FEATURES EXTRACTION

This part aims to extract features on the signals. Once extracted, these characteristics will be useful in observing the breathing sound differences between preterm neonates who need SRT, those who do not need it and those who have had it. Being blind to these data, my work was to gather enough features, so that some of them can make a distinction.

# Methods

The main characteristics identified are spectral. Some are directly taken from the power spectrum, and others are derived from spectrum fits. Finally, the MFCC coefficients as well as the LPC coefficients are calculated.

## Features on Power Spectrum

Some basic features based on the power spectrum were implemented at first. They are described below. Figures 1, 2, 3 and 4 illustrate these features.

**Mean (meanPSD) -** The average frequency calculated as the sum of product of the power spectrum and the frequency divided by the total sum of the power spectrum.

**Median (medPSD) -** The frequency which separates higher half of the spectral power from the lower half.

**Power bandwidth (bw) -** The difference between the upper frequency where the power is 3 dB lower maximum and the lower frequency where response is 3 dB lower.

**stdPSD -** The deviation of the spectrum frequencies from the mean frequency. It is calculated as follows:

\sqrt{\frac{\sum pxx\*(f-meanPSD)^2}{\sum pxx}}

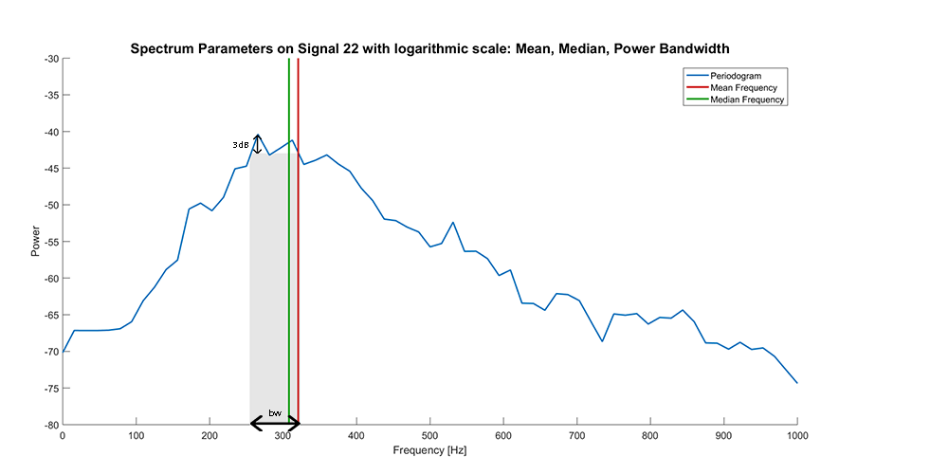


Figure 1: Frequency Mean and Median as well as Power Bandwidth in the Logarithmic Periodogram of Signal 22

**p25 -** The frequency below which a quarter of the spectral power lies.

**p75 -** The frequency below which three quarter of the spectral power lies.

**IQR -** Interquartile range, i.e. the frequency range between p25 and p75.

Mettre figure à faire sur autre ordi

**TP -** Total power in the 100-1000 Hz range.

**p100-200 -** Power in the 100-200 Hz range divided by TP.

**p200-400 -** Power in the 200-400 Hz range divided by TP.

**p400-800 -** Power in the 400-800 Hz range divided by TP.

**p800-1200 -** Power in the 800-1200 Hz range divided by TP.

­

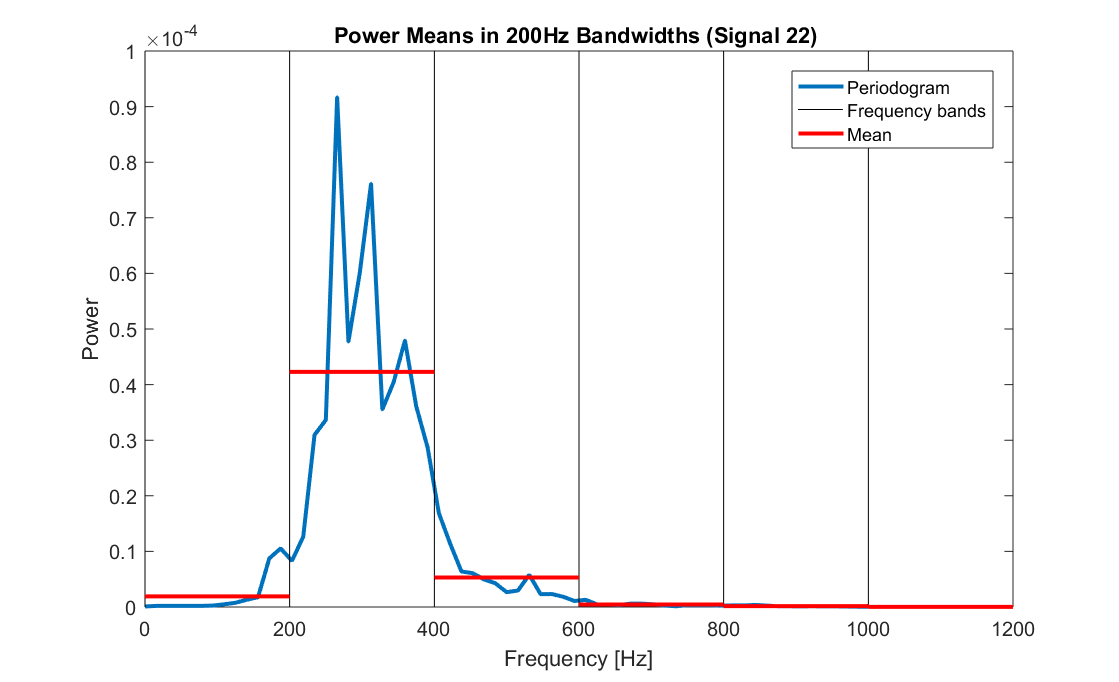


Figure 2: Power Means in 200Hz Bandwiths on Signal 22

**spectrum-slope -** The rate at which the sound spectrum power tails off or decreases from mean frequency to higher frequencies. The value represents the gradient of the linear regression line fitted to the power in logarithmic octave scale.

**r-square2 -** Statistical measure of how close the data is to the fitted regression line.

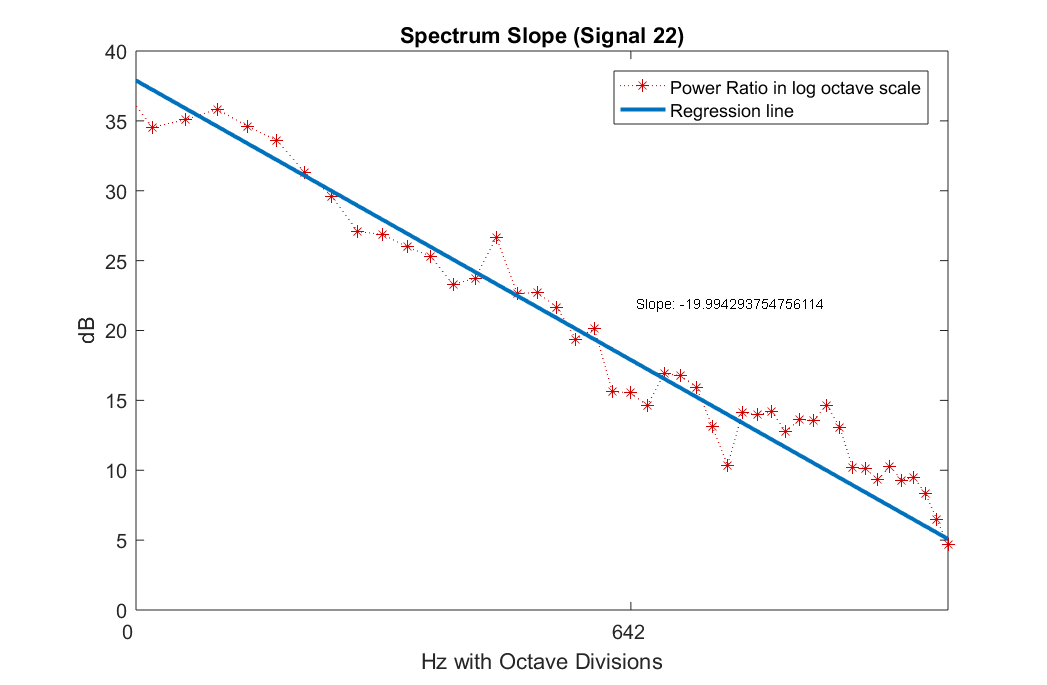


Figure 3: Power ratio and regression line on an octave scale

­

## Welch Periodogram Fit

In order to obtain further information on power spectral density, Welch periodogram models were generated. The filtering by moving average and the Gaussian Mixture model were used.

Sur ces modèles, il a été décidé de se concentrer sur des caractéristiques liées à leur forme, et notamment à leurs pics. Elles sont décrites ci-dessous :

**nb\_pks –** Nombre de peaks dans le modéle

**f\_higherPk–** Fréquence du maximum

**dif\_higherPks–** Différence relative en fréquence entre les 2 peaks les plus important

### MAF: Moving Average Filter

La moving average permet de lisser le periodogramme  en supprimant les [fluctuations](https://fr.wiktionary.org/wiki/fluctuation) [transitoires](https://fr.wiktionary.org/wiki/transitoire) de façon à en souligner les tendances globales.

Cette moyenne est dite *mobile* parce qu'elle est recalculée de façon continue, en utilisant à chaque calcul un sous-ensemble d'éléments dans lequel un nouvel élément remplace le plus ancien ou s'ajoute au sous-ensemble. Ici, le span est de 5. The calculation of elements is given below, with $x$ the starting signal and $y$ the smoothed signal.

$y(1) = x(1)$

$y(2) = \frac{x(1) + x(2) + x(3)}{3}$

y(n)=\frac{\sum\_{i=-\left\lfloor\dfrac{span}{2}\right\rfloor}^{\left\lfloor\dfrac{span}{2}\right\rfloor}{x(n+i)}}{span}\\

with $1 \leq n \leq N-\left\lfloor\dfrac{span}{2}\right\rfloor, N the number of samples

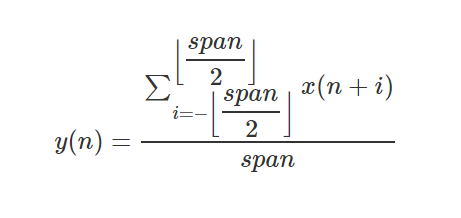


Figure … shows how the moving average fits the periodogram in Signal 22. The peaks were also represented as it is from them that the features are extracted (cf beginning of part 1).

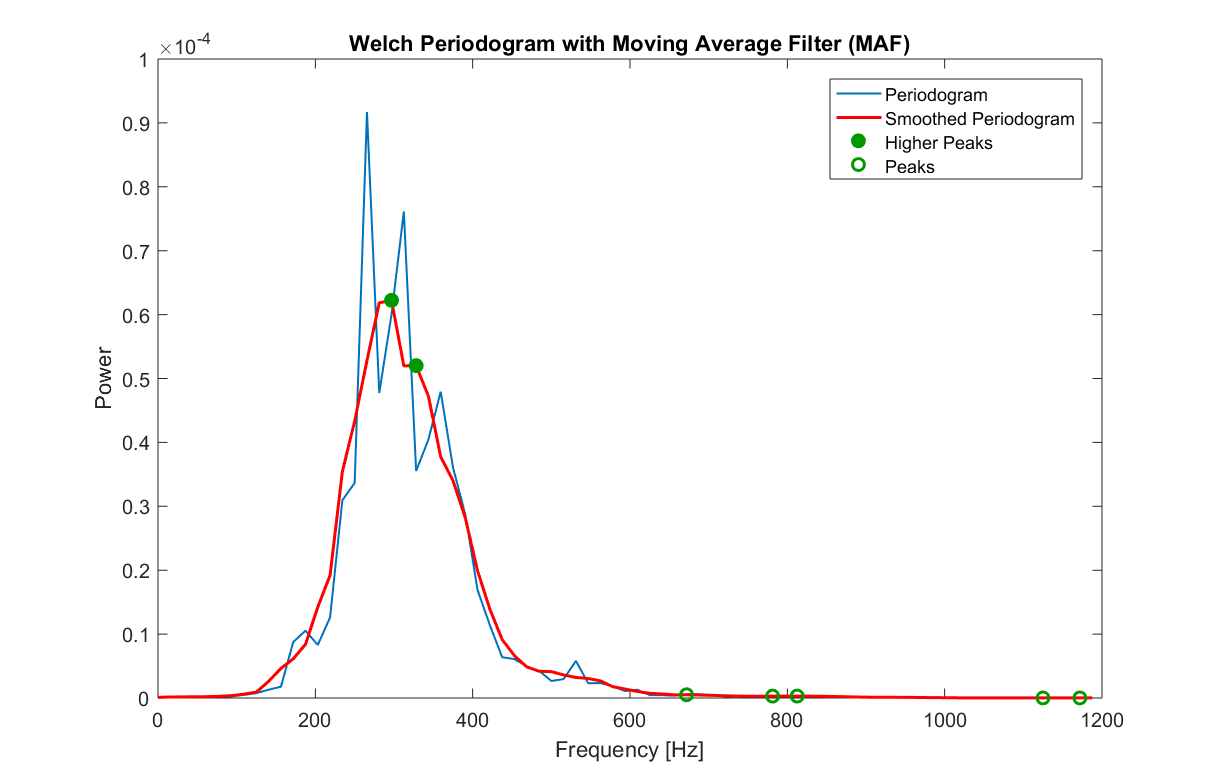


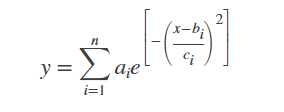
Figure 4: Welch Periodogram and Smoothed Periodogram by a Moving Average Filter on Signal 22

### Gaussian Mixture Model

Un Gaussian Mixture Model est souvent utilisé en statistique comme a [probabilistic model](https://en.wikipedia.org/wiki/Probabilistic_model) for representing the presence of [subpopulations](https://en.wikipedia.org/wiki/Subpopulation) within an overall population. Il est ici utile pour l’extraction de caractéristiques. En effet, en plus des pics, les paramètres de chaque gaussiennes vont être extraites pour une analyse future, ce qui donnera de nouvelles informations sur le périodogramme.

Les periodogrammes des différents signaux sont souvent composés de pics principaux entre 100 et 500 Hertz, ainsi que de quelques pics secondaires. Le Gaussian Mixture Model a été choisi pour fits 4 peaks, correspondant à l’addition de 4 gaussiennes.

The Gaussian model is given by



where $a$ is the amplitude, $b$ is the centroid, $c$ is related to the peak width and $n=4$ is the number of peaks to fit.

La figure … ci-dessous represente chacune des gaussiennes, avec également le somme de toutes ces gaussiennes (en rouge) et le periodogramme. Commenter figure

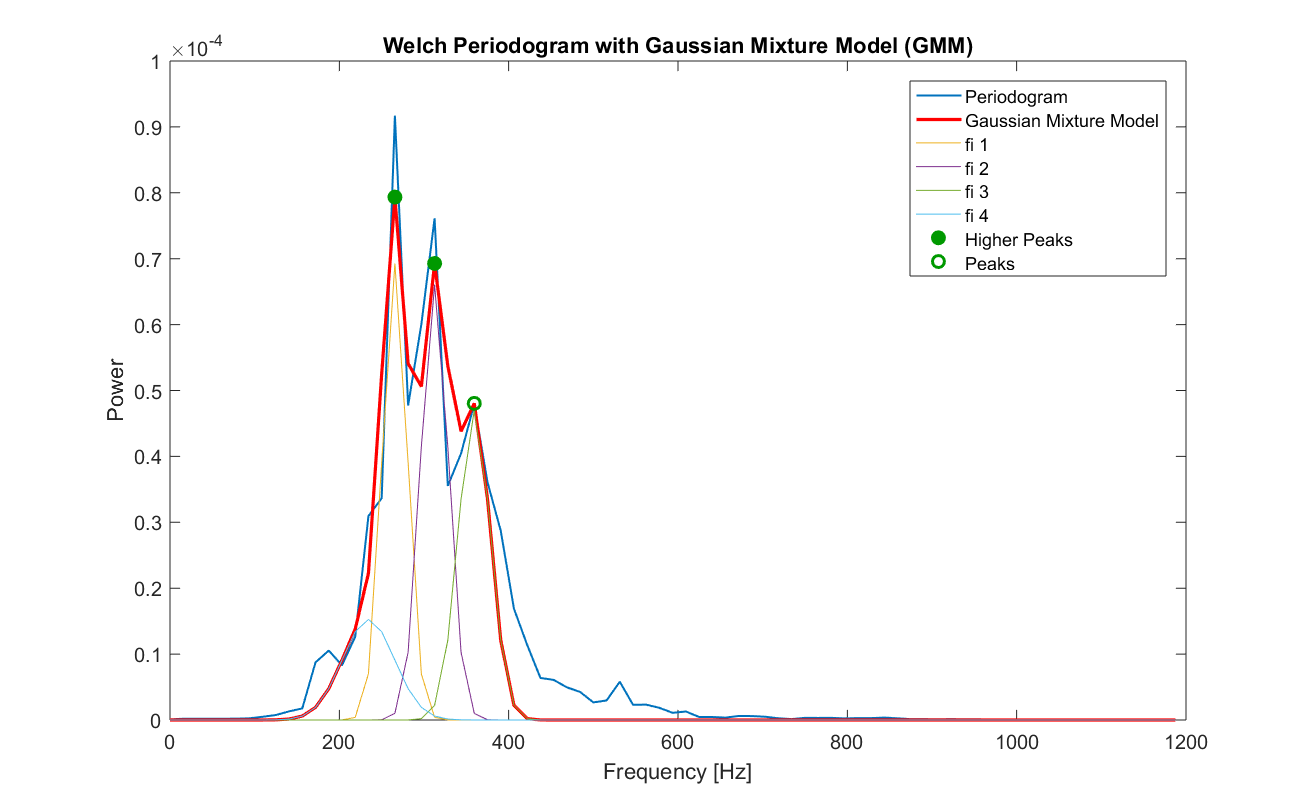


Figure 5: Welch Periodogram and Gaussian Mixture Model on Signal 22

Les paramètres a, b et c de chaque gaussienne ont été gardé afin de voir si une différence été remarquable suivant les bébés et leur traitement.

## MFCC Coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are a feature widely used in automatic speech and speaker recognition. They aim at identifying the linguistic content and discarding all the other noises. Even though the signals being studied are not words but respiratory sounds, these coefficients are state of the art in audio signal features and could be interesting to differentiate the babies who had an SRT from the others. The calculation of MFCC coefficients is performed in several steps[[1]](#endnote-1), described below (figure .. and …).

Figure 1 illustrates the process performed to generate the MFCC coefficients of each signal. It is composed of the following five stages:

1. The temporal signal ${s(n)} is framed into short window ${s\_w(n)}. The frame size is usually between 20 and 40ms, a duration for which the stationarity is assumed. The frame size chosen is ${f\_n\*0.03} with an overlap of ${f\_n\*0.02}, where ${f\_n} is the sampling frequency.
2. In each frame, MFCC coefficients are computed. A detailed explanation can be found in Figure 2.
3. Only the first 6 coefficients are kept. EXPLIQUER POURQUOI on a utilisé les 14 premiers, voir avec matlab les filtres qui correspondent aux fréquences. On a pris les 14 premiers coefficients, qui correspondent aux 14 premiers filtres. Les bandes passanets de chaque filtres se trouve en annexe.
4. To avoid frequency peaks associated with discontinuities present in the signal without crying, the MFCCs were extracted from the complete record ${s(n)}. Once the 6 coefficients of each frame have been recovered, those corresponding to crying frames were removed.
5. The coefficients of all frames are then averaged to obtain the mean first 6 MFCCs of the complete signal without crying.

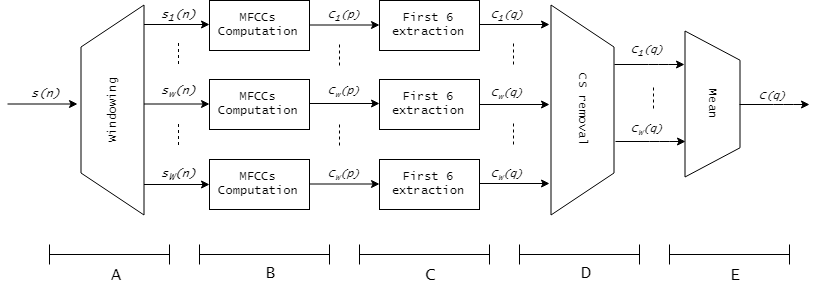
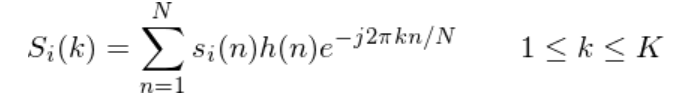


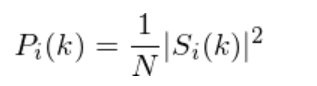
Figure 6: General diagram of MFCC coefficients generation

In figure 1,  $1 \leq n \leq N$ with N the number of samples, $1 \leq w \leq W$ with W the number of frames, $1 \leq p \leq P$ with P the number of MFCCs, $1 \leq q \leq Q$ with $Q=6$ the number of MFCCs kept.

Figure 2 illustrates the MFCC coefficients computation, with a construction based on the perception of sounds by humans. It is composed of the following six stages.

1. On each frame, the complex DFT ${S\_w(f)} (cf equation …) and the periodogram estimate of the power spectrum ${P\_w(f)} are computed (cf equation …). It enables to identify which frequencies are present in each frame, as the human cochlea (an organ in the ear).





Changer les lettres pour les 2 équations!

1. A mel-spaced filterbank is applied to the periodogram ${P\_w(f)}. This is a set of triangular filters which are wider as the frequencies get higher. Multiplying each filter with the periodogram enables to obtain as much periodogram sections as filters, which are wider in high frequencies. By analogy with humans, the cochlea cannot discern the difference between two closely spaced frequencies, especially when the frequencies increase.
2. The coefficients obtained in each periodogram section are summed to get the energy.
3. The logarithm function is applied to these energies. This is also motivated by human hearing: the perceived volume of sound is not linear with the energy it contains.
4. The Discrete Cosine Transform (DCT) of each log energy gives one MFCC coefficient.
5. There are as many coefficients as filters, but only the first 12 are kept. The higher DCT coefficients represent fast changes in the filterbank energies and deteriorate the recognition.



Figure 7: Diagram of MFCCs computation process

In figure 2, $1 \leq n \leq N$ with N the number of samples, $1 \leq w \leq W$ with W the number of frames, $1\leq f \leq F$ with F the maximum frequency and $1\leq j \leq J$ with J the number of filters.

## LPC Coefficients

A voir. Discoutinuité peut poser des porblemes mais pas très graves. Il faut sinon le coder à la main mais pas le temps.

Certains LPC coeff ne servent à rien (les premiers car pas dan 100 1200)

Linear Predictive Coding with its LPCs (Linear Predictive Coefficients) are a tool used mostly in audio signal processing and speech processing for representing the signal spectral envelope with a reduced number of parameters.

Ces coefficients sont intéressants dans le projet car il donne des informations sur le signal, notamment sur les formants et …. C’est pour cette raison qu’il a été choisi de les implémenter.

**Theorical calculations**

They are based on a prediction of the signal value at time t approximated with linear combination of real signal values in previous moments, which can be expressed as follows (eq. 1):



where t is a discrete time moment, ft is the original signal, ftprime an approximation of the original signal, ak the LPCs 1≤k≤p and p the number of LPCs.

They are calculated so that they minimize the error between the real signal and the one calculated using LPCs over the interval of interest to minimize to mean squared error as defined in (1.3 ) .



where ts is the starting point in time of the interval for which the error is being calculated, te the ending and E the measured error on the interval t∈[ts; te].

To minimize this error, the ak are calculated in such a way that the derivative of E is equal to 0 (…).

avec =0.

Finally, the expression of the LPCs are as follow (…):

METTRE L’EXPRESSION

More details one the calculation of the coefficients can be found on the paper[[2]](#endnote-2).

**Implementation**

L’implémentation se fait simplement avec la fonction lpc sous Matlab. Le choix de 6 coefficients a été retenu car …

Mettre la représentation du signal avec LPC et sans

# Results

Partie résultats compliqué car j’étais aveugle sur les labels des signaux pendant toute la durée de mon stage. Je n’ai donc pas pu voir quelles caractéristiques différenciait au mieux les SRT de non SRT. Aucune figure n’a donc pu être effectué pour montrer les différences ou similitudes.

Apres la fin de mon stage, Arabella a fait une étude statitique, et voici les résultats.

Dire quelles feautures marchent le mieux (time or spectral)

# Discussion

Discussion d’un point de vue ingénieur. Dans le future, d’autres caractéristiques pourraient être implémentées, notamment dans les domaine spectro-temporal. Le papier … a par exemple aussi regarder les Spectral shape (scales estimate how broad or narrow the spectral profile is) and the temporal modulations (rates captures how fast and slow the frequency contents change with time and in which phase).

La qualité des recordings, du aux measures de precautions sanitaires qui devrait être amélioré d’apres le paper …[[3]](#endnote-3) : 'lb obtain similar lung sound amplitude for the infants, we recorded for as long as necessary until the infant spontaneously sighed or otherwise took deep breaths that resulted in clearly audible lung sounds. This occurred frequently in the term infants but often after 10 or more minutes in the premature infants.

Discussion médicale.

# Annex – MFCC filters

<https://www.mathworks.com/help/audio/ref/mfcc.html>

|  |  |
| --- | --- |
| **Filters** | **Passband Edges (Hz)** |
| Filter 1 | [133 267] |
| Filter 1 | [200 333] |
| Filter 1 | [267 400] |
| Filter 1 | [333 467] |
| Filter 1 | [400 533] |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
| Filter 1 |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# REFERENCES

1. <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>

   <http://recherche.ircam.fr/anasyn/peeters/pub/cours/20171031_Peeters_20172018_ATIAM_Cours_Structure.pdf> [↑](#endnote-ref-1)
2. <http://www.ivoronline.com/Science/Signals/LPC%20-%20Linear%20Predictive%20Coefficients/LPC%20-%20Linear%20Predictive%20Coefficients.pdf> [↑](#endnote-ref-2)
3. Comparison of the Lung Sound Frequency Spectra of Infants and Adults.Jamshed F., et. al.1986 Pediatr Pulmonol 1986; 2:292-295 [↑](#endnote-ref-3)