Selection and combination of acoustic descriptors for the discrimination between normal and pathological speakers

T. Dubuisson[1], T. Dutoit[1], B. Gosselin[1], M. Remacle[2]

[1]Faculté Polytechnique de Mons, [2]Université Catholique de Louvain

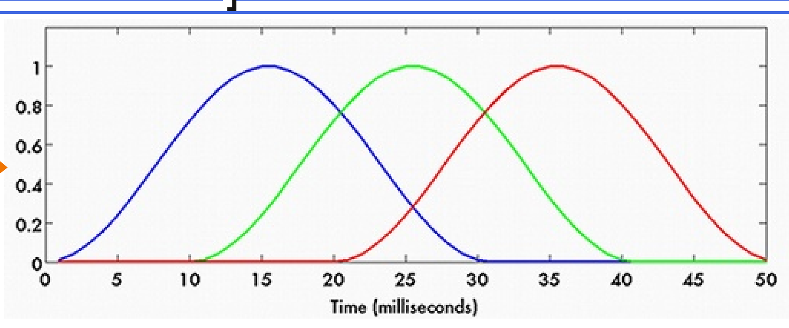
thomas.dubuisson@umons.ac.be

POLYTECH.MONS

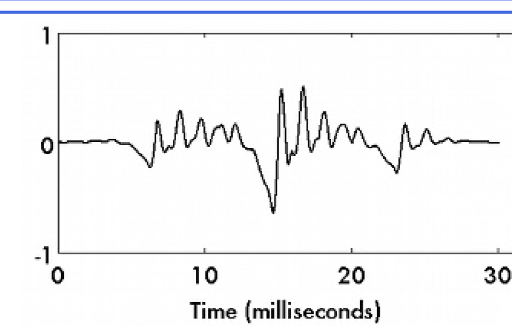
Context of Research

- Voice pathologies are assessed by clinician perceptually (highly dependent on the experience of the listener) and objectively (cumbersome and expensive equipment).
- Aims of the automatic analysis of voice pathologies.
- Drawbacks of existing methods.
- Aims of this study:
    1. Proposing descriptors NOT based on fundamental frequency.
    2. Combining descriptors from Music Information Retrieval (MIR) domain and from voice pathologies literature.
    3. Using the correlation between descriptors instead of a complex classifier.

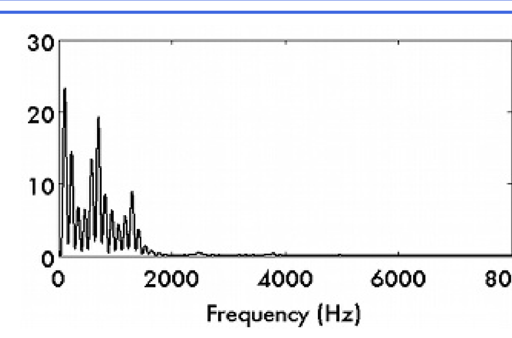
Principles of Computation

Signal →  → Frame → Temporal Domain ① / Spectral Domain ②

Windowing operation

① 

FTemporal frame of speech

② 

Fourier Transform of the frame of speech

Some Examples of Descriptors

Temporal Domain

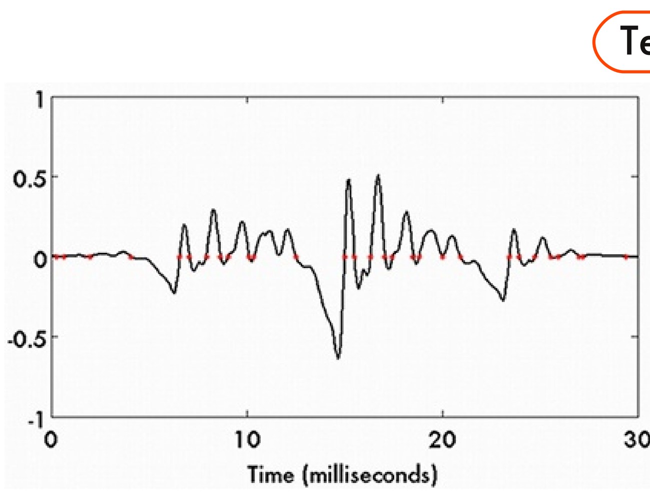


Illustration of the Zero Crossings Detection

$$E_T(dB) = 10 \times log_{10}(\sum_{i=1}^{N} x(n)^2) = 10.10dB$$

$$\mu_T = \frac{1}{N} \times \sum_{i=1}^{N} x(n) = 0$$

$$\sigma_T = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(n)-\mu_T)^2} = 0.1461$$

$$Zero\ Crossing\ Rate(Hz) = \frac{\#Zero\ Crossing}{0.03} = 1000Hz$$

Spectral Domain

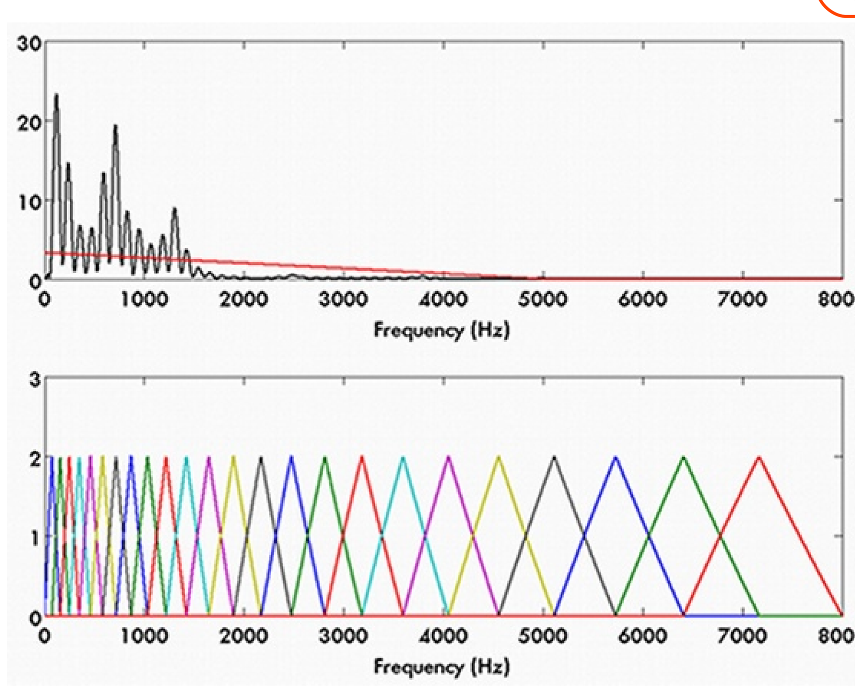


Illustration of the Spectral Slope computation and the MEL scale of frequencies

$$COG(Hz) = \frac{\sum_{f=1}^{8000} f \times X(f)}{\sum_{f=1}^{8000} X(f)} = 749Hz$$

$$Decrease = \frac{\sum_{f=2}^{8000} \frac{X(f)-X(1)}{f-1}}{\sum_{f=2}^{8000} X(f)} = 0.0028$$

$$\widehat{X}(f) = Slope \times f + K \ (Slope = -6.62 \times 10^{-4})$$

$$E_0 = \frac{\sum_{f=60}^{f=400} X(f)}{\sum_{f=60}^{f=8000} X(f)} = 0.339$$

$$E_2 = \frac{\sum_{f=2000}^{f=5000} X(f)}{\sum_{f=60}^{f=8000} X(f)} = 0.035$$

$$T_1 = \frac{\sum_{MEL_{[1]}} X(f)}{\sum_{MEL_{[1...24]}} X(f)} = 0.082$$

$$T_2 = \frac{\sum_{MEL_{[2,3,4]}} X(f)}{\sum_{MEL_{[1...24]}} X(f)} = 0.254$$

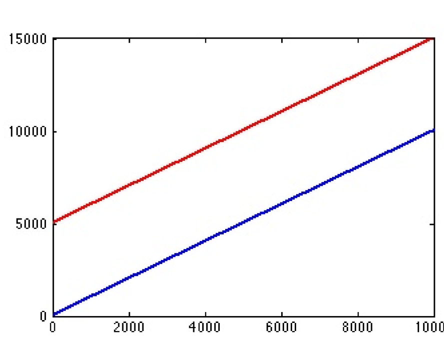
Analysis of speech pathologies

Database: Kay Elemetrics MEEI Database consisting on 53 normal and 657 pathological sustained vowels /a/ (sampling frequency: 16 kHz; quantification: 16 bits).
Aim: extracting information from speech signal for finding significant differences between normal speakers and pathological speakers.
Principle: Use of the correlation between 87 acoustic descriptors for discriminating normal and pathological voices.
Database is split into training set (randomly chosen 65% of the two classes) and test set (remaining 35%).

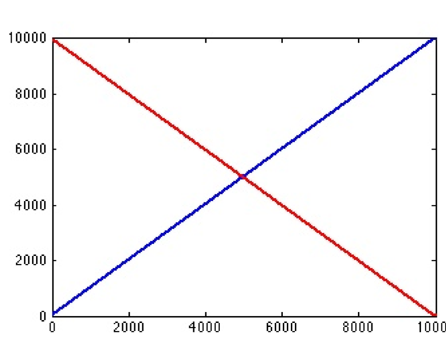
Computation of the Correlation Matrix

$$R_{xy} = \frac{\sum_{i=1}^{N}(x_i-\overline{x}) \times (y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\overline{x})^2} \times \sqrt{\sum_{i=1}^{N}(y_i-\overline{y})^2}}$$

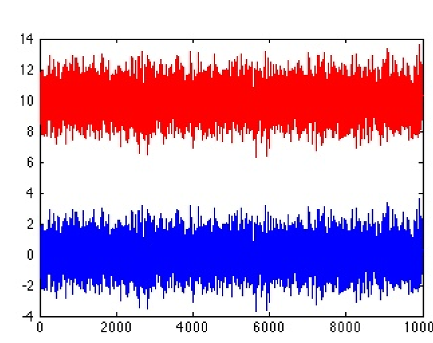
Pearson Coefficient of Correlation



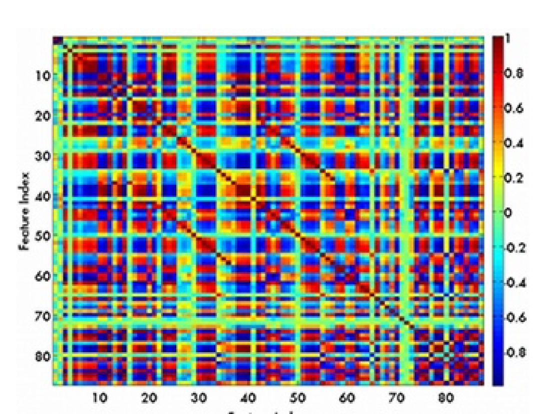
Correlation = 1



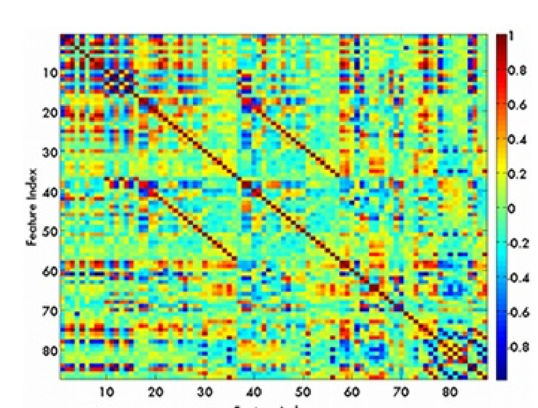
Correlation = - 1



Correlation = - 0.08



Correlation matrix for a normophonic voice

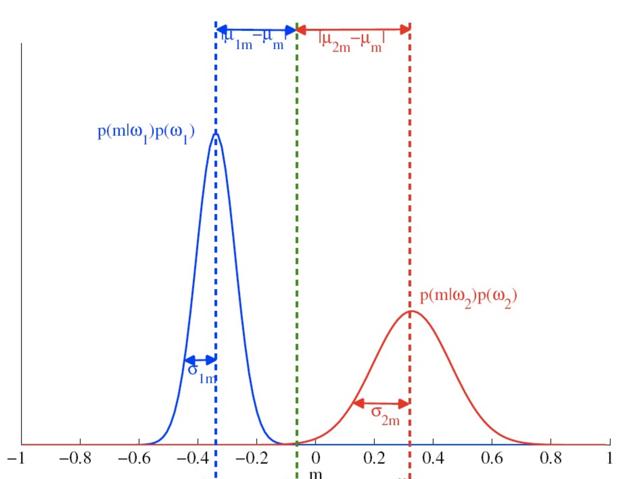


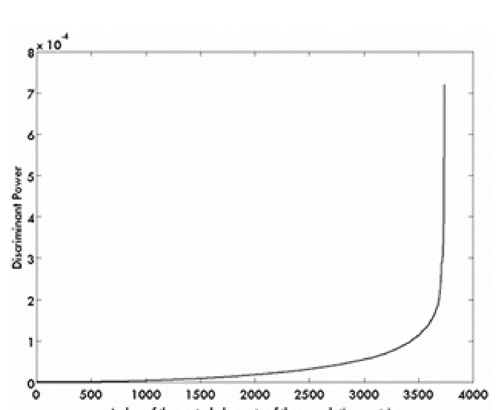
Correlation matrix for a pathological voice

Selection of the Most Discriminant Correlation

$$D_k = \frac{\sum_{c=1}^{C} p(\omega_c)(\mu_{ck}-\mu_k)^2}{\sum_{c=1}^{C} p(\omega_c)\sigma_{ck}^2}$$

Discriminant power by Fisher Analysis

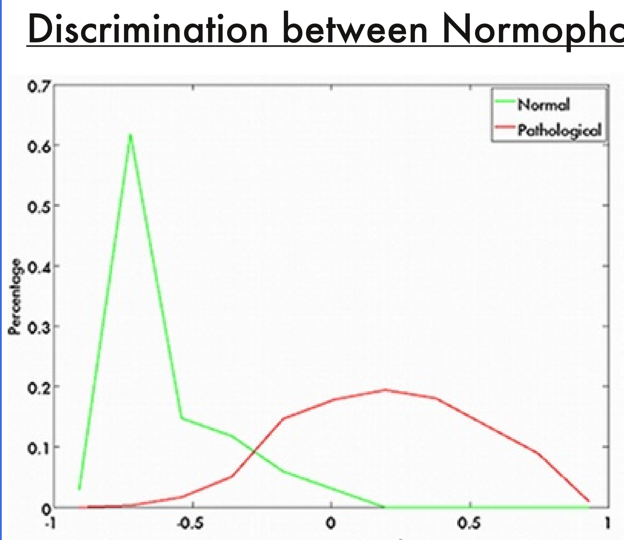




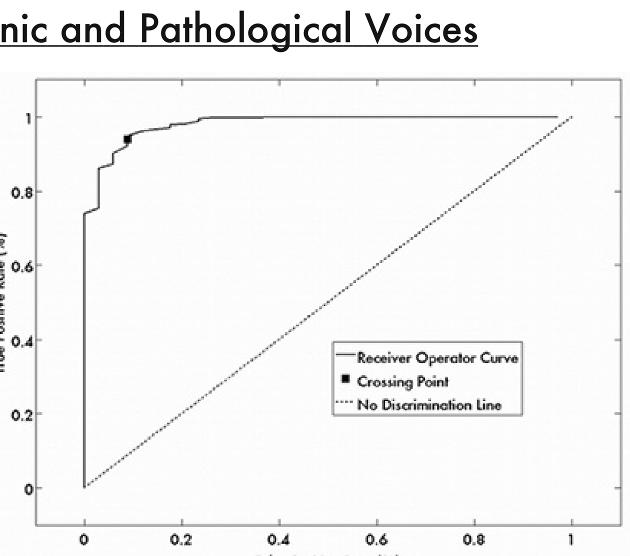
Discriminant power of the correlations (in ascending order)

The correlation between the specral decrease and the first tristimulus in Bark frequency bands is the most discriminant between the two populations.

Discrimination between Normophonic and Pathological Voices



Distribution of the most discriminant correlation for the normal and pathological voices (training set).



Receiver-Operator Curve for the discrimination between the normal and pathological voices (training set).

CONFUSION MATRIX FOR THE ONE CORRELATION CASE (TRAINING SET)

	Manual Pathological	Manual Normal
Auto Pathological	0.947	0.088
Auto Normal	0.053	0.912

ACCURACY FOR THE 10 PAIRS OF TRAINING AND TEST SETS (ONE CORRELATION CASE)

#	Training Set	Test Set
1	0.946	0.947
2	0.942	0.947
3	0.942	0.947
4	0.940	0.951
5	0.940	0.951
6	0.938	0.955
7	0.929	0.971
8	0.938	0.955
9	0.940	0.963
10	0.942	0.945

MEAN CONFUSION MATRIX FOR THE 10 TRAINING SETS (ONE CORRELATION CASE)

	Manual Pathological	Manual Normal
Auto Pathological	0.943	0.109
Auto Normal	0.057	0.891

MEAN CONFUSION MATRIX FOR THE 10 TEST SETS (ONE CORRELATION CASE)

	Manual Pathological	Manual Normal
Auto Pathological	0.955	0.074
Auto Normal	0.045	0.926

T. Dubuisson, T. Dutoit, B. Gosselin, M. Remacle, "On the Use of Correlation between Acoustic Descriptors for the Normal/Pathological Voices Discrimination", (accepted for publication in) EURASIP Journal on Advances in Signal Processing, Special Issue on Analysis and Signal Processing of Oesophageal and Pathological Voices.