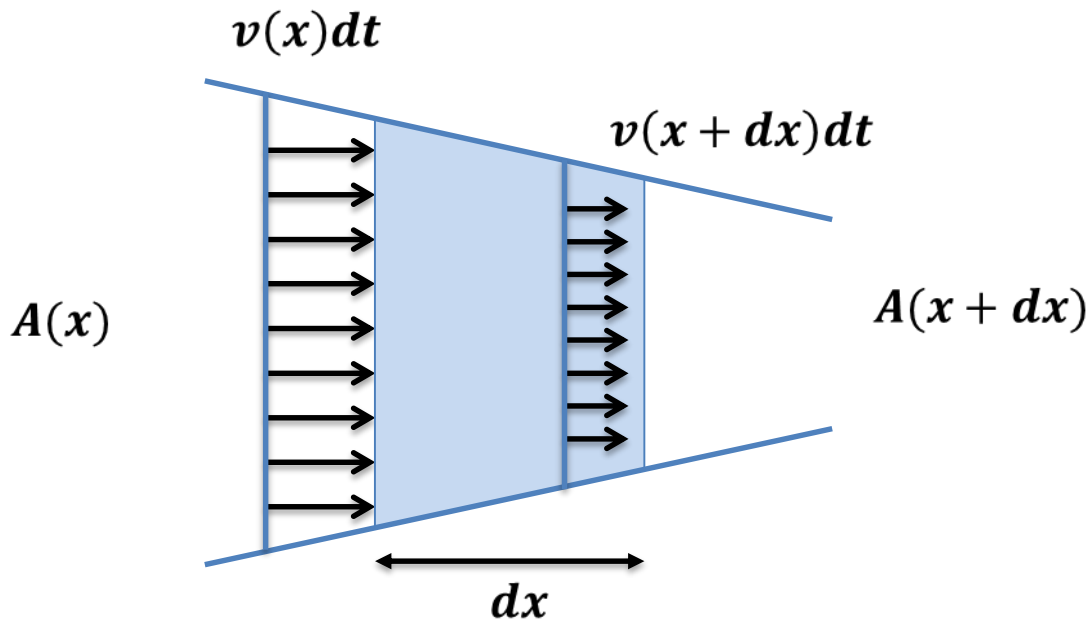
- v is a function of space and time. Hence $\mathrm{dv} = \frac{\partial v}{\partial t}dt + \frac{\partial v}{\partial x}dx$ or equivalently $\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x}\frac{dx}{dt}$ or:

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x}$$

- $A(x,t)\rho(x,t)\left(\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x}\right) = -\frac{\partial}{\partial x}(Ap) - kv$

- The sound pressure variation is about 1 Pa (<< atmospheric pressure)
- Velocity 10^{-7}ms^{-1} at the hearing threshold
- Classical simplifications:
  - A is constant in a uniform tube
  - ρ is almost constant $\rho = \rho_0$
  - $\frac{\partial v}{\partial x}$ is very small since velocity is small, and $v\frac{\partial v}{\partial x}$ is negligible

Equation of continuity



- Increase of mass within dt:

$$A(x)\rho(x,t)v(x,t) - A(x+dx,t)\rho(x+dx,t)v(x+dx,t) =$$
$$-(\frac{\partial A}{\partial x}\rho(x,t)v(x,t) + \frac{\partial \rho}{\partial x}A(x)v(x,t) + \frac{\partial v}{\partial x}A(x)\rho(x,t)dxdt$$

- $\frac{\partial \rho}{\partial x}$ is small and the second term is thus negligible. The flow of mass is thus:

$$-\rho_0\left(\frac{\partial A}{\partial x}v(x,t) + \frac{\partial v}{\partial x}A(x)\right)dxdt$$

- $\frac{\partial}{\partial t}(A\rho)dt\,dx$ is the variation of mass in the volume and thus

$$-\rho_0\left(\frac{\partial A}{\partial x}v(x,t)+\frac{\partial v}{\partial x}A(x)\right)dxdt=\frac{\partial}{\partial t}(A\rho)dt\,dx$$

Or equivalently:

$$-\rho_0\left(\frac{\partial A}{\partial x}v(x,t)+\frac{\partial v}{\partial x}A(x)\right)=\frac{\partial}{\partial t}(A\rho)$$

3 equations and 3 unknowns, solving is thus possible…

Properties of the wall

Vibration of the wall: $m\ddot{y} + b\dot{y} + k(y - y_0) = p(x,t)S(x,t)$ with

$$A(x,t) = A_0(x,t) + y(x,t)S_0(x,t)$$

Dynamic vocal tract $A = A(x,t)$ thus $\frac{\partial A}{\partial t} \neq 0$

Boundary conditions:

$$p(x = 0, t) = P_{subglottic}$$

$$p(x = lips, t) = 0$$

Nasal coupling: conservation of the airflow, continuity of the pressure

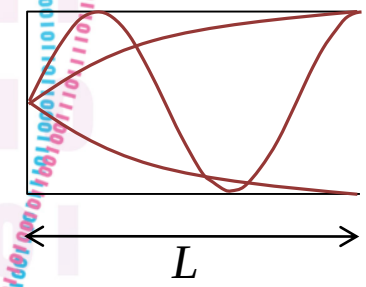
Radiation at lips: the sound perceived is not that at the very output of lips

Losses: due to viscosity and/or vibration of the vocal tract wall.

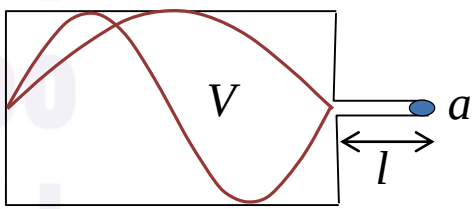
Solving equations of acoustics

- Finite difference equations (time and space)
- Equivalence between acoustics and electricity.

Tubes forming the vocal tract



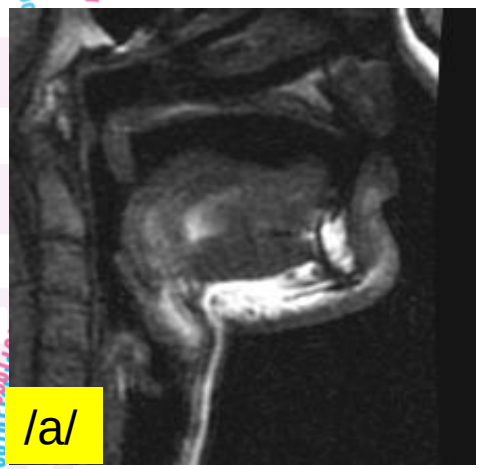
- Tube closed at one end and open at the other:
  - quarter wavelength resonator
  - resonance frequencies: $(2n-1)\frac{c}{4L}$
  - Exercise: find resonance frequencies for L=17 cm and c = 350m/s

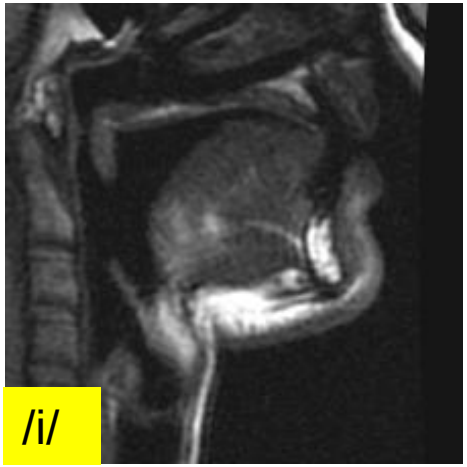


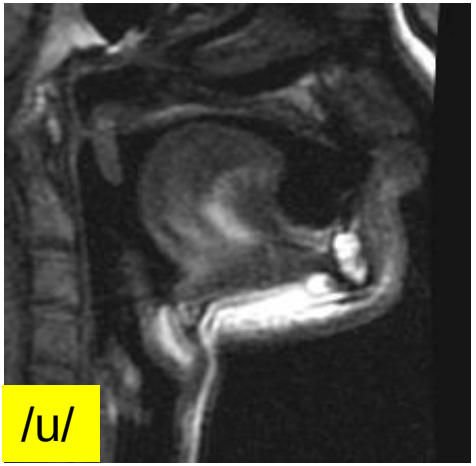
- Tube (almost close at both end)
  - half-wavelength resonator
  - Resonance frequencies: $n\frac{c}{2L}$
  - Helmholtz frequency at low frequency: $\frac{c}{2\pi}\sqrt{\frac{a}{lV}}$

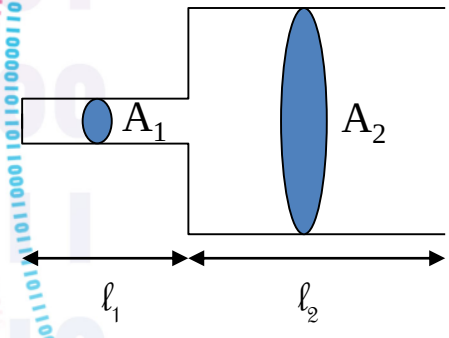
    where V is the volume of big tube, a the area and l the length of the small tube (the neck).

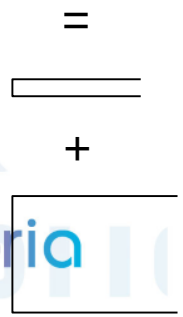
Acoustic properties of vowels

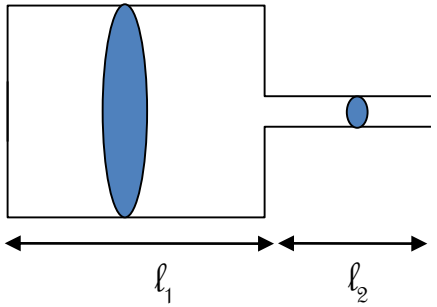


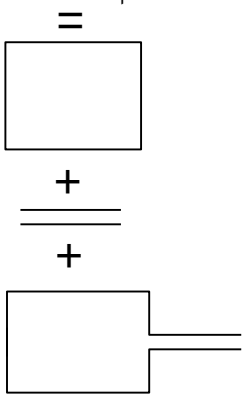


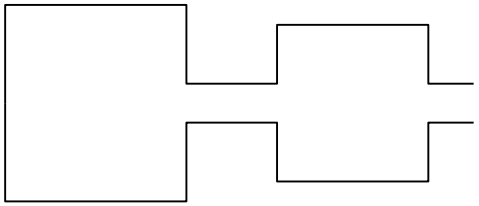




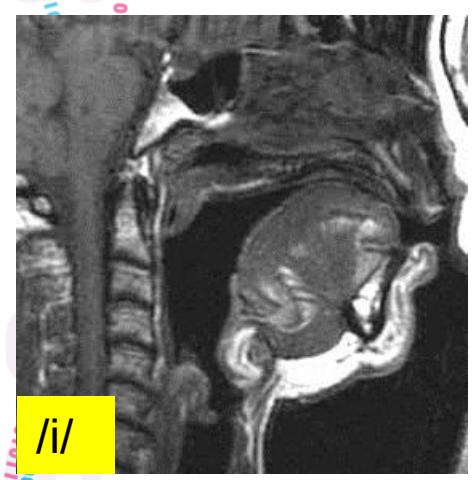


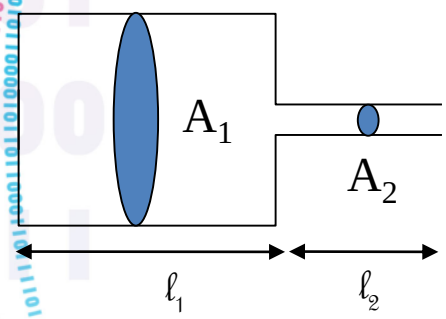


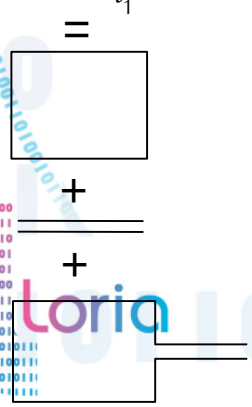




Exercises: vowel /i/

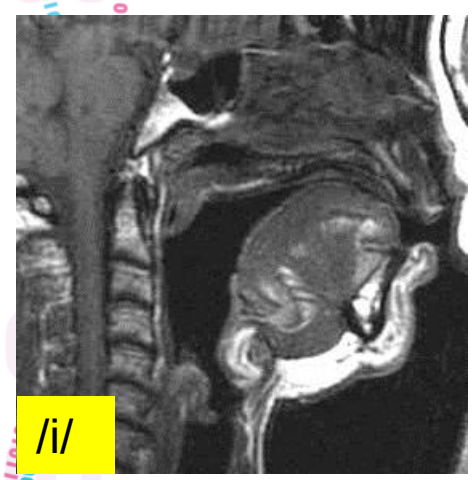


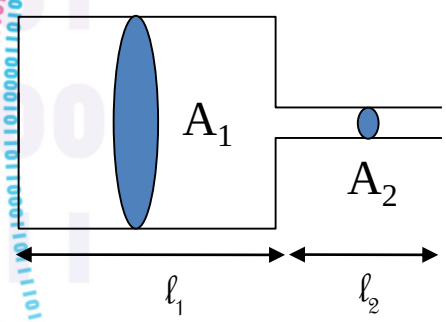


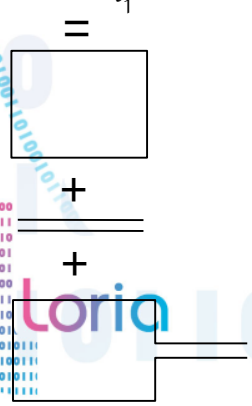


- l_1 = 9cm, l_2= 6cm, A_1= 8cm^2, A_2=1cm^2
- calculate F1, F2, F3

Exercises: vowel /i/







- l_1 = 9cm, l_2= 6cm, A_1= 8cm^2, A_2=1cm^2
- calculate F1, F2, F3

$$F_1 = \frac{c}{2\pi}\sqrt{\frac{a}{lV}} = \frac{340}{2\pi}\sqrt{\frac{0.00015}{0.06 \times 0.0008 \times 0.09}} = 319Hz$$

Helmholtz resonance

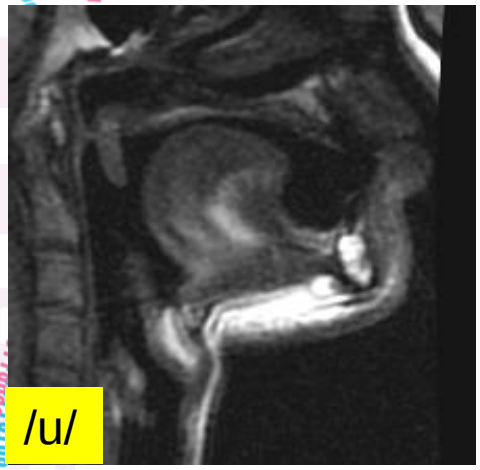
$$F_3 = \frac{340}{2 \times 0.06} = 2833Hz$$

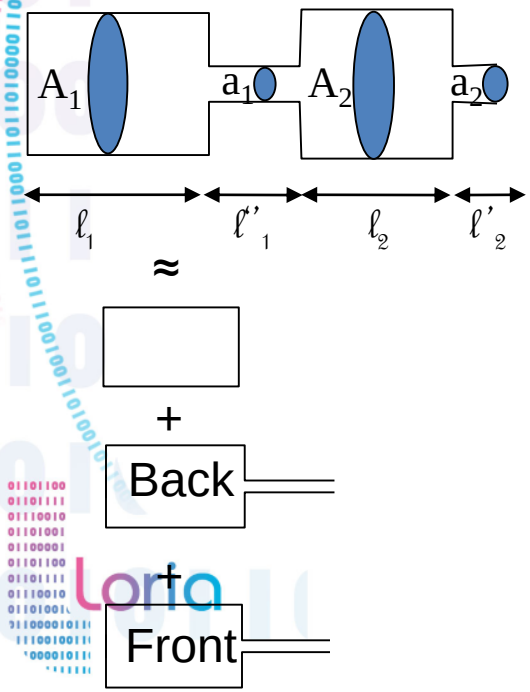
Half-wavelength front cavity

$$F_2 = \frac{340}{2 \times 0.09} = 1888Hz$$

Half-wavelength pharynx cavity

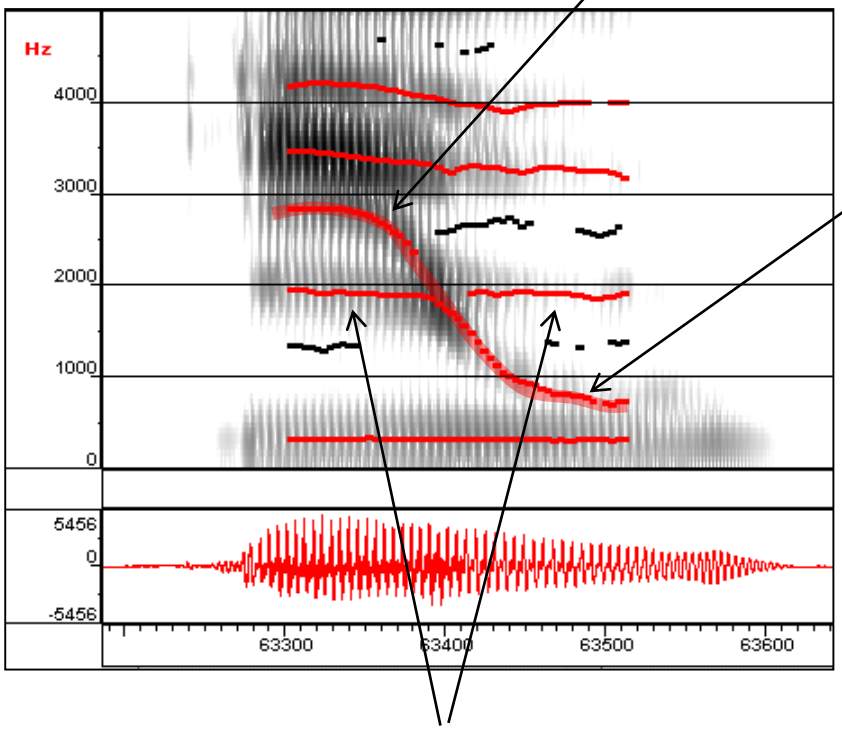
Vowel /u/





- From /i/ to /u/

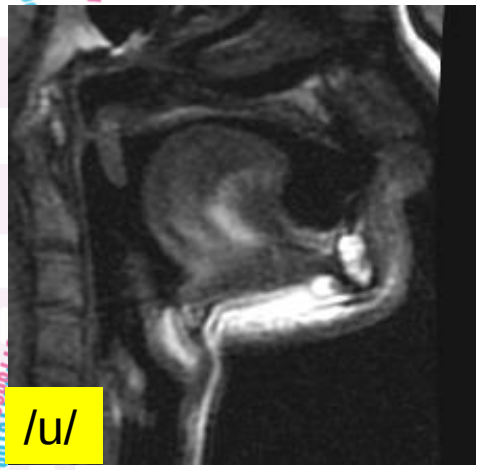
Half-wavelength mouth cavity (front cavity)

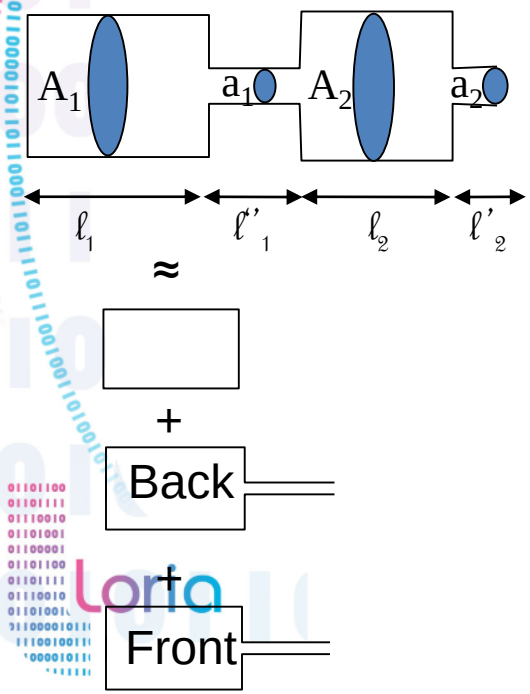


Mouth cavity by continuity, but low frequency resonance.
→ Helmholtz resonance of the front cavity.

Nothing to deduce from the continuity.

Vowel /u/





- Mouth cavity
  - Helmholtz resonator: A_2=7cm^2,l_2=5cm,l'_2=1,5cm,a_2=1cm^2 The frequency of the front cavity is 747 Hz.
- Pharynx cavity
  - Helmholtz resonator: A_1=8cm^2,l_1=8 cm,l'_1=3 cm,a_1=0.7cm^2 The Helmholtz frequency of the front cavity is 326 Hz.
  - Half-wavelength 2125 Hz