

## TP Biométrie Reconnaissance du locuteur

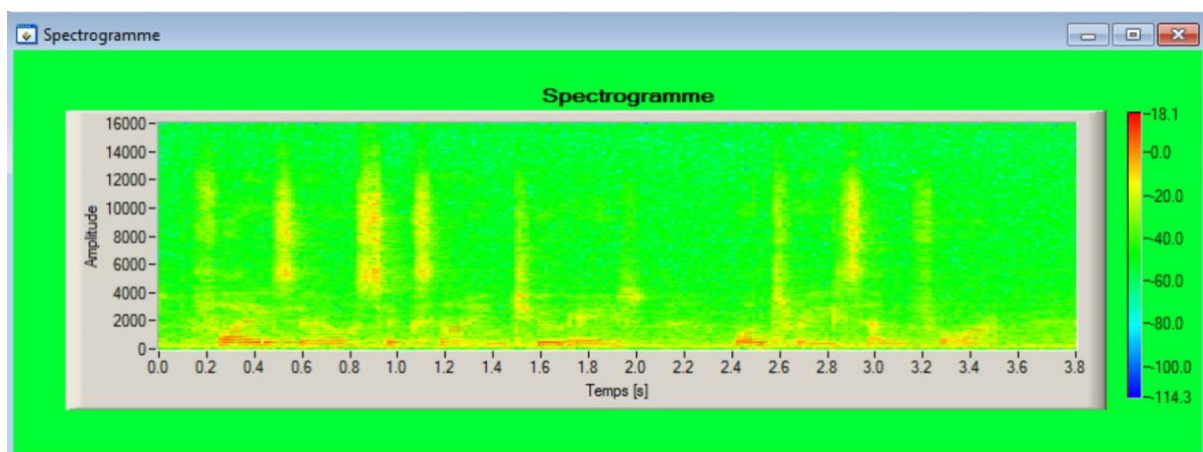
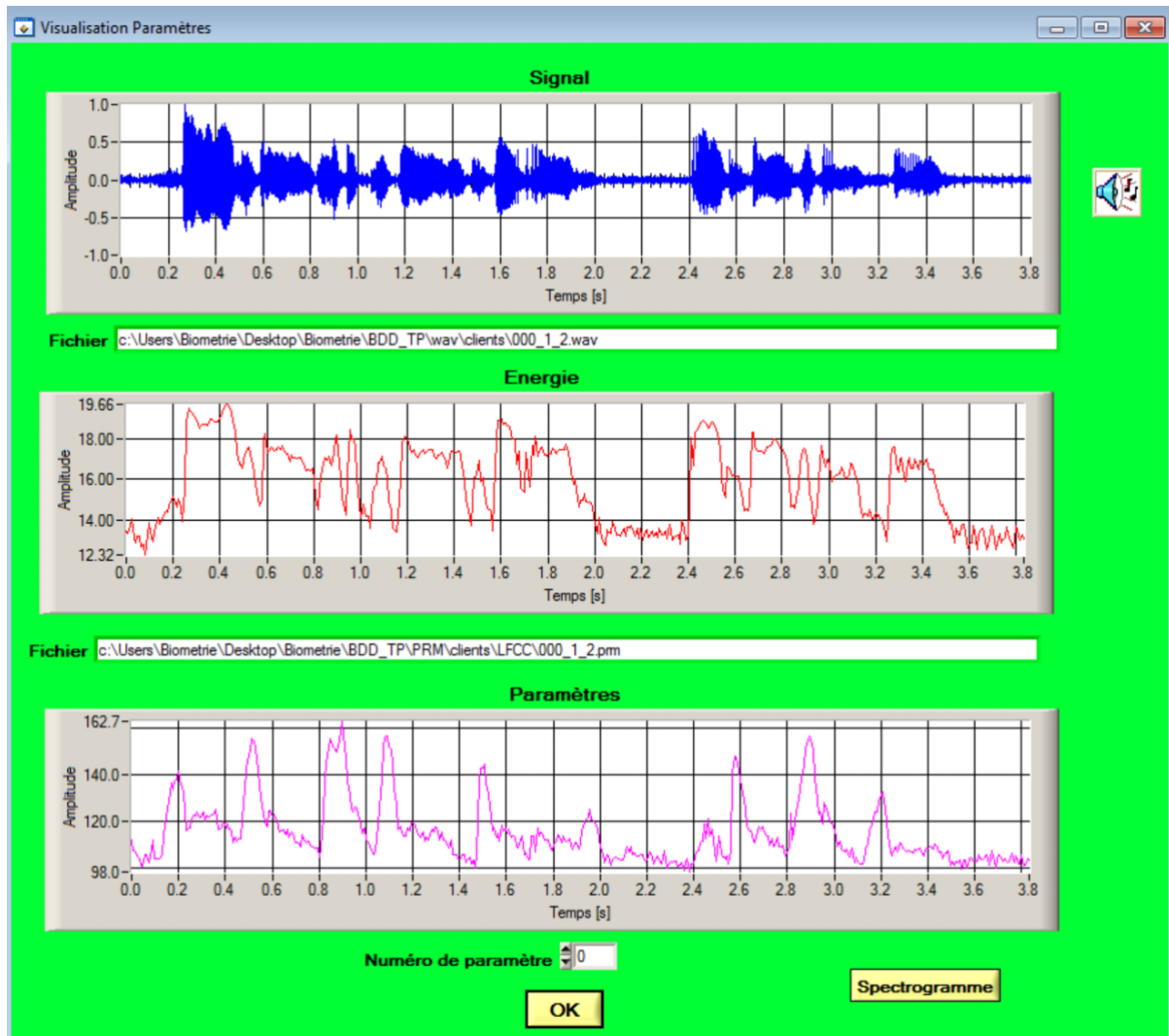
You, ZUO

École Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise

Date: 02/05/2020

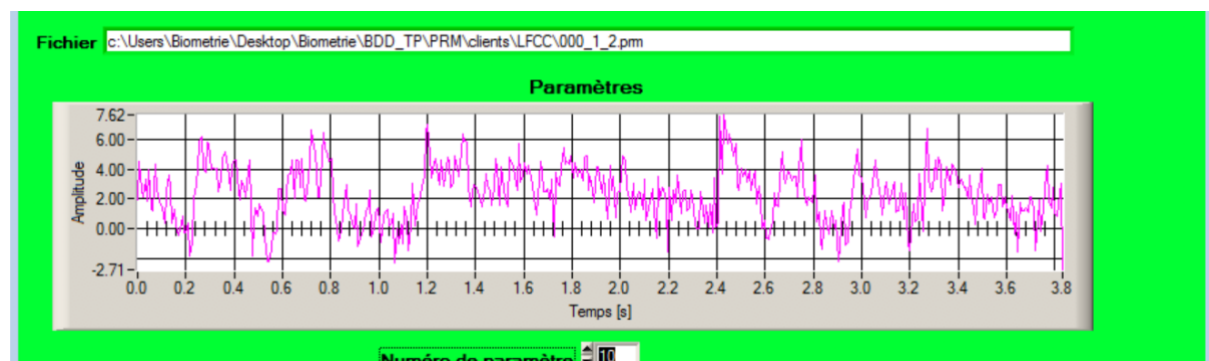
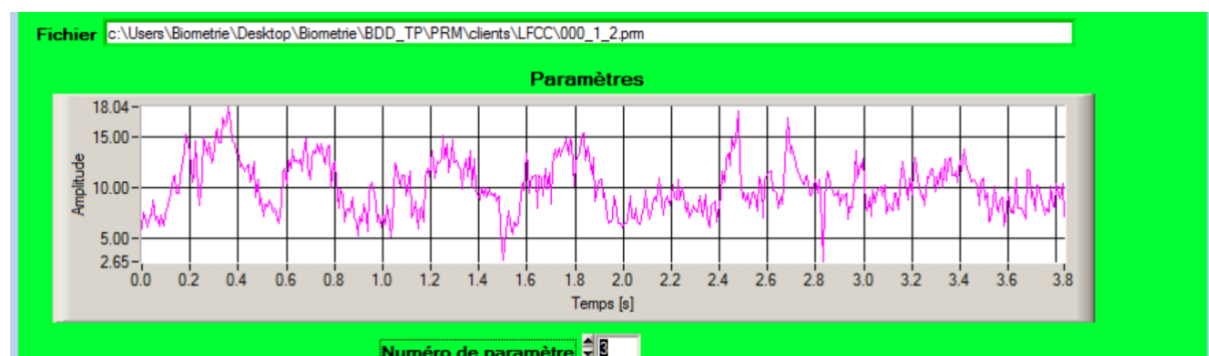
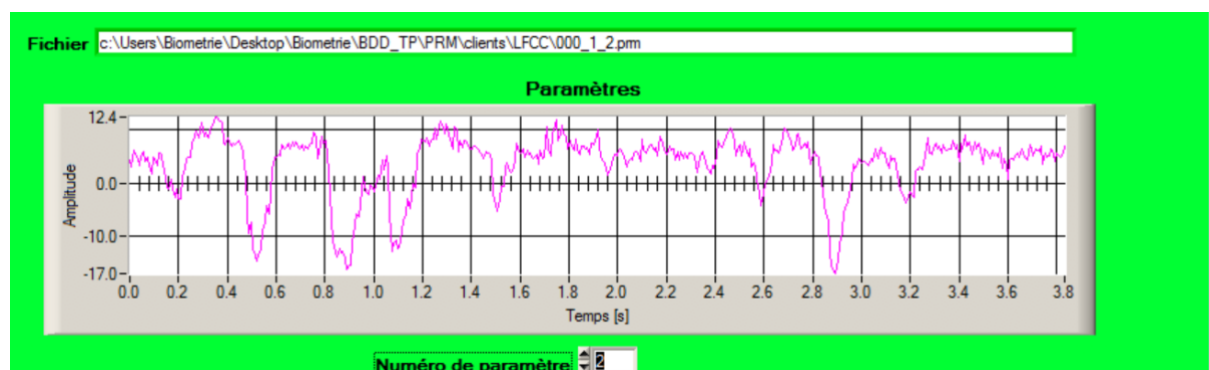
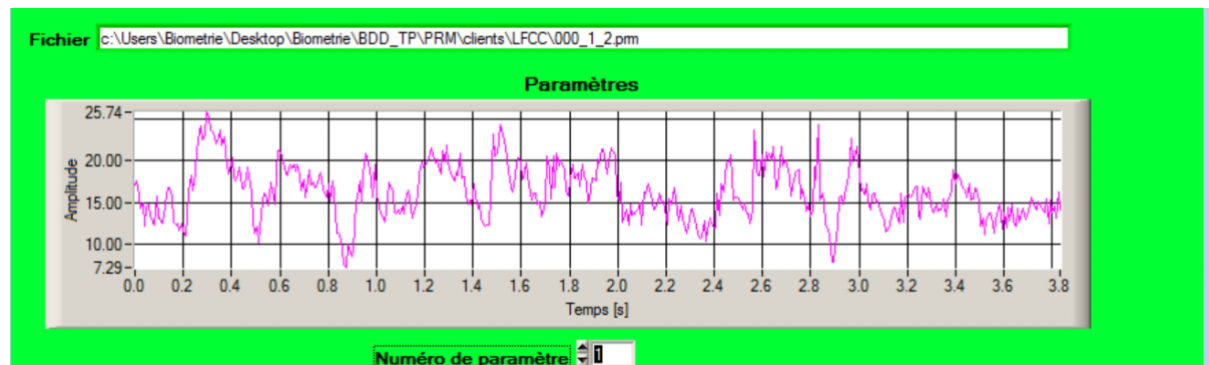
## I. Visualisation paramètres

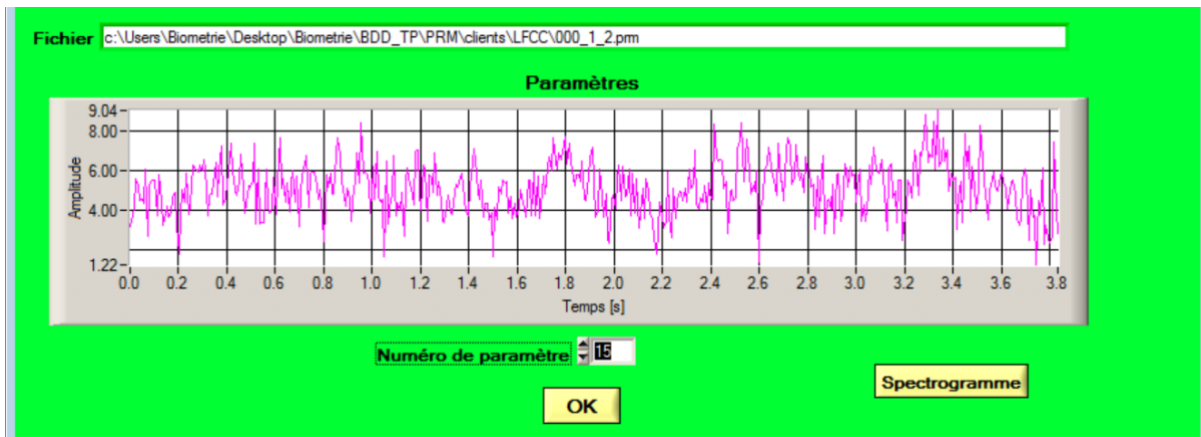
On a sélectionné d'abord le record du 1er client pour le phrase de la série de chiffres «5069281374» :



La figure ci-dessus nous montre que : lorsque l'amplitude du signal augmente ou diminue, la courbe d'énergie montre la même tendance. Dans le spectrogramme, nous pouvons voir que les parties en bas avec une petite amplitude ont toujours une luminosité plus élevée, et en outre, chaque syllabe (ou la prononciation du chiffre) correspond à une colonne de collecte d'énergie.

Ensuite on va analyser les courbes correspondant à différents paramètres :



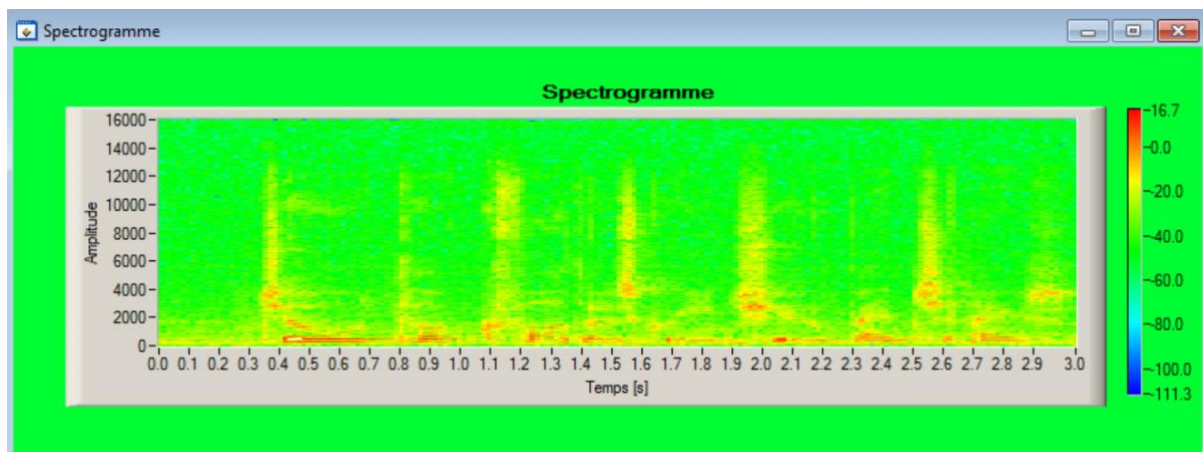
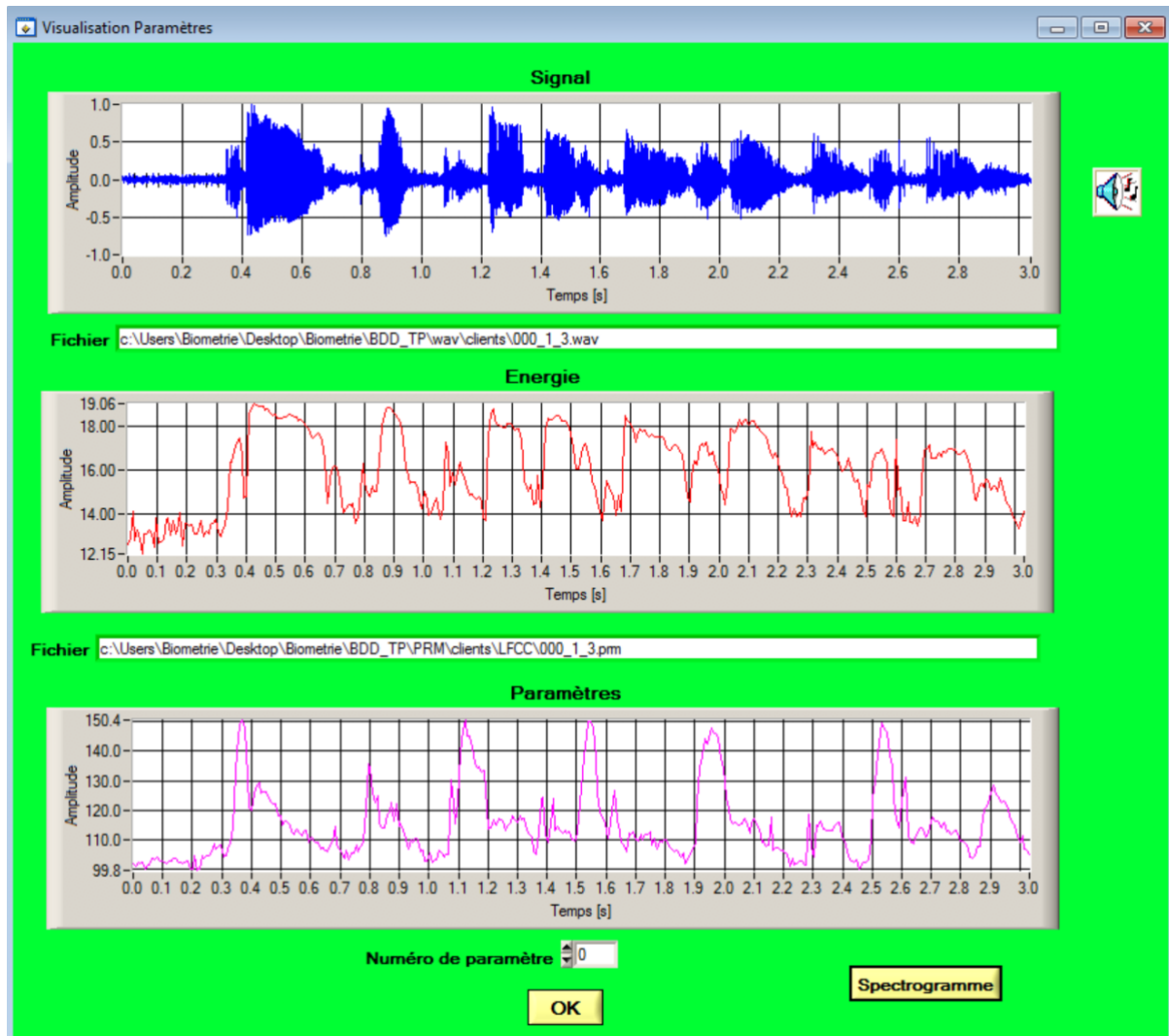


J'ai arrangé les paramètres de petit à grand, et j'ai trouvé que la fréquence de l'onde augmente de plus en plus, et la forme de l'onde devient de plus en plus compliquée. Mais nous pouvons encore voir qu'il existe un certain niveau de convergence entre l'amplitude de l'onde et l'amplitude du signal d'origine.

De plus, l'amplitude correspondant au premier paramètre est la plus grande, et l'amplitude du paramètre suivant sera considérablement réduite, et l'amplitude diminuera progressivement dans un ordre alternatif (devient d'abord plus grande, puis plus petite, puis plus grande, puis plus petite).

La raison pour ça est que les paramètres LFCC se calculent de la même manière que les MFCC, mais avec la seule différence que les filtres sont uniformes et non pas sur une échelle de Mel. Donc les premiers paramètres correspondent aux signaux basse fréquence, tandis que les derniers paramètres correspondent aux signaux haute fréquence.

Ensuite, pour la phrase « Joe took father's green shoe bench out » de la même personne et dans la même condition d'enregistrement nous avons les résultats comme suivants :

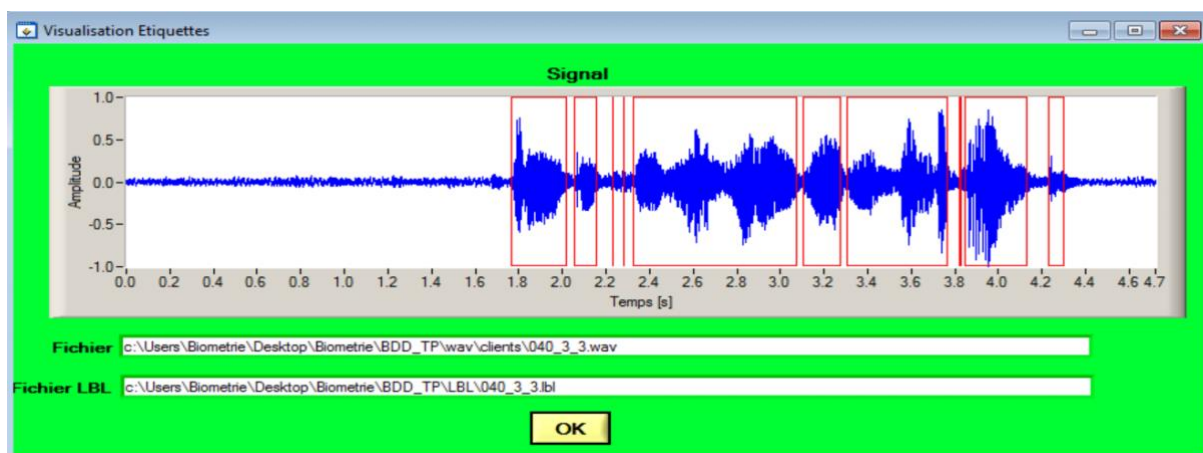
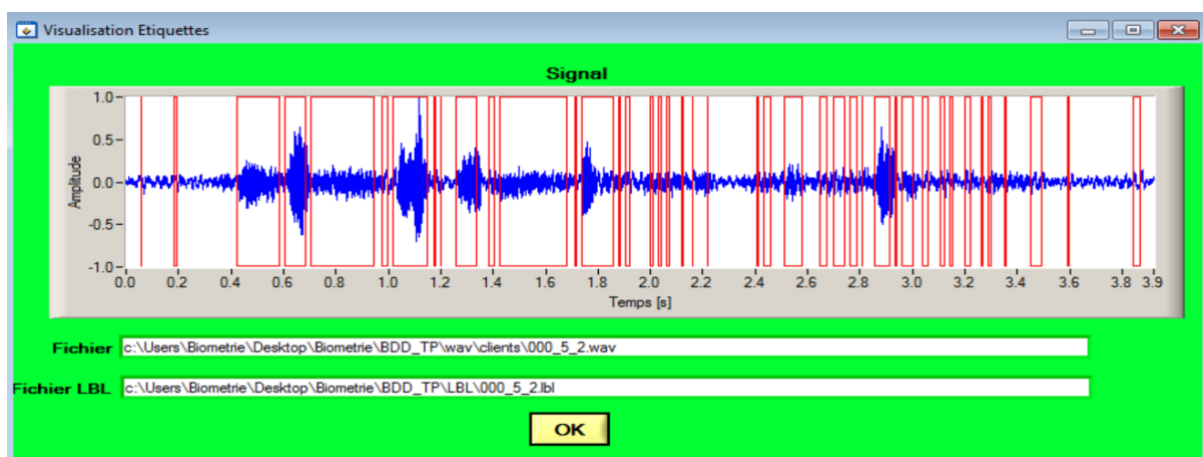
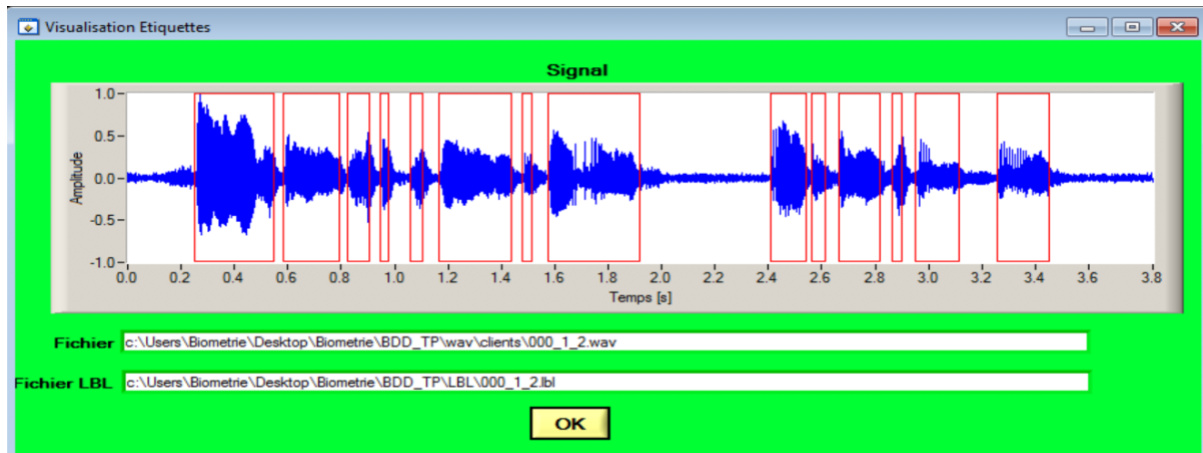


Pour les résultats de la phrase, j'ai trouvé que la partie de faible amplitude du spectrogramme est plus continue et dense que la partie de faible amplitude du spectrogramme de la phrase de chiffres précédent, et le temps de collecte d'énergie correspondant aux hautes fréquences est

relativement plus long. De même, plus le nombre de paramètres est grand, plus la fréquence de courbe correspondante est grande et complexe.

## II. Visualisation étiquettes

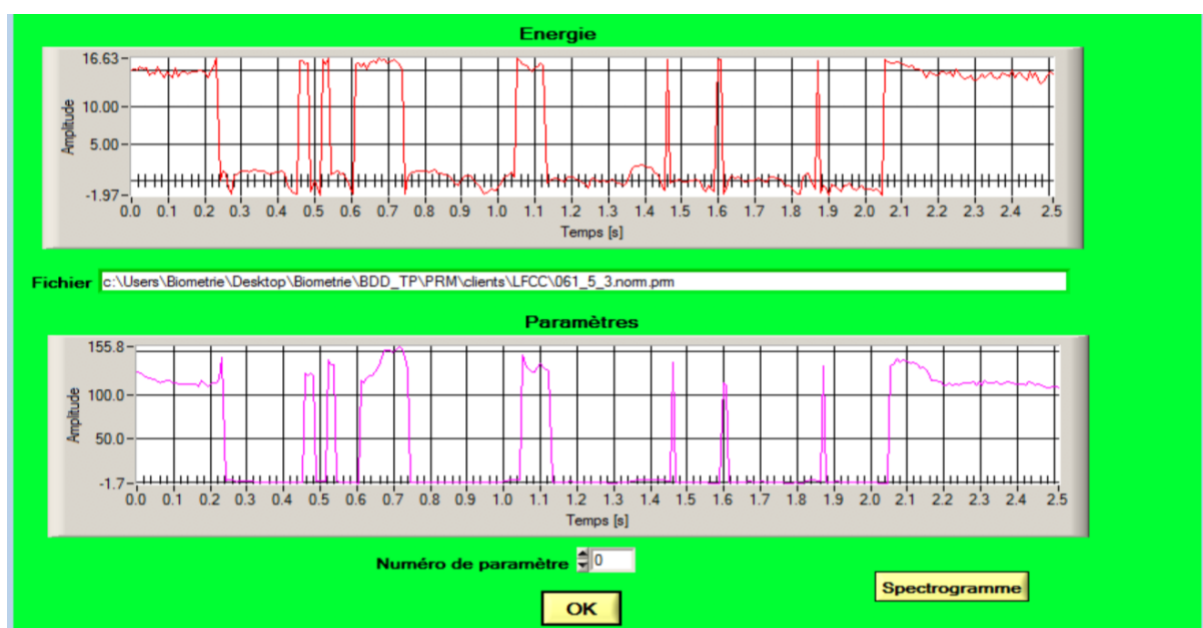
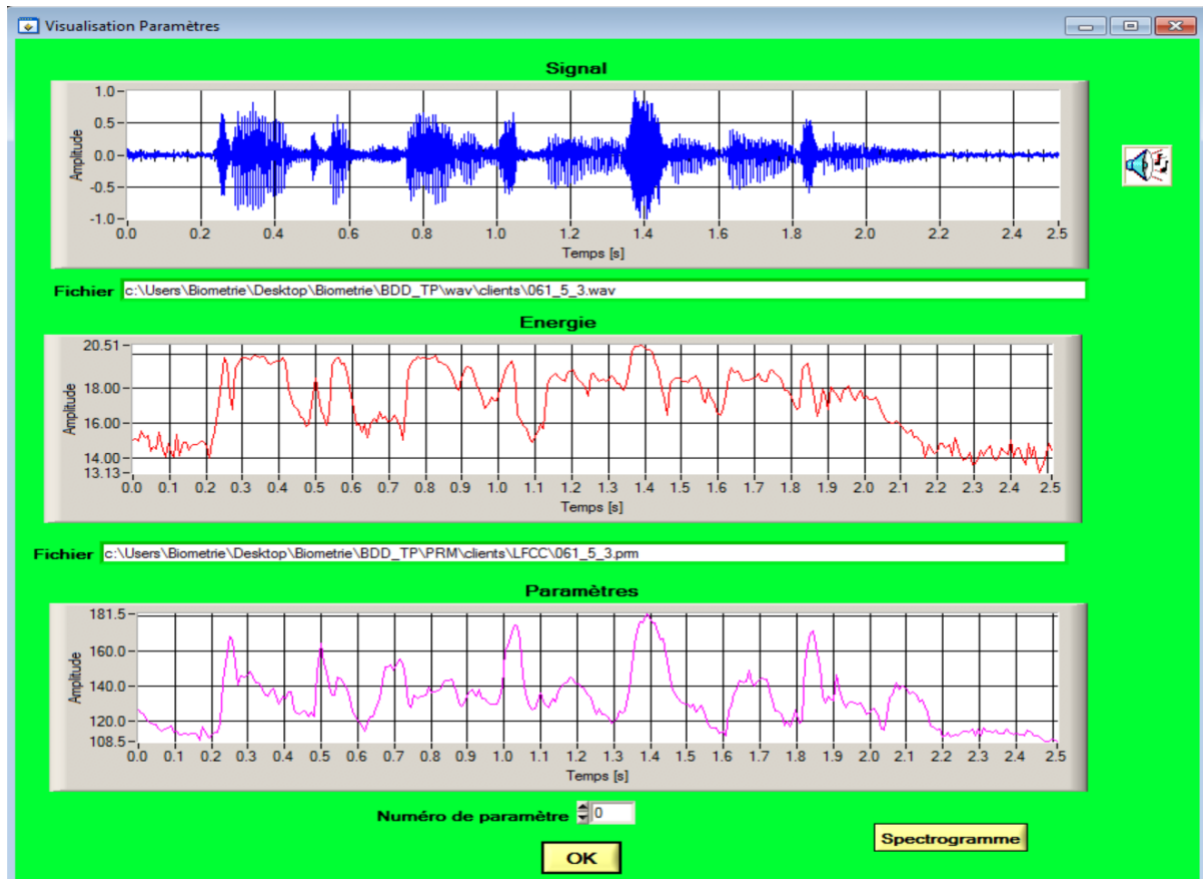
Cette étape vise à détecter les périodes de silence dans les fichiers. J'ai sélectionné aléatoirement trois fichiers d'étiquette :



Les fenêtres indiquent les zones avec signal, c'est-à-dire que les fenêtres ont filtré les signaux avec une petite amplitude, ce qui sont les moments de silence. Et ces fenêtres esclaves sont séparées par ces moments de silence.

### III. Normalisation des paramètres

Cette étape effectue une normalisation des paramètres acoustiques.





La courbe d'énergie normalisée montre une tendance opposée à la courbe d'énergie d'origine, et sa valeur est également polarisée (soit proche de 0 soit proche de la valeur maximale), et sa valeur maximale est probablement la position moyenne de la courbe d'énergie d'origine.

J'ai aussi remarqué que la valeur minimale du premier paramètre est dans un état stable de façon discontinue. Par conséquent, le point central (la moyenne) de normalisation est déterminé en fonction du premier paramètre.

#### IV. Analyse du modèle du monde

Il existe deux modèles de mélange gaussiens dans notre dossier de modèles, l'un composé de 16 distributions gaussiennes et l'autre composé de 128 distributions gaussiennes.

Parmi les modèles à 16 distributions gaussiennes, le plus représentatif est le 12ème ( $i = 11$ ), car il a le poids le plus élevé ( $\text{weight} = 0,21807601490352108$ ), le deuxième plus représentatif est le 11ème ( $i = 10$ ), et le poids  $\text{weight} = 0,15089360981325356$ . En d'autres termes, environ 35% des échantillons appartiennent à ces deux distributions gaussiennes.

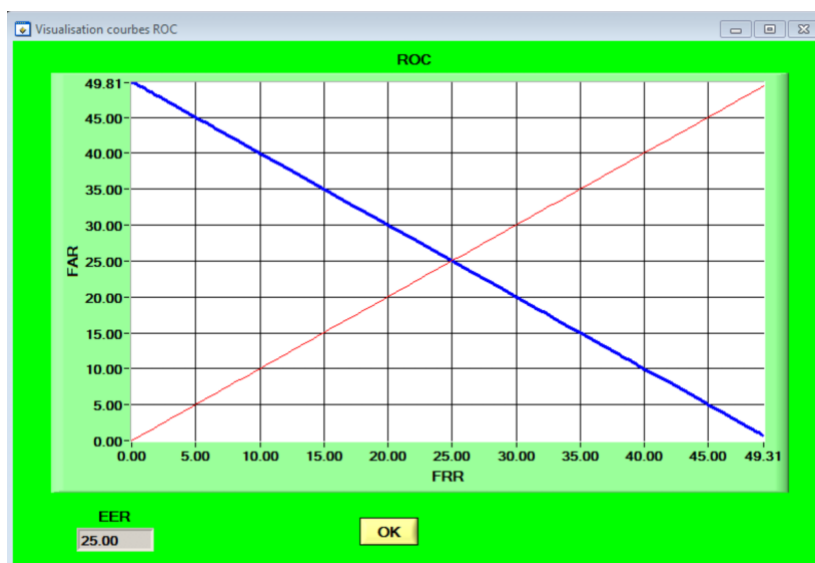
Pour le modèle de mélange gaussien avec 128 distributions gaussiennes, la plus significative est la 47ème distribution ( $i = 46$ ), avec un poids  $\text{weight} = 0.10338898250222654$ .

#### V. Analyse du modèle de clients

En comparant le modèle du client avec le modèle du monde, pour la distribution de même rang, nous constatons que les poids et les covariances de leurs distributions gaussiennes sont complètement égaux. Bien que les moyennes comportent une certaine erreur, elles sont globalement très proches.

#### VI. Système indépendant du texte

Il convient de noter que la courbe ROC est en fait une courbe convexe, mais à cause de l'échelle de représentation automatique, notre courbe semble une droite.



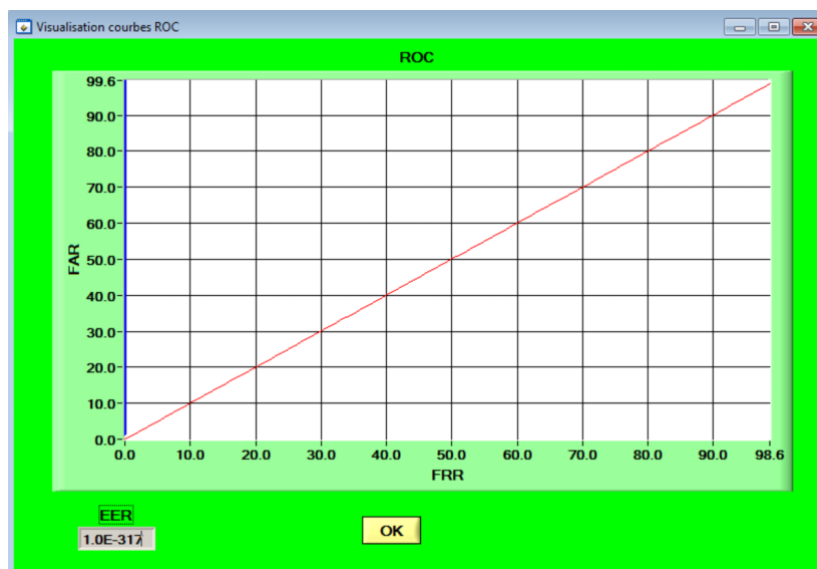


Nous pouvons voir sur la figure ci-dessus que les FAR et les FRR montrent des tendances de croissance opposées, et l'augmentation de l'un d'entre eux entraînera la diminution de l'autre. Mais globalement, nous voulons que ces deux taux d'erreur soient plus petits si possible, donc nous utilisons la technique de EER, qui est la valeur lorsque FAR et FRR sont égaux. Ici nous avons  $EER = 25.00\%$ .

Si nous avons un FAR de 10%, dans les mêmes conditions de seuil, alors nous aurons 40% de FRR. C'est-à-dire que bien que la probabilité d'accepter un imposteur par erreur soit faible (garantissant la sécurité du système), la probabilité de rejeter par erreur un client augmente également.

Les limites de variations du seuil pour obtenir la courbe ROC est que chaque point de la courbe ROC représente un seuil fixe correspondant au FRR et au FAR. On parcourt tous les valeurs de seuil de 0 à 1 et calcule le FRR et le FAR correspondants comme un point, et finalement on obtient la courbe formée de ces points.

## VII. Système dépendant du texte



A cause du problème de mise à l'échelle des coordonnées, nous ne pouvons pas bien voir la tendance de la courbe, mais nous pouvons constater que le système dépendant du texte a une FRR très faible.

C'est-à-dire qu'après avoir éliminé le facteur d'interférence des textes différentes, le taux de rejeter par erreur un client est fortement réduite. Mais cela a également entraîné une augmentation des FAR. Nous ne sommes pas sûrs, mais je suppose que c'est parce que l'apprentissage du modèle dépendant a pris les échantillons avec un texte fixé, ce qui conduit à un taux de réussite très élevé pour un échantillon de ce texte, même si c'est un imposteur.