



**HAL**  
open science

# Méthode de Détection et de Reconnaissance de Bruits Electromagnétiques permettant la Prédiction de leurs Effets sur les Transmissions GSM-R

Stephen Dudoyer

► **To cite this version:**

Stephen Dudoyer. Méthode de Détection et de Reconnaissance de Bruits Electromagnétiques permettant la Prédiction de leurs Effets sur les Transmissions GSM-R. Electronique. UNIVERSITE DES SCIENCES ET TECHNOLOGIE DE LILLE 1, 2013. Français. NNT : . tel-00930294

**HAL Id: tel-00930294**

**<https://theses.hal.science/tel-00930294>**

Submitted on 14 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**IFSTTAR**

INSTITUT FRANÇAIS  
DES SCIENCES  
ET TECHNOLOGIES  
DES TRANSPORTS,  
DE L'AMÉNAGEMENT  
ET DES RÉSEAUX

## Thèse de doctorat

---

Soutenue le 17 septembre 2013

# Méthode de Détection et de Reconnaissance de Bruits Electromagnétiques permettant la Prédiction de leurs Effets sur les Transmissions GSM-R

par **Stephen Dudoyer**

en vue de l'obtention du titre de docteur de  
**l'Université Lille 1** dans le cadre de  
**l'école doctorale n°072 – SPI**

**Structure de recherche d'accueil : LEOST**



**THESE**

*Présentée à*

**L'UNIVERSITE des SCIENCES et TECHNOLOGIES de LILLE  
(LILLE 1)**

*Pour l'obtention du grade de*  
**DOCTEUR en ELECTRONIQUE**

Par  
**Stephen DUDOYER**

**Méthode de Détection et de Reconnaissance de Bruits  
Electromagnétiques permettant la Prédiction de leurs Effets  
sur les Transmissions GSM-R**

Soutenue le 17 septembre 2013

Membres du jury :

**Flavio CANAVERO**  
**Marc HELIER**

*Rapporteur*  
*Rapporteur*

**Françoise PALADIAN**  
**Andrea MARISCOTTI**  
**Noël HADDAD**

*Examinatrice*  
*Examineur*  
*Examineur*

**Virginie DENIAU**  
**Sébastien AMBELLOUIS**

*Encadrante Ifsttar*  
*Encadrant Ifsttar*

**Marc HEDDEBAUT**

*Directeur de Thèse*





# REMERCIEMENTS

Je tiens à adresser mes remerciements à toutes les personnes qui ont, de près ou de loin, contribué au bon déroulement de cette thèse.

En commençant par toute l'équipe "encadrante" qui m'a suivie durant ces trois années (voire plus pour certains). Merci à mon directeur de thèse, Marc Heddebaut, pour ses conseils avisés, ses remarques toujours constructives et ses relectures attentives. Je remercie également Andrea Mariscotti qui, malgré la distance, a toujours suivi l'avancée de mes travaux et s'est régulièrement déplacé pour nos réunions d'avancement. Un grand merci à Sébastien Ambellouis dont on peut dire maintenant (alors que c'était loin d'être gagné au départ) qu'il aime la CEM ! Tu m'as fait découvrir plein de choses sur la classification dont je ne connaissais rien à la base et tu as su m'orienter sans rien m'imposer (je pense aux nombreux compromis que tu as fait...). Enfin, mes plus sincères remerciements vont à Virginie Deniau avec qui j'ai eu la chance de travailler depuis mon stage de fin d'études et qui a toujours su me guider tout en me laissant une grande autonomie dans mon travail. On a toujours très bien travaillé ensemble mais les choses ne s'arrêtent pas là car, ce n'est un secret pour personne, on a depuis longtemps développé de véritables liens amicaux. Je remercie au passage toute ta petite famille pour tous les bons moments que nous avons passés ensemble.

Je remercie messieurs Flavio Canavero et Marc Hélier d'avoir accepté d'être les rapporteurs de ma thèse ainsi que madame Françoise Paladian et monsieur Noël Haddad de m'avoir fait l'honneur d'assister à ma soutenance de thèse.

J'ai bien évidemment une pensée pour le reste de l'équipe CEM sur qui j'ai toujours pu compter et avec qui je suis régulièrement partie en déplacement pour des conférences ou des campagnes de mesures : Jean, Hassene et ma chère Sousou (on restera en contact quoi qu'il arrive ne t'inquiète pas).

Je n'oublie pas les nombreux collègues qui ont partagé (et ceux qui partagent encore) mon bureau (dans lequel a toujours régné la bonne humeur) durant toutes ces années à l'Ifsttar ainsi que toutes celles et ceux que j'ai sollicités au quotidien. Je pense notamment à l'indispensable secrétaire du LEOST, Corinne, et l'irremplaçable responsable RH de proximité, Isabelle (et son fameux hologramme !), pour les tâches administratives, Manu et Olivier (que j'ai régulièrement embêtés) pour l'informatique, Daniel pour la reprographie, Bernard pour les envois postaux et Jean-Pierre que j'allais voir lorsque j'avais besoin de conseils pour une manip. Je pense également à la célèbre "M" qui a toujours surveillé ses "agents" du couloir D. J'en profite pour remercier les stagiaires que j'ai encadrés et qui ont développé les applications informatiques utilisées dans cette thèse : Benjamin, Nabil et Rida.

Je remercie également Elodie pour les indispensables moments de détente/délire passés ensemble en compagnie d'Elody que je remercie encore une fois pour le magnifique pot de thèse qu'elle a organisé.

Merci aussi à tous mes amis de Polytech' Lille avec qui je suis toujours restée en contact depuis ma sortie de l'école : Catherine, Pierre, Pablo, Manue, Sylvie, Blaise et Nathalie.

Je pense également à mes amis d'Alsace, Anne et Jonathan, que j'ai toujours plaisir à retrouver et qui m'ont régulièrement reçue chez eux. Je n'oublie pas leurs familles qui sont, elles aussi, vraiment adorables.

Je ne peux bien évidemment pas achever ces remerciements sans citer ma sœur, Emeline, mon beau-frère, Cédric et ma petite filleule adorée, Célia, que je remercie pour les nombreux bons moments (je pense aux vacances, anniversaires, soirées "console" ...) passés en leur compagnie.

Enfin, je remercie mon chéri, Thierry, pour le soutien qu'il m'a apporté (et qu'il continue de m'apporter) et surtout pour l'amour qu'il me donne au quotidien.



# SOMMAIRE

<b>REMERCIEMENTS</b>	<b>3</b>
<b>INTRODUCTION GENERALE</b>	<b>9</b>
<b>Chapitre 1 : La CEM dans le domaine ferroviaire</b>	
<b>I - Les travaux de David Middleton</b>	<b>14</b>
<b>II - Standards et méthodologies de mesure existants</b>	<b>16</b>
II.1 - Les standards de compatibilité électromagnétique ferroviaires	16
II.2 - Les standards CEM	17
II.2.a - Définition des signaux de perturbation	17
II.2.b - Méthode d'injection des signaux de perturbation	19
<b>III - Le GSM-R et les sources de perturbation ferroviaires internes</b>	<b>20</b>
III.1 - Le système GSM-R	20
III.2 - Les perturbations EM transitoires	22
III.3 - Autres sources de perturbation	23
<b>IV - Vulnérabilité électromagnétique du GSM-R</b>	<b>24</b>
IV.1 - Méthodologie d'étude	24
IV.2 - Résultats des tests d'immunité	25
<b>V - Les besoins en termes de mesures d'émissions EM rayonnées dans le but de protéger le GSM-R et l'approche proposée dans cette thèse</b>	<b>26</b>
V.1 - Les besoins	26
V.2 - L'approche proposée	26
<b>Conclusion</b>	<b>28</b>
<b>Chapitre 2 : Rappels sur les méthodes de classification et les transformées temps-fréquence/temps-échelle</b>	
<b>I - La classification</b>	<b>30</b>
I.1 - Principe général de la classification	30
I.2 - Présentation de différents classifieurs existants	30
I.2.a - Méthodes de classification supervisées	30
I.2.b - Méthodes de classification non-supervisées	34
<b>II - Principe général d'une transformation temps-fréquence</b>	<b>37</b>
<b>III - Présentation de différentes transformations TF existantes</b>	<b>38</b>
III.1 - Représentations temps-fréquence linéaires	39
III.1.a - Transformée de Fourier à court terme (TFCT)	39
III.1.b - Transformée de Gabor	40
III.1.c - Comparaison de fenêtres	40
III.1.d - Transformée en ondelettes (WT : Wavelet Transform)	42
III.2 - Représentations temps-fréquence quadratiques	46
III.2.a - Spectrogramme (SPEC)	47
III.2.b - Distribution de Wigner-Ville (WD : Wigner-Ville Distribution)	47
III.2.c - Distribution de Choï-Williams (CWD : Choï-Williams Distribution)	48
III.3 - Bilan	49
<b>Conclusion</b>	<b>50</b>

<b>Chapitre 3 : Mises en œuvre préliminaires dans un contexte CEM _____</b>		<b>51</b>
<b>I - Application 1 : détection de signaux transitoires noyés dans du bruit blanc gaussien _____</b>		<b>52</b>
I.1 - Construction des fichiers de test _____		52
I.1.a - Principe _____		52
I.1.b - Variation des paramètres _____		53
I.2 - Méthode de détection utilisant la transformée de Gabor ou les ondelettes _____		53
I.2.a - Seuillage des coefficients de la matrice _____		53
I.2.b - Détection des positions des transitoires potentiels _____		55
I.2.c - Regroupement des colonnes de détection successives et détermination des instants d'apparition associés _____		56
I.2.d - Elimination des doublons _____		56
I.2.e - Affichage des résultats _____		58
I.3 - Comparaison des résultats obtenus par les deux transformées _____		58
I.3.a - Variations indépendantes des trois paramètres _____		59
I.3.b - Variations simultanées de deux des trois paramètres _____		59
I.3.c - Bilan de la comparaison des résultats _____		60
<b>II - Application 2 : classification de signaux _____</b>		<b>61</b>
II.1 - Principe de calcul _____		61
II.1.a - Calcul et représentation de la décomposition par paquets d'ondelettes _____		61
II.1.b - Calcul et représentation de la cartographie de l'énergie _____		63
II.1.c - Décomposition en valeurs singulières _____		66
II.1.d - Sélection des nœuds significatifs sur la cartographie _____		68
II.1.e - Construction du diagramme de dispersion _____		68
II.1.f - Utilisation de la règle du plus proche voisin _____		69
II.2 - Résultats obtenus sur un jeu de données réelles _____		70
II.2.a - Présentation du jeu de données _____		70
II.2.b - Résultats _____		70
<b>III - Bilan des deux applications préliminaires _____</b>		<b>72</b>
<b>IV - Principe de la méthodologie de classification proposée _____</b>		<b>73</b>
IV.1 - Objectif de notre méthodologie _____		73
IV.2 - Banc de mesures, configuration employée et protocole suivi _____		74
IV.2.a - Banc de mesures _____		74
IV.2.b - Configuration des équipements _____		75
IV.2.c - Protocole de collecte des fichiers temporels _____		75
<b>Conclusion _____</b>		<b>77</b>
<b>Chapitre 4 : Méthode de classification développée _____</b>		<b>79</b>
<b>I - Critères de sélection et définition du descripteur _____</b>		<b>80</b>
I.1 - Critères de sélection du descripteur _____		80
I.2 - Définition du descripteur _____		80
<b>II - Description de la méthode _____</b>		<b>82</b>
II.1 - Transformée de Gabor optimisée _____		82
II.2 - Recueil du descripteur _____		84
II.2.a. Détermination de $P_{\text{GSM-R}}$ et $P_{\text{max sur les autres canaux}}$ à chaque instant $t$ _____		84
II.2.b. Calcul du descripteur _____		85
II.3 - Extraction des paramètres _____		87
II.4 - Construction de la base de référence et visualisation du diagramme de dispersion _____		88
II.5 - Définition du seuil de détection : importance pour la méthode de classification et choix retenu _____		92
<b>III - Présentation de l'interface de calcul et visualisation _____</b>		<b>95</b>
III.1 - Description générale de l'interface _____		95

---

III.2 - Description des différents champs, outils et commandes	96
III.2.a. Champs d'entrée des paramètres de calcul	96
III.2.b. Choix du couple (fen, fpas) respectant l'inégalité d'Heisenberg	96
III.2.c. Détermination et application du seuil de détection des minima	97
III.2.d. Utilisation de l'intervalle de suivi dans le temps	99
III.2.e. Correspondance entre descripteur et diagramme de dispersion	100
III.2.f. Barre d'outils	101
<b>Conclusion</b>	<b>102</b>
<b>Chapitre 5 : Résultats de la méthode de classification</b>	<b>103</b>
<b>I - Construction de la base de référence avec les différents scénarios créés/simulés</b>	<b>104</b>
I.1 - Présentation des scénarios considérés	104
I.2 - Diagramme de dispersion des éléments de la base de référence	105
I.3 - Analyse et amélioration du diagramme de dispersion de référence	105
<b>II - Tests sur des signaux réellement acquis à bord d'un train</b>	<b>109</b>
II.1 - Présentation des scénarios créés à partir de signaux transitoires réels	109
II.2 - Diagrammes de dispersion et analyse des résultats	109
II.2.a. Base de développement	110
II.2.b. Tracé de la frontière entre les deux classes	112
II.2.c. Base de test	113
II.3 - Bilan et discussion des résultats	114
II.3.a. Augmentation du nombre de voisins	114
II.3.b. Compensation de la perte de dynamique liée au sous-échantillonnage lors de la collecte des fichiers à l'oscilloscope	115
<b>III - Tests sur des signaux acquis dans le métro de Barcelone</b>	<b>117</b>
III.1 - Présentation des signaux de test utilisés	117
III.2 - Analyse de la couverture spectrale des signaux de test utilisés	117
III.3 - Courbes obtenues pour le descripteur	119
<b>IV - Vers l'application automatique d'un seuil pour détecter et compter les minima</b>	<b>122</b>
<b>Conclusion</b>	<b>125</b>
<b>CONCLUSION GENERALE ET PERSPECTIVES</b>	<b>127</b>
<b>GLOSSAIRE</b>	<b>131</b>
<b>LISTE DES FIGURES</b>	<b>133</b>
<b>LISTE DES TABLEAUX</b>	<b>136</b>
<b>REFERENCES</b>	<b>137</b>
<b>BIBLIOGRAPHIE DE L'AUTEUR</b>	<b>145</b>

---



# INTRODUCTION GENERALE

Avec la multiplication des systèmes électroniques analogiques et numériques embarqués dans les moyens de transports actuels, nous assistons à une prolifération des sources de bruit ElectroMagnétique (EM) aussi bien à l'intérieur qu'à l'extérieur des véhicules. Ainsi, l'environnement EM devient de plus en plus riche en signaux de toutes sortes et, par conséquent, s'avère plus complexe à caractériser.

Dans le cadre de ces travaux de thèse, nous nous intéressons à un système en particulier. Il s'agit du GSM-R (Global System for Mobile communications - Railways). Le GSM-R est le nouveau système de communication numérique sans fil et de signalisation actuellement en cours de déploiement sur le réseau ferroviaire européen afin d'assurer l'interopérabilité des trains à grande vitesse en Europe. A terme, tous les pays d'Europe utiliseront ce même système, ce qui facilitera la circulation des trains d'un pays à l'autre. Il constitue l'un des composants du standard ERTMS (European Rail Traffic Management System) destiné à remplacer les 27 systèmes de signalisation ferroviaire actuellement en service en Europe. En effet, jusqu'à présent, chaque Etat possédait son propre système de signalisation et de gestion de trafic, de ce fait les trains transfrontaliers devaient être équipés de chacun des systèmes exploités dans les pays traversés. Avec le GSM-R comme support radio, un unique système de contrôle-commande (ETCS : European Train Control System) permettra la gestion de trafic ferroviaire sur l'ensemble du territoire européen.

Le GSM-R constitue une plate-forme sécurisée pour la transmission de la voix et des informations de trafic entre les différents membres des équipes ferroviaires. A terme, il assurera la transmission de la voix entre les agents (conducteurs et personnels des centres de contrôle) et des informations de signalisation qui s'afficheront directement sur le tableau de bord du conducteur. Il permettra un échange d'informations continu entre le train et le centre de régulation des trains avec un haut niveau de priorité. Si la connexion est perdue, le train s'arrête automatiquement.

De plus, afin de garantir un niveau de sécurité suffisant, la qualité de transmissions est évaluée en permanence et ne doit pas descendre en-deçà d'une limite fixée par les opérateurs ferroviaires.

La qualité et la continuité de service étant essentielles pour la bonne opération du système de gestion de trafic, il est donc indispensable de garantir une transmission correcte des données échangées via le GSM-R.

En environnement ferroviaire, le GSM-R doit faire face à différentes sources de bruit EM internes (matériel roulant, infrastructure d'alimentation électrique...) ou externes (autres systèmes de communication par exemple). Dans le cadre de cette thèse, nous limitons notre étude aux seules perturbations EM transitoires provenant du contact glissant entre la caténaire et le pantographe. Ces perturbations couvrent de larges bandes de fréquences dont, potentiellement, celles du GSM-R. Plusieurs campagnes de mesures ont été menées à bord de trains afin de collecter ces signaux et de procéder à leur caractérisation.

En raison de leur proximité avec la caténaire, les antennes GSM-R (fixées en toiture) reçoivent une partie des signaux transitoires à chacune de leurs apparitions. Ces derniers peuvent donc affecter les transmissions GSM-R.

Une étude de la susceptibilité du système GSM-R en présence de signaux transitoires a été précédemment menée au laboratoire. Celle-ci sert de point d'entrée à nos travaux.

Habituellement pour ce genre d'étude, on se réfère aux normes ferroviaires et CEM (Compatibilité ElectroMagnétique). Or, nous le verrons, ces dernières se révèlent peu adaptées pour un système mobile de transmission numérique en présence de ce type de perturbations. Les différentes méthodologies qu'elles décrivent ne permettent pas nécessairement de recueillir les caractéristiques des bruits EM les plus influentes sur la susceptibilité d'un dispositif électronique ou de télécommunication donné et plus particulièrement lorsqu'il s'agit de caractériser un environnement

mobile. En effet, les moyens et les méthodes de mesures sont initialement établis pour correspondre à des situations statiques et pour garantir la répétabilité et la reproductibilité des essais CEM. Ainsi, ils sont souvent peu adaptés face au caractère mobile des véhicules et à l'évolution de leur environnement EM. Il nous faut donc penser une nouvelle méthode de caractérisation de l'environnement EM adaptée au système radio mobile GSM-R que l'on cherche à protéger.

Dans ce travail de thèse, nous proposons d'établir un lien entre les conditions de bruit EM transitoire dans lesquelles se déroulent les transmissions GSM-R et l'impact de celles-ci sur la qualité des transmissions. Il s'agit de classer les bruits EM transitoires en deux catégories (qualité de communication satisfaisante pour l'opérateur ferroviaire et qualité en deçà de celle attendue) en fonction de leur impact (niveau de dégradation) sur la qualité d'une communication GSM-R. La méthodologie proposée est indépendante du système GSM-R à protéger puisqu'elle repose intégralement sur des mesures temporelles, réalisées à l'aide d'un oscilloscope numérique et un post-traitement informatique.

D'un point de vue pratique, cette méthode de classification pourrait être mise en place à bord des trains où elle permettrait de prédire l'effet des perturbations EM transitoires sur la qualité des transmissions GSM-R tout au long du trajet afin d'identifier et localiser les zones critiques pour le système. Les résultats fournis pourraient aussi bien servir au conducteur du train, pour une vérification in situ de la fiabilité des informations reçues, qu'à l'opérateur ferroviaire pour une identification des zones à surveiller ou investiguer.

Pour ce qui est de la standardisation, ces travaux de recherche peuvent contribuer à l'évolution nécessaire des équipements et des méthodologies définies dans les normes CEM afin de s'adapter aux nouvelles problématiques émanant de la multiplicité des systèmes et protocoles de communication sans fils employés dans le monde des transports.

Ce manuscrit est organisé de la façon suivante. Dans le chapitre 1, nous passons en revue les standards et méthodes de mesure habituellement appliqués pour des études CEM. En particulier, nous y expliquons en quoi les méthodes classiques se révèlent peu adaptées à nos travaux de recherche sur le GSM-R. Nous présentons ensuite le système GSM-R et les sources de perturbations EM considérés dans cette thèse. Après une synthèse des résultats de l'étude de la susceptibilité du GSM-R en présence de signaux transitoires, nous définissons les besoins en termes de mesures d'émissions rayonnées dans le but de le protéger et introduisons l'approche proposée dans ce travail de thèse.

Le chapitre 2 est consacré aux méthodes de classification et aux transformées temps-fréquence. Il s'agit des deux principaux outils employés pour cette étude. Nous commençons par rappeler le principe général de la classification, utilisée habituellement dans d'autres domaines que la CEM, avant de décrire quelques-unes des méthodes de classification supervisées et non-supervisées. Nous justifions également notre choix en ce qui concerne la méthode de classification retenue. Dans la seconde partie de ce chapitre, après un rappel sur le principe général d'une transformation temps-fréquence, nous présentons différentes transformations existantes. Lors du bilan, nous expliquons quelles sont celles que nous retenons et pourquoi nous effectuons ce choix.

Dans le chapitre 3, nous procédons à des mises en œuvre préliminaires des transformées temps-fréquence et de la méthode de classification retenues afin de nous assurer que ces deux outils conviennent bien à une application CEM comme la nôtre. Deux applications, développées par nos soins, sont présentées. La première concerne la détection de signaux transitoires noyés dans du bruit blanc gaussien et la deuxième consiste en de la classification de signaux.

La méthode de classification développée dans cette thèse fait l'objet du chapitre 4. Nous y décrivons, en détails, l'ensemble des étapes sur un scénario de bruit EM transitoire que nous avons construit.



Le chapitre 5 est consacré aux résultats de notre méthode de classification. Nous présentons et analysons les résultats obtenus avec les différents scénarios créés/simulés à partir de modèles de transitoire, générés sous Matlab<sup>®</sup>, ou de signaux réellement acquis à bord de véhicules ferroviaires (train et métro). Pour finir, nous concluons sur la pertinence et la robustesse de notre méthode de classification des bruits EM transitoires en fonction de leur impact sur la qualité des communications GSM-R.



# Chapitre 1

## La CEM dans le domaine ferroviaire

---

### SOMMAIRE

---

<i>I - Les travaux de David Middleton</i>	<b>14</b>
<i>II - Standards et méthodologies de mesure existants</i>	<b>16</b>
<i>III - Le GSM-R et les sources de perturbation ferroviaires internes</i>	<b>20</b>
<i>IV - Vulnérabilité électromagnétique du GSM-R</i>	<b>24</b>
<i>V - Les besoins en termes de mesures d'émissions EM rayonnées dans le but de protéger le GSM-R et l'approche proposée dans cette thèse</i>	<b>26</b>
<i>Conclusion</i>	<b>28</b>

---

L'étude de l'impact des signaux de perturbation EM sur des transmissions numériques a été introduite assez récemment en CEM. De manière générale, on se limite à l'étude des défauts électriques ou électroniques introduits par les perturbations. Or, nous allons le voir dans ce chapitre, avec le GSM-R, les méthodes classiques définies dans les normes ferroviaires et CEM s'avèrent insuffisantes. Il apparaît donc nécessaire d'explorer d'autres pistes en utilisant d'autres outils, ce qui constitue une proposition de remise en question des méthodes existantes [1.1].

Dans ce chapitre, après avoir rappelé le contexte des travaux menés par David Middleton, nous commençons par présenter les standards et méthodologies de mesure qui existent dans le domaine ferroviaire et en CEM et montrons en quoi ils sont peu adaptés au système que nous cherchons à protéger. Nous décrivons ensuite le GSM-R et les perturbations EM auxquelles nous nous intéressons, qui peuvent affecter les communications GSM-R. Enfin, nous expliquons ce qui est nécessaire à la protection du GSM-R en termes de mesures d'émissions EM rayonnées dans l'environnement ferroviaire et, introduisons la nouvelle approche que nous proposons pour répondre aux besoins d'un système de transmission numérique.

## I - Les travaux de David Middleton

David Middleton (1920-2008) a proposé et formulé des modèles canoniques (c'est-à-dire, dont les formulations analytiques ne varient pas selon la source du bruit, ni ses caractéristiques). Ceux-ci sont donc indépendants des mécanismes à l'origine du bruit EM [1.2]. Ainsi, ils permettent de modéliser les différentes interférences EM, qu'elles soient d'origine naturelle ou humaine, qui peuvent être rencontrées dans un environnement quelconque.

Pour ce faire, Middleton a défini trois classes d'interférence EM en fonction de la largeur spectrale du signal de perturbation ( $\Delta f_{\text{bruit}}$ ) vis-à-vis de celle de l'étage d'entrée du récepteur ( $\Delta f_{\text{récepteur}}$ ) [1.2] :

- la classe A :  $\Delta f_{\text{bruit}} \leq \Delta f_{\text{récepteur}}$ ,
- la classe B :  $\Delta f_{\text{bruit}} > \Delta f_{\text{récepteur}}$ ,
- la classe C : bruit correspondant à la somme d'un élément de la classe A et d'un élément de la classe B.

Notons qu'une formulation équivalente à celle d'un bruit de classe B peut toujours être trouvée pour une interférence de classe C. Ainsi, pour la plupart des applications, seules les classes A et B sont considérées [1.2]. Dans le cadre de nos travaux, les signaux de perturbation EM considérés sont des signaux large bande, donc appartenant à la classe B.

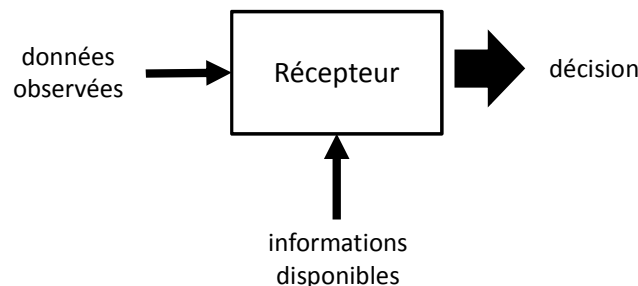
Middleton a construit un modèle différent pour chacune de ces classes. Le modèle de la classe A est décrit par une seule fonction caractéristique du premier ordre avec trois paramètres alors que pour la classe B, il faut une paire de fonctions caractéristiques et donc, six paramètres. Le lecteur pourra se reporter à [1.3] et [1.4] pour une description détaillée de la procédure de détermination de ces paramètres.

Ces modèles ont été construits pour, principalement, trois raisons [1.4] :

- fournir des descriptions réalistes des interférences EM d'origine naturelle et humaine,
- spécifier des méthodes de mesure de ces interférences,
- déterminer une structure optimale pour les systèmes de communication et évaluer et comparer leurs performances avec des systèmes non optimisés en environnement bruité.

Pour ce qui est de l'étude de l'effet des interférences EM sur les performances des systèmes de communication, et ceci rentre dans le cadre de ce travail de thèse, celle-ci se limite aux « problèmes de réception ».

Un problème de réception survient quand il faut prendre une décision, à propos d'un signal, sur la base des observations de ce signal (éventuellement complétée des informations disponibles à propos du signal), après contamination par du bruit [1.5]. Une illustration de ce processus est proposée en figure 1.1.



**Figure 1.1. Illustration d'un problème de réception**

Cette décision peut être de deux types [1.5] :

- déterminer s'il y a présence ou absence d'un (ou plusieurs) signal(aux) parmi le bruit. C'est ce que l'on appelle la « détection ».
- estimer certaines données ou paramètres (amplitude, fréquence, forme d'onde...) d'un signal bruité. Il s'agit de l'« extraction ».

Dans les deux cas, la règle de décision est basée sur des probabilités. Le but étant de minimiser le risque d'erreur, soit encore de mauvaise décision.

Bien que les travaux de David Middleton soient très complets, aboutis et applicables à tous les types d'interférences EM dont celles que nous considérons dans cette thèse, nous ne pouvons toutefois pas nous y référer pour notre étude. En effet, ces travaux utilisent les modèles de bruit construits pour définir des récepteurs optimaux. Le récepteur optimal étant défini pour sa part comme celui qui minimise les mauvaises décisions.

Dans notre cas il s'avère en effet exclu d'apporter quelque modification que ce soit au récepteur GSM-R élaboré selon des spécifications générales ne prenant pas nécessairement en compte les spécificités du bruit EM provenant du contact glissant fil caténaire-pantographe. Par ailleurs, notre but n'est pas d'optimiser le récepteur mais d'étudier l'impact des perturbations EM considérées dans cette thèse sur la qualité des transmissions GSM-R. Nous devons donc comparer l'information reçue par le mobile GSM-R, en présence de perturbations EM avec celle émise au départ et non pas extraire certaines données du signal bruité à l'entrée du récepteur.

Afin d'obtenir plus de précision sur l'environnement EM ferroviaire général dans lequel doivent opérer les récepteurs, rappelons maintenant quelques standards et méthodologies de mesure habituellement considérés.

## II - Standards et méthodologies de mesure existants

### *II.1 - Les standards de compatibilité électromagnétique ferroviaires*

Les standards communément appliqués dans le ferroviaire font partie de la série des normes EN 50121 [1.6] qui regroupe 5 parties :

- EN 50121-1 : Généralités.

Celle-ci fournit une description du comportement électromagnétique d'un système ferroviaire et spécifie le critère de performance employé dans cette série de normes.

- EN 50121-2 : Emission du système ferroviaire dans son ensemble vers le monde extérieur.

Elle fixe les limites d'émission du système ferroviaire vers le monde extérieur et définit les méthodes de test à appliquer pour vérifier les émissions.

- EN 50121-3 : Matériel roulant (section 1 : trains et véhicules complets, section 2 : appareils).

La section 1 spécifie les besoins en termes d'émission et d'immunité pour tous les types de matériel roulant. En particulier, elle fixe des valeurs-limites d'émission des véhicules à l'arrêt et en déplacement à vitesse lente.

La section 2 s'applique aux équipements électriques et électroniques intégrés dans le matériel roulant.

Elle fixe des valeurs-limites d'émission de ces équipements et définit les tests d'immunité qui peuvent être pratiqués sur ces mêmes équipements.

- EN 50121-4 : Emission et immunité des appareils de signalisation et de télécommunication.

Elle spécifie les limites d'émission et d'immunité EM des appareils de signalisation et de télécommunication et définit les tests d'immunité pouvant être pratiqués sur les différents ports d'entrée/sortie.

- EN 50121-5 : Emission et immunité des installations fixes d'alimentation de puissance et des équipements associés.

Elle décrit les essais d'émission et d'immunité pour les appareils et les systèmes utilisés dans les installations fixes associées à l'alimentation.

Les tests et méthodes de mesure qui y sont décrits peuvent être appliqués en laboratoire ou le long des voies ferrées. Par exemple, les émissions du véhicule ferroviaire vers le monde extérieur sont mesurées à 10 mètres des voies à l'aide d'un analyseur de spectre configuré en mode "maxhold". Le mode maxhold ne conserve que la valeur maximale sur chaque fréquence durant la totalité de la mesure. On récupère ainsi, pour chaque fréquence de la bande d'étude, la puissance maximale des signaux de bruit EM à cette fréquence qui est représentative du pire cas. Les mesures sont ensuite comparées à des gabarits de limites d'émission tels que celui défini dans la norme EN 50121-2 dont la représentation est fournie en figure 1.2 (en page suivante), afin de s'assurer que les spécifications sont bien respectées.

La limitation introduite par ce type de mesure, uniquement fréquentielle, est que, bien évidemment, l'aspect temporel n'est absolument pas pris en compte. En effet, en procédant de la sorte il est impossible de savoir à quels instants se sont produits les signaux qui ont généré la puissance maximale enregistrée à une fréquence donnée et si ce phénomène s'est répété plusieurs fois, ou s'il s'agit d'un événement isolé. Cette méthode est donc peu adaptée aux systèmes qui nécessitent un suivi dans le temps des perturbations EM présentes dans leur environnement de fonctionnement.

C'est cependant le cas des systèmes de communication numérique mobiles tels que le GSM-R auquel nous nous intéressons dans ce travail de thèse. En effet, pour ce système de radiocommunication mobile (qui sera décrit dans la section III.1), la puissance du signal utile varie au cours du temps. Ainsi, connaître la puissance maximale des perturbations ne suffit pas. Le niveau de puissance ramené par les perturbations EM sur les bandes de fonctionnement du GSM-R doit être relié à une durée d'exposition et une fréquence d'apparition du phénomène, ce qui est impossible avec la méthode

préconisée dans la norme EN 50121-2 qui se cantonne à la détermination du pire cas d'exposition aux perturbations.

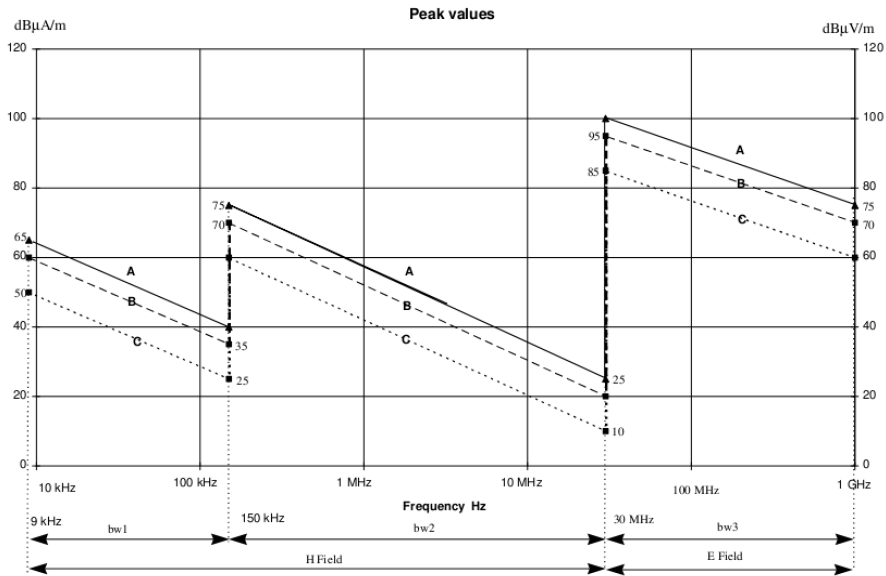


Figure 1.2. Gabarit d'émission, A=25 kV AC - B = 15 kV AC, 3 kV AC ou 1.5 kV DC - C = 750 V DC et bw1 = 200 Hz - bw2 = 9 kHz - bw3 = 120 kHz

## II.2 - Les standards CEM

### II.2.a - Définition des signaux de perturbation

Après avoir procédé à la caractérisation de l'environnement de fonctionnement d'un équipement, on se penche sur l'étude de son immunité aux différents signaux de perturbation identifiés via des essais CEM. Ces essais sont définis dans des normes qui décrivent la méthodologie et les signaux de perturbation à employer. Ces derniers correspondent aux signaux de perturbation les plus courants dans un environnement industriel. Chaque perturbateur possède d'ailleurs sa propre norme :

- décharges électrostatiques → EN 61000-4-2 [1.7],
- champs EM rayonnés aux fréquences radioélectriques → EN 61000-4-3 [1.8],
- transitoires électriques rapides en salves → EN 61000-4-4 [1.9],
- ondes de choc → EN 61000-4-5 [1.10].

Or, force est de constater que parmi tous ces signaux, aucun ne correspond exactement aux signaux de perturbation dont nous cherchons à évaluer l'impact sur les communications GSM-R, à savoir les signaux transitoires qui trouvent leur origine dans le contact glissant entre le fil caténaire et le pantographe. Il existe bien les signaux de la norme EN 61000-4-4, qui correspondent à des perturbations provenant de transitoires de commutation (switch, rebondissements de contacts de relais, connexion d'un équipement "lourd" au réseau électrique...), qui s'en rapprochent. Cependant leurs caractéristiques temporelles diffèrent de celles des transitoires observés à bord des trains car ce sont des phénomènes notablement plus lents qui en sont à l'origine. En effet, si le décollement du pantographe (par rapport à la caténaire) est équivalent, d'un point de vue mécanique, à ce phénomène de commutation, en revanche, vient se rajouter la décharge électrostatique qui est bien plus rapide. A titre de comparaison, nous présentons figure 1.3 les caractéristiques temporelles des transitoires définis dans la norme 61000-4-4 et celles les plus fréquemment rencontrées pour des transitoires collectés à

bord d'un train. Les représentations des formes d'onde correspondantes dans les domaines temporel et fréquentiel y figurent également.

La première caractéristique temporelle comparée est la durée à 50% du transitoire. Celle-ci correspond à l'intervalle de temps durant lequel l'amplitude du transitoire est supérieure ou égale à 50% de son amplitude crête [1.9]. Dans la norme, cette valeur est fixée à 50 ns ± 30%, alors qu'à bord d'un train la valeur la plus fréquente est de 5 ns, soit dix fois moins.

La deuxième caractéristique temporelle est le temps de montée qui correspond à l'intervalle de temps durant lequel l'amplitude du transitoire passe, pour la première fois, de 10% à 90% de son amplitude crête [1.9]. Dans la norme, le temps de montée est fixé à 5 ns ± 30%, alors qu'à bord d'un train la valeur la plus fréquemment rencontrée vaut 0.4 ns, soit 12.5 fois moins.

Comme le montre la figure 1.3, la conséquence de ces différences de valeur est visible en premier lieu dans le domaine temporel. Nous commençons par construire un modèle de transitoire sous la forme d'une double exponentielle en utilisant les deux couples de caractéristiques (durée, temps de montée). Avec les caractéristiques temporelles extraites de la norme 61000-4-4, le signal possède une durée totale plus longue que celui construit à partir des valeurs observées à bord d'un train, à cause de sa durée à 50%, dix fois plus importante.

Les bandes de fréquences couvertes par ces signaux perturbateurs seront donc par conséquent très différentes. Pour le vérifier, nous calculons dans un second temps les spectres de ces deux signaux transitoires à l'aide d'une FFT (Fast Fourier Transform). Sur la représentation spectrale présentée figure 1.3 nous faisons également apparaître les bandes de fréquences du GSM-R pour plus de clarté. Nous constatons que leurs couvertures spectrales respectives sont différentes. Si l'allure générale reste la même (comme pour la représentation temporelle), les niveaux de puissance ramenés par les deux transitoires se distinguent nettement. Cette différence est particulièrement visible dans les bandes de fréquences du GSM-R pour lesquelles on relève presque 30 dB d'écart. Le bruit EM généré par le modèle utilisant les caractéristiques relevées à bord d'un train est beaucoup plus élevé. Cela signifie qu'en utilisant les valeurs de durée et de temps de montée de la norme EN 61000-4-4, pour modéliser les transitoires issus du contact caténaire-pantographe, on sous-estimerait le niveau de bruit EM qu'ils produisent sur les bandes du GSM-R à chaque apparition. Il nous est donc impossible de nous référer aux valeurs fournies dans cette norme.

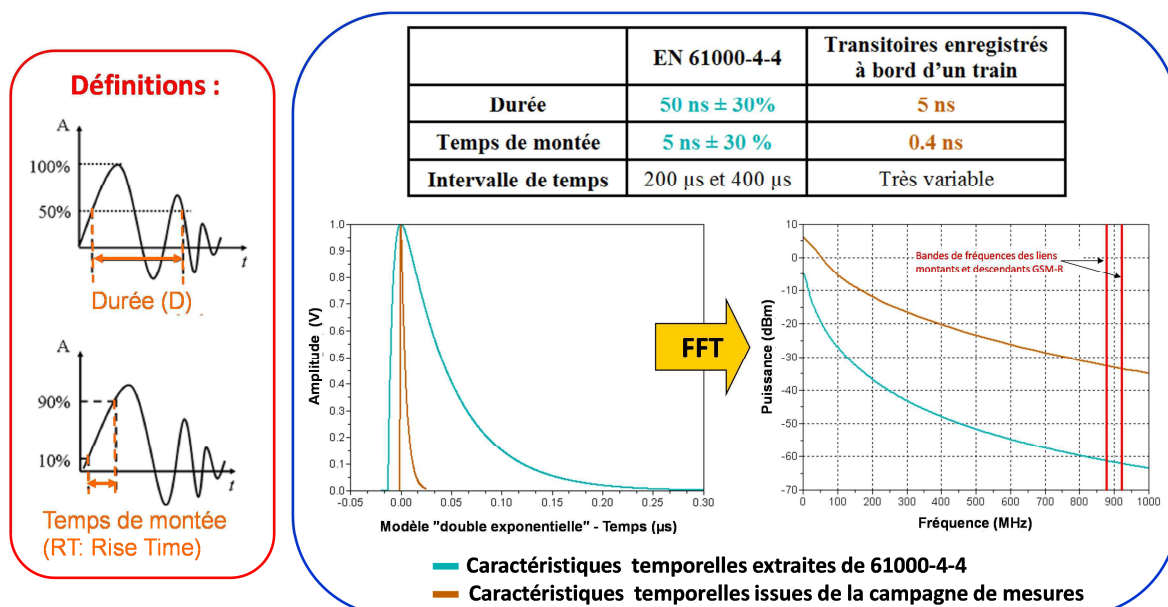
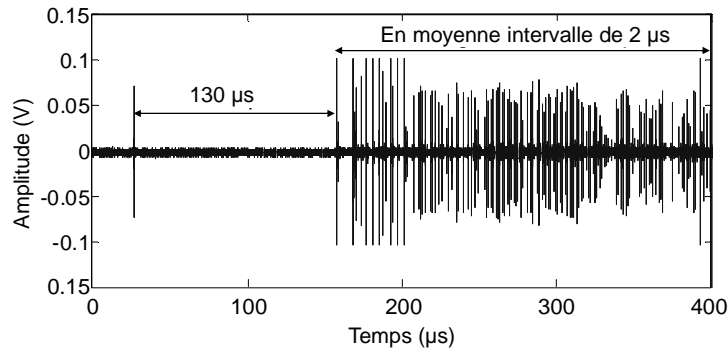


Figure 1.3. Caractéristiques temporelles des signaux transitoires définis dans la norme 61000-4-4 et des transitoires collectés à bord d'un train



De plus, une troisième caractéristique de temps, que nous n'avons pas encore introduite, présente elle aussi des différences significatives en termes de valeurs. Il s'agit de l'intervalle de temps entre deux transitoires successifs. Dans la norme, l'apparition des transitoires est considérée comme un phénomène très répétitif dans la mesure où seulement deux valeurs d'intervalle sont définies. La réalité est en revanche toute autre. En effet, à bord d'un train il s'avère qu'il s'agit d'un phénomène très aléatoire dont la récurrence dépend de nombreux paramètres (vitesse du train, usure du fil caténaire, conditions climatiques, présence de givre sur la caténaire...). L'intervalle de temps peut descendre jusqu'à quelques  $\mu\text{s}$  dans le cas d'une salve de transitoires, comme il peut monter à plusieurs centaines de  $\mu\text{s}$ , voire plus, lorsque les transitoires sont moins fréquents. Cette situation est illustrée par la figure 1.4 qui présente un exemple d'enregistrement de transitoires à bord d'un train.



**Figure 1.4. Illustration de la variation de l'intervalle de temps entre deux transitoires successifs à bord d'un train**

Impossible donc de fixer arbitrairement une valeur ou deux valeurs d'intervalle de façon figée pour quantifier la fréquence d'apparition des transitoires à bord d'un train. Or, nous verrons dans la suite de ce chapitre que la récurrence constitue un paramètre essentiel à l'étude de l'impact des transitoires sur le système GSM-R.

### II.2.b - Méthode d'injection des signaux de perturbation

Les signaux de perturbation sont injectés soit directement sur l'enveloppe de l'équipement (cas des décharges électrostatiques), soit sur ses ports d'entrée/sortie et/ou d'alimentation électrique. Le phénomène transitoire qui fait l'objet de notre étude peut s'apparenter à une décharge électrostatique. Cependant, dans notre cas, c'est la perturbation rayonnée induite par la décharge électrostatique qui intervient et non pas la décharge elle-même. De ce fait, nous ne pouvons pas nous satisfaire d'envoyer des décharges sur un mobile GSM-R car cela ne serait pas représentatif de la présence de signaux transitoires tels qu'ils apparaissent à bord des trains.

De plus, habituellement avec ce type d'essais CEM, on s'intéresse aux défauts électriques ou électroniques des équipements introduits par les perturbations EM. Or, avec le GSM-R il s'agit d'étudier le lien radio et non pas l'électronique du système, comme cela se fait couramment. Il devient donc nécessaire de s'éloigner de ce qui se pratique habituellement et d'imaginer un autre mode d'injection des perturbations EM transitoires qui permette d'affecter le lien radio afin d'évaluer l'impact de ces perturbations sur la qualité des transmissions GSM-R.

En résumé, que ce soient les modèles de signaux de test, les méthodes de mesure ou celles d'injection des perturbations, aucune méthode ne convient idéalement à notre étude. Il nous faut proposer une nouvelle méthodologie adaptée aux caractéristiques à la fois du système GSM-R et des perturbations EM considérées dans cette thèse. Nous la présentons dans la section V de ce chapitre.

### III - Le GSM-R et les sources de perturbation ferroviaires internes

#### III.1 - Le système GSM-R

Le GSM-R est un élément clé du nouveau standard ERTMS / ETCS. Il s'agit du système de communication radio numérique paneuropéen portant la signalisation actuellement en cours de déploiement sur tout le réseau ferroviaire européen. Tous les Etats membres européens se sont engagés à utiliser ce système afin d'assurer notamment l'interopérabilité des trains à grande vitesse sur l'ensemble du territoire européen. A terme, il assurera la transmission de la voix entre les agents (conducteurs et personnels des centres de contrôle) et des informations de signalisation qui s'afficheront également sur le tableau de bord du conducteur.

Le GSM-R constitue une plate-forme sécurisée pour les communications de voix et de données entre les différents membres des équipes ferroviaires : conducteurs, régulateurs, membres de l'équipe de manœuvre, personnels à bord du train, chefs de gare. Il apporte des dispositifs spécifiquement développés pour le domaine ferroviaire tels que les appels de groupes (VGCS : Voice Group Call Services), les annonces ou appels diffusés (VBS : Voice Broadcast Services), les appels localisés, et la préemption des appels en cas d'urgence.

En pratique, le système GSM-R utilise des mobiles embarqués à bord des trains connectés à des antennes fixées en toiture ainsi que des stations de base (BTS : Base Transceiver Station) installées spécifiquement et régulièrement le long des voies ferrées. Le mobile GSM-R est connecté à un réseau avec lequel il échange en permanence des informations. Lorsque les données sont transmises du train vers la BTS, on dit que la communication se produit sur le lien montant (à l'inverse, de la BTS vers le mobile, c'est le lien descendant qui est employé).

La couche radio GSM-R est basée sur la norme GSM phase 2 + [1.11], [1.12]. Cependant le système possède ses propres bandes de fréquences de fonctionnement. Deux bandes lui ont spécifiquement été allouées en Europe :

- 876-880 MHz, pour le lien montant,
- 921-925 MHz, pour le lien descendant.

Ces bandes sont directement adjacentes à celles employées par le GSM public, comme le montre la figure 1.5.



Figure 1.5. Bandes de fréquences allouées au GSM-R

Comme dans le cas du GSM, l'espacement de fréquence entre chaque canal physique est de 200 kHz (on compte donc 38 canaux au total). La modulation est de type GMSK (Gaussian Minimum Shift Keying).

Le GSM-R utilise un mode de multiplexage temporel, permettant de transmettre plusieurs signaux sur un seul canal, de type TDMA (Time Division Multiple Access). La transmission des données est organisée pour chaque porteuse (ou canal physique) par l'envoi périodique de trame TDMA, de période 4.615 ms.

Chaque trame TDMA est découpée en 8 intervalles de temps (nommés time slots) également appelés canaux logiques, d'une durée de 577  $\mu$ s chacun. La durée d'une trame TDMA est donc égale à 8 fois la durée d'un slot, soit pratiquement 4.6 ms.

Chaque utilisateur utilise un slot par trame TDMA. Une même trame peut ainsi compter au plus 8 utilisateurs sur un même canal physique. Chaque utilisateur a donc accès à tour de rôle au canal de communication pendant 577  $\mu$ s, comme représenté sur la figure 1.6.

Un slot accueille un élément de signal radioélectrique, appelé burst, composé de 156 bits dont 148 bits d'information. Par conséquent, la durée de transmission d'un bit GSM-R est de 3.7  $\mu$ s.

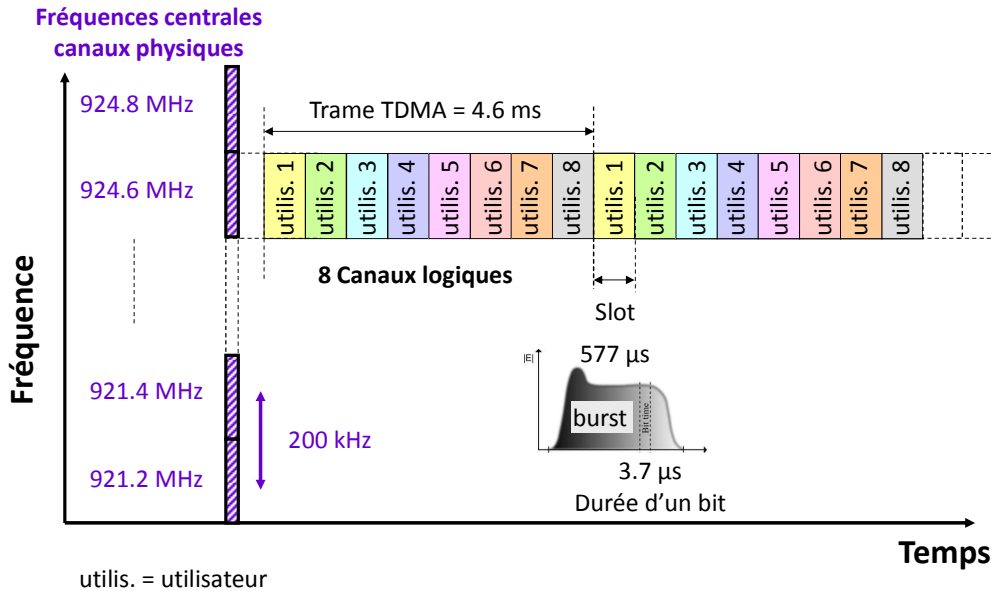


Figure 1.6. Description du système TDMA employé par le GSM-R

Pour ce qui concerne le niveau de couverture radioélectrique des signaux GSM-R le long de la ligne et sa dynamique, nous présentons en figure 1.7, un résultat d'une campagne de mesures réalisée à bord de trains sur des lignes ferroviaires équipées du système GSM-R [1.13]. Cette figure montre l'évolution de la puissance maximale de réception (lien descendant) mesurée en dBm, avec un analyseur de spectre sur l'ensemble de la bande GSM-R au cours du déplacement du train sur 16 km. Comme nous pouvons le constater, ce niveau varie de -20 dBm, lorsque l'antenne GSM-R est très proche de la station de base, à -90 dBm, lorsque le train (donc l'antenne GSM-R fixée sur son toit) est à mi-chemin entre deux stations de base successives.

Ces valeurs sont en accord avec les spécifications [1.14], selon lesquelles la puissance des signaux reçus par les antennes GSM-R peut varier de -35 dBm à -92 dBm.

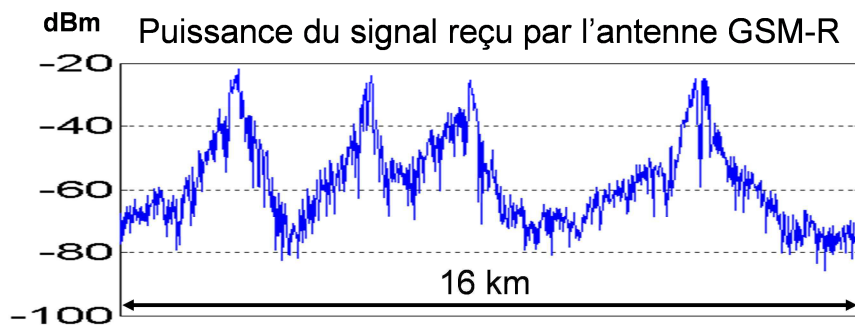
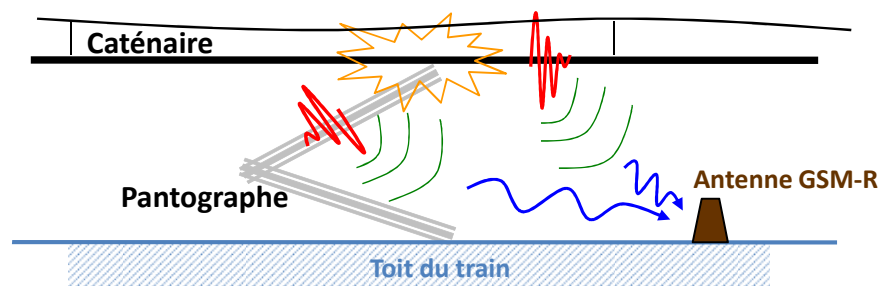


Figure 1.7. Evolution de la puissance du signal GSM-R reçu à bord du train au cours de son déplacement

En revanche, les signaux du lien montant sont émis par les antennes GSM-R avec une puissance beaucoup plus élevée (jusqu'à 8 W) que les niveaux reçus sur le lien descendant. Par conséquent, les signaux de perturbation présents dans l'environnement ferroviaire peuvent, à condition de produire un niveau de bruit suffisant, perturber la réception des signaux GSM-R mais n'affectent pas l'émission. C'est la raison pour laquelle nous limitons l'étude de la susceptibilité du GSM-R au seul lien descendant (921-925 MHz).

### III.2 - Les perturbations EM transitoires

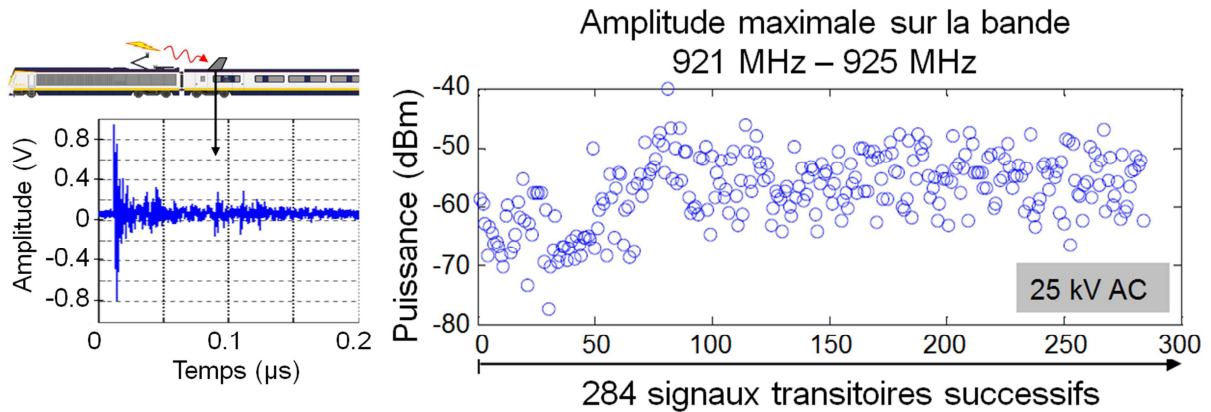
Les perturbations EM transitoires prennent leur source dans le contact glissant qui existe entre le fil caténaire et le pantographe. Lorsqu'il se produit une perte de contact, ou plus généralement un mauvais contact, apparaît un courant transitoire sur la caténaire et le pantographe en raison de la différence de potentiel créée entre ces deux éléments. Ce courant transitoire est conduit par les éléments métalliques constituant le fil caténaire et le pantographe qui rayonnent à leur tour une partie de cette énergie. L'antenne GSM-R, située sous la caténaire et à proximité du pantographe, met en évidence une partie de ce rayonnement. Ce phénomène est illustré par la figure 1.8.



**Figure 1.8. Origine des perturbations EM transitoires reçues par l'antenne GSM-R fixée en toiture**

On peut se demander s'il ne suffirait pas d'éloigner l'antenne GSM-R du pantographe pour éviter qu'elle ne réceptionne ces signaux de perturbation. Or, nous l'avons dit, l'antenne est fixée sur le toit du train. Dans ces conditions, peu importe la distance qui la sépare du pantographe, elle se trouvera toujours sous la caténaire qui conduit, elle aussi, les courants transitoires. Donc, bien que la précaution d'éloigner, autant que possible, l'antenne GSM-R du pantographe soit déjà prise par les constructeurs de matériels ferroviaires, elle ne résout pas totalement le problème.

Les signaux transitoires sont des événements qui couvrent de larges bandes de fréquences, dont potentiellement celles du GSM-R. Pour estimer le niveau de bruit EM généré par ces signaux dans la bande 921-925 MHz, nous utilisons des enregistrements de transitoires collectés à bord d'un train sur une ligne non équipée en GSM-R. Un traitement basé sur un calcul de FFT leur est appliqué de manière à extraire le niveau de puissance maximal généré par chacun sur la bande 921-925 MHz [1.13]. L'étude réalisée sur 284 transitoires successifs a conduit au résultat présenté en figure 1.9. Sur ce graphe, il apparaît que les transitoires étudiés produisent du bruit EM dont le niveau maximum varie entre -40 et -80 dBm. Or, nous venons de le voir, la puissance des signaux GSM-R reçus à bord des trains peut varier de -35 dBm à -92 dBm. Par conséquent, les niveaux de variation du bruit transitoire sont comparables aux niveaux de variation du signal utile. Ceci a pour conséquence que les signaux transitoires sont susceptibles de perturber les transmissions GSM-R lors de leur apparition.



**Figure 1.9. Étude de la puissance générée par les perturbations EM transitoires sur la bande du lien descendant du GSM-R**

### *III.3 - Autres sources de perturbation*

Les signaux transitoires ne sont pas les seules sources de perturbation du GSM-R. En effet, ce dernier doit cohabiter avec d'autres systèmes de communication et notamment le GSM public dont les bandes de fréquences sont adjacentes à celles du GSM-R. Il arrive également fréquemment que, pour des raisons de coût, une antenne GSM-R et une antenne GSM soient colocalisées sur le même mât, le long des voies ferrées.

Une étude a été menée sur l'impact des signaux GSM public sur les communications GSM-R [1.15]. Celle-ci a montré que les transmissions sur le premier canal GSM public (925.2 MHz) peuvent perturber les communications GSM-R qui ont lieu sur le dernier canal du lien descendant (924.8 MHz). En effet ces deux canaux ne sont espacés que de 400 kHz, ce qui fait qu'une partie de la puissance émise sur le canal GSM public peut subsister sur celui du GSM-R. Ces problèmes sont abordés dans le document ETSI TS 145 005 [1.16] et peuvent entraîner notamment des phénomènes de désensibilisation des récepteurs.

Il existe d'autres sources de perturbations EM qui peuvent être citées et qui sont décrites dans [1.17] :

- issues du matériel roulant,
- provenant de l'infrastructure d'alimentation électrique,
- externes au domaine ferroviaire.

Dans le cadre de cette thèse, nous limitons notre étude aux seuls signaux transitoires. Toutefois, la méthodologie que nous proposons peut très bien être adaptée pour procéder à l'étude des perturbations engendrées par les signaux du GSM public ou des autres perturbations précédemment listées.

## IV - Vulnérabilité électromagnétique du GSM-R

La susceptibilité EM des communications GSM-R face aux signaux transitoires issus du contact caténaire-pantographe a fait l'objet de travaux antérieurs menés au laboratoire [1.18], [1.19] et [1.20]. Elle a notamment mis en évidence les paramètres de bruit transitoire les plus influents sur la qualité des communications GSM-R.

Nous présentons ici la méthodologie et les principaux résultats de cette étude de susceptibilité qui a servi de point de départ à ce travail de thèse.

### IV.1 - Méthodologie d'étude

Le principe consiste à simuler la présence de signaux de perturbations transitoires durant une communication GSM-R, comme indiqué sur la figure 1.10. Dans un premier temps, une communication est établie entre un mobile GSM-R et une station de base ou un simulateur de réseau GSM-R. Nous optons pour un simulateur dans la mesure où cela nous permet de contrôler les signaux émis (canal d'émission, puissance, paramètres du réseau...) et d'effectuer des mesures du BER (Bit Error Rate) qui constitue le critère employé pour juger de la qualité de la communication GSM-R. Une explication sur la façon dont le simulateur de réseau détermine le BER est fournie dans le chapitre 3. Durant la communication, un générateur de signaux arbitraires est utilisé pour générer des signaux transitoires et venir ainsi perturber les transmissions.

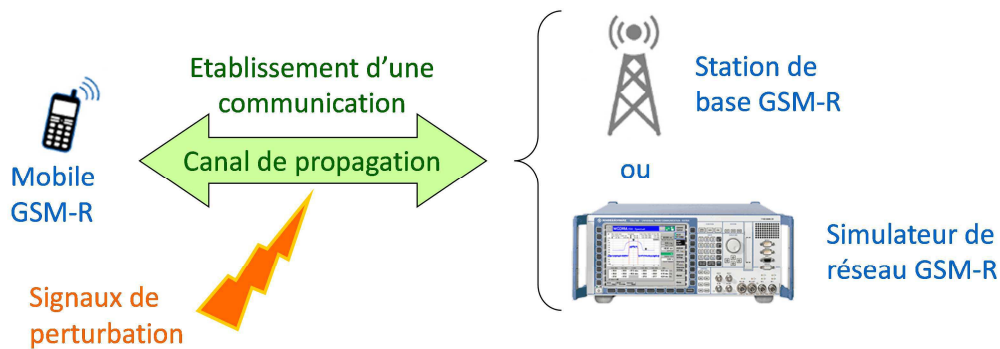


Figure 1.10. Principe de la mesure

Notre hypothèse de travail initiale est de considérer divers scénarios d'apparition des transitoires en jouant sur plusieurs paramètres : type de transitoire, récurrence des signaux, puissance générée sur le canal GSM-R et d'estimer l'impact de chaque scénario de bruit sur la communication GSM-R en cours, grâce à une mesure du BER.

La puissance des signaux GSM-R qui parviennent au mobile est réglée de manière à fixer le rapport signal à bruit (ou SNR, pour Signal-to-Noise Ratio) à la valeur désirée.

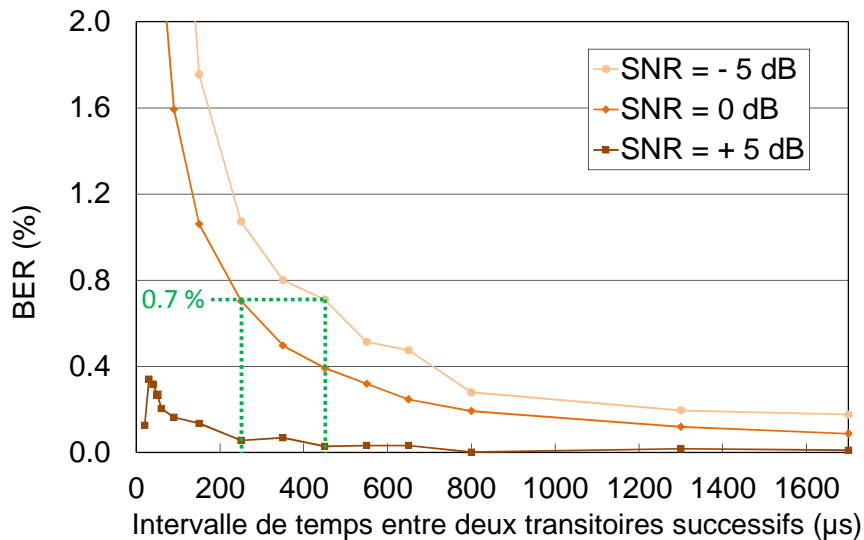
Les signaux transitoires sont générés avec un intervalle de temps TI (Time Interval) constant entre deux événements successifs. Plusieurs modèles de transitoire ont été définis sur la base des résultats présentés dans [1.21]. Les valeurs de TI et du SNR ainsi que le modèle de transitoire sont les trois paramètres que nous faisons varier lors de nos essais pour créer les différents scénarios de test.

Dans la section IV.2, nous donnons les résultats obtenus, en procédant de la sorte, pour trois valeurs différentes du SNR.

## IV.2 - Résultats des tests d'immunité

La figure 1.11 présente l'évolution du BER en fonction de l'intervalle de temps entre deux transitoires successifs et pour trois valeurs du SNR : -5, 0 et +5 dB. Ce graphe montre, premièrement, que le BER augmente avec la fréquence d'apparition des transitoires quel que soit le SNR. Donc, dans tous les cas, plus les transitoires sont récurrents, plus la communication GSM-R est dégradée. Toutefois, en fonction de la valeur du SNR cette dégradation s'avère plus ou moins forte. En effet, lorsque le SNR vaut +5 dB (c'est-à-dire lorsque la puissance du signal de communication GSM-R excède de 5 dB celle du bruit transitoire), l'augmentation de la fréquence d'apparition des transitoires possède très peu d'effet sur le BER qui demeure inférieur à 0.4 %. En revanche, dans le cas contraire (SNR = -5 dB), le BER augmente plus rapidement avec la récurrence des transitoires que dans le cas d'un SNR à 0 dB (pour lequel la puissance du signal GSM-R est égale à celle du bruit transitoire). Il existe donc un lien entre l'évolution du BER, la récurrence d'apparition du perturbateur et le SNR.

Plus précisément, nous constatons qu'une même valeur de BER peut être obtenue pour deux configurations différentes (SNR, TI) d'apparition des transitoires. Par exemple, une valeur de BER de 0.7 % peut être atteinte avec un SNR de -5 dB associé à un TI de 450  $\mu$ s mais également avec un SNR de 0 dB associé cette fois-ci à un TI de 250  $\mu$ s.



**Figure 1.11. Evolution du BER en fonction de l'intervalle de temps entre deux transitoires successifs pour trois valeurs du SNR**

Nous en concluons que pour les perturbations transitoires, les deux paramètres les plus influents sur la qualité d'une communication GSM-R sont :

- la récurrence des signaux,
- la dégradation du SNR à chaque apparition d'un signal transitoire.

Il nous faut donc trouver un moyen de suivre simultanément l'évolution de ces deux paramètres afin d'être capable de prévoir correctement un impact sur la qualité de réception des signaux GSM-R.

## V - Les besoins en termes de mesures d'émissions EM rayonnées dans le but de protéger le GSM-R et l'approche proposée dans cette thèse

### *V.1 - Les besoins*

Le GSM-R présente la particularité d'être à la fois un système de radiocommunication et un système radio mobile exploité avec des vitesses de mobilité élevées, actuellement jusque 350 km/h. Ces deux aspects doivent être pris en compte pour définir la méthode de mesure à employer dans le but de déterminer l'impact des perturbations EM qu'il rencontre dans son environnement de fonctionnement. D'une part, nous l'avons montré, le niveau de puissance du signal de communication GSM-R n'est pas constant mais évolue rapidement en fonction de la distance entre le mobile et la station de base. D'autre part, l'impact du bruit EM transitoire sur les communications GSM-R dépend non seulement du niveau d'interférence qu'il induit sur les bandes de fréquences GSM-R mais également du niveau des signaux de communication.

Il n'est donc pas envisageable de simplement mesurer les signaux de perturbations transitoires et de leur appliquer des valeurs-limites d'émission. Il devient par contre indispensable d'observer et de suivre l'évolution, au cours du temps, du rapport entre le signal utile (signal de communication GSM-R) et le bruit EM transitoire afin de conclure sur la qualité des transmissions.

Cela suppose d'être capable de mesurer en permanence la puissance à la fois des signaux de communication GSM-R et des signaux de perturbations transitoires. Il s'agit véritablement d'une remise en question de nos méthodes car la vulnérabilité du système GSM-R ne peut être exprimée qu'en fonction de l'évolution au cours du temps et non pas sur la base d'une "photographie" des conditions EM à un instant donné comme le préconisent les normes ferroviaires.

De plus, il faut également prendre en compte la fréquence d'apparition des transitoires qui, comme nous l'avons montré, est l'autre élément influant sur la qualité des communications GSM-R. Nous devons donc trouver un moyen de les prendre en compte, tout en nous assurant de ne considérer que les événements transitoires qui couvrent les bandes du GSM-R.

Nous avons donc besoin de suivre l'évolution temporelle des signaux de communication et transitoires afin de connaître leurs niveaux de puissance respectifs à chaque instant. Dans le même temps, il nous faut aussi connaître le contenu spectral des événements transitoires qui se produisent pour savoir si le bruit EM qu'ils génèrent couvre les fréquences du GSM-R. Ceci n'est pas réalisable avec une étude uniquement temporelle d'un côté et fréquentielle de l'autre. Il nous faut travailler dans les deux domaines simultanément en réalisant une analyse temps-fréquence.

Cette analyse sera réalisée, en post-traitement, sur des enregistrements temporels effectués à l'aide d'un oscilloscope numérique afin d'utiliser un véritable équipement de mesure répondant aux exigences des normes CISPR [1.22]. Le fait d'utiliser un équipement de mesure complètement indépendant du système GSM-R, que l'on cherche à protéger, nous permet de rester dans une démarche CEM. Dans la section suivante, nous décrivons l'approche que nous envisageons pour prédire l'impact des perturbations EM transitoires sur les transmissions GSM-R.

### *V.2 - L'approche proposée*

L'approche que nous proposons consiste à établir un lien entre les conditions de bruit EM dans lesquelles se déroulent les transmissions GSM-R et l'impact de cet environnement EM sur la qualité de ces transmissions. Notre proposition consiste à classer les bruits EM transitoires en fonction de leur impact sur la qualité d'une communication GSM-R. Plusieurs classes peuvent être définies selon le



niveau de dégradation de qualité. Dans un premier temps nous nous limiterons à deux classes : une première associée à une qualité de communication satisfaisante pour l'opérateur ferroviaire et l'autre à une qualité en-deçà de celle attendue.

Cette méthodologie s'inspire de techniques de classification habituellement utilisées dans d'autres domaines que la CEM, tels que l'acoustique ou le traitement d'images. Elle repose entièrement sur des enregistrements temporels (collectés à l'aide d'un oscilloscope numérique) des signaux de communication échangés entre le mobile et son réseau (ou simulateur de réseau pour les mesures effectuées au sein du laboratoire) en présence de signaux transitoires. A ces enregistrements est appliqué un post-traitement basé sur l'utilisation d'un descripteur de bruit EM dont l'extraction de certains paramètres permet d'estimer l'impact des perturbations EM sur la qualité d'une communication radio numérique sol-trains.

Le calcul du descripteur sera réalisé à partir des données issues d'une transformée temps-fréquence (TF), comme illustré sur la figure 1.12. Les différentes étapes de cette approche seront décrites en détail dans le chapitre 4.

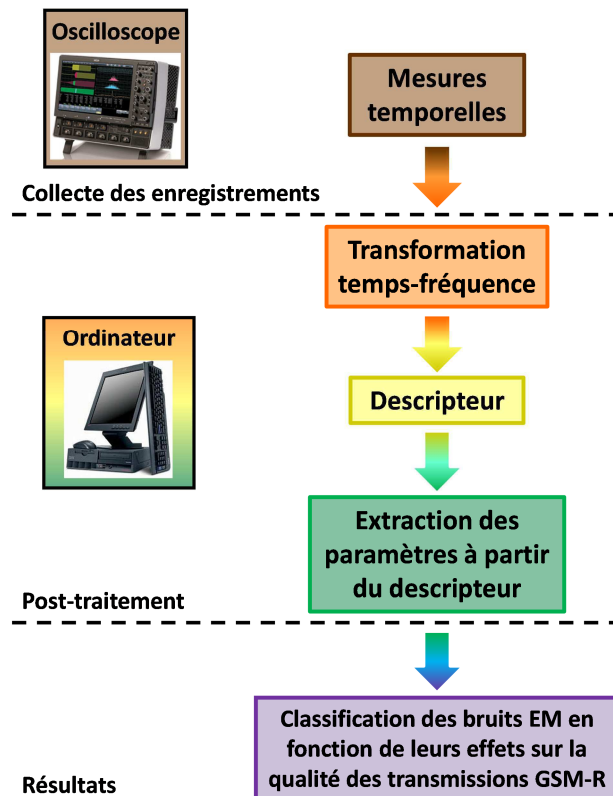


Figure 1.12. Les différentes étapes de l'approche proposée dans cette thèse

## Conclusion

Les modèles de bruit EM construits par Middleton permettent de modéliser tout type d'interférence EM quel qu'en soit le mécanisme à l'origine. Middleton a proposé d'appliquer ces modèles à l'étude des systèmes de communication. Toutefois, la portée de l'étude concerne la détermination de récepteurs optimaux qui minimisent les erreurs lorsqu'il s'agit de déterminer s'il y a présence ou absence d'un signal parmi le bruit ou d'estimer certaines données d'un signal bruité. La méthode que nous visons dans ce travail de thèse doit cependant permettre d'identifier et de quantifier l'impact produit par les perturbations EM sur la qualité des communications GSM-R. Ne pouvant reprendre directement les travaux de Middleton, nous nous sommes donc intéressés dans un premier temps aux normes ferroviaires et CEM existantes afin de procéder à une première approche de l'environnement EM rencontré par les équipements.

Habituellement, pour l'étude de la CEM dans le domaine ferroviaire, on se réfère dans un premier temps aux normes ferroviaires qui décrivent les méthodes à appliquer pour la mesure des émissions électromagnétiques et l'étude de l'immunité de l'ensemble des équipements et des installations. Dans un deuxième temps, ce sont les normes CEM qui sont considérées. Ces normes s'intéressent principalement aux défauts électriques ou électroniques introduits, sur les équipements, par les perturbations EM. Or, pour un système de communication numérique comme le GSM-R, c'est le lien radio qui doit être étudié et non simplement l'électronique, ce qui est assez éloigné de ce qui se pratique habituellement en CEM. De plus, la méthode de mesure des émissions rayonnées décrite dans les normes ferroviaires n'est pas adaptée pour caractériser les conditions EM dans lesquelles le GSM-R sera amené à fonctionner. Pour étudier la susceptibilité des nouveaux systèmes de transmissions numériques en général, il nous faut ainsi proposer une nouvelle approche, plus spécifique et qui prenne en compte leurs spécificités. Cela suppose de se tourner vers de nouveaux outils peu ou pas utilisés actuellement en CEM.

Dans le chapitre 2, nous présentons et décrivons ces outils, à savoir les transformées temps-fréquence ainsi que les méthodes de classification. Nous en rappelons le principe général et passons en revue différentes méthodes. Nous y expliquons et justifions également nos choix en matière de transformées TF et de méthode de classification retenues.

# Chapitre 2

## Rappels sur les méthodes de classification et les transformées temps-fréquence/temps-échelle

### SOMMAIRE

---

<b><i>I - La classification</i></b>	<b>30</b>
<b><i>II - Principe général d'une transformation temps-fréquence</i></b>	<b>37</b>
<b><i>III - Présentation de différentes transformations TF existantes</i></b>	<b>38</b>
<b><i>Conclusion</i></b>	<b>50</b>

---

La caractérisation de l'environnement EM ferroviaire nécessite de tenir compte des caractéristiques des perturbations EM que l'on y rencontre. Dans le cadre de nos travaux, nous nous intéressons aux perturbations EM transitoires issues du contact glissant entre le fil caténaire et le pantographe afin de mesurer leurs impacts sur la qualité d'une communication GSM-R. Ces dernières présentent deux particularités principales : elles sont de courte durée ( $\leq 20$  ns) et par conséquent large bande.

Lorsqu'elles sont observées dans le domaine fréquentiel, il apparaît qu'elles produisent des interférences (ou bruit) EM qui couvrent de larges bandes de fréquences, parmi lesquelles celles employées par le système de communication sol-trains GSM-R. Dans le domaine temporel, en raison de leur durée brève, il est nécessaire d'avoir une résolution temporelle suffisamment élevée pour être capable de les observer. Or, les expérimentations menées précédemment ont montré qu'étudier ces perturbations EM transitoires séparément en fréquence et en temps ne suffit pas. Il est nécessaire de relier le moment d'apparition d'un événement transitoire et sa durée aux fréquences couvertes par l'évènement. C'est pourquoi nous proposons de collecter les signaux transitoires à l'aide d'un oscilloscope utilisant une fréquence d'échantillonnage élevée et d'appliquer à ces enregistrements un traitement basé sur l'utilisation d'une transformée temps-fréquence. Nous utilisons une méthode de classification pour prédire l'impact des bruits EM transitoires sur la qualité d'une communication GSM-R.

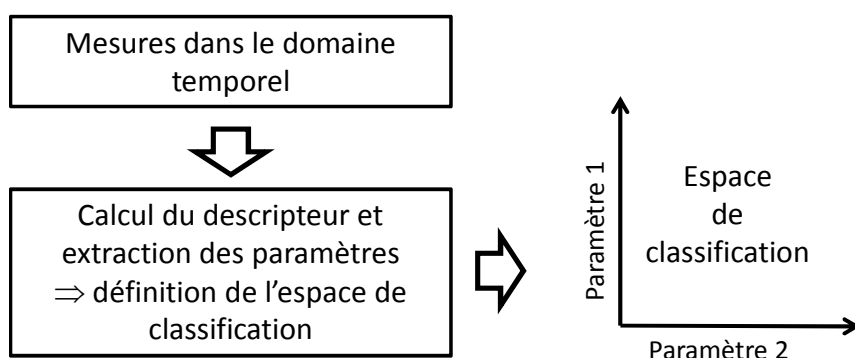
Ce chapitre est consacré au rappel des principaux éléments théoriques mis en œuvre dans ce travail de thèse en matière de classification et de transformées temps-fréquence/temps-échelle. Nous le débutons par une introduction à la classification, puis nous décrivons le principe général d'une transformation temps-fréquence et nous terminons par une présentation de différentes transformées existantes.

## I - La classification

### *1.1 - Principe général de la classification*

La classification peut être définie comme la recherche d'une répartition d'un ensemble d'éléments en plusieurs catégories. Chaque catégorie, appelée classe, regroupe des individus partageant des caractéristiques semblables. L'objectif est d'obtenir des classes les plus homogènes et les plus distinctes possibles.

Identifier des catégories nécessite de définir soigneusement un espace dans lequel le problème de classification doit être résolu. Un tel espace est souvent représenté par des vecteurs de paramètres extraits des éléments à classer et la classification est réalisée en adoptant une approche probabiliste, discriminative, neuronale ou encore stochastique. Une description précise de ces méthodes est fournie dans [2.1]. La figure 2.1 décrit le principe de la classification dans le cas particulier d'un espace de deux paramètres.



**Figure 2.1. Illustration du principe de la classification dans le cas où deux paramètres sont utilisés pour définir l'espace dans lequel le problème de classification doit être résolu**

De nombreuses méthodes de classification ont été proposées dans divers contextes d'applications tels que la reconnaissance de la parole [2.2], l'analyse des signaux électro-encéphalographiques [2.3] ou la reconnaissance de modulation d'un signal de télécommunication [2.4].

Dans la section suivante, nous présentons quelques-unes des méthodes de classification les plus utilisées.

### *1.2 - Présentation de différents classifieurs existants*

Il existe deux types de méthode de classification : les méthodes dites "supervisées" et dites "non-supervisées".

#### 1.2.a - Méthodes de classification supervisées

Les méthodes supervisées opèrent en deux étapes. La première étape, dite étape d'apprentissage, consiste à caractériser chaque classe à partir d'un premier ensemble de données dont la classe est connue. Cet ensemble constitue la base d'apprentissage. La seconde étape a pour objectif de déterminer la classe à laquelle appartiennent des éléments inconnus à partir des modèles estimés lors de l'étape précédente.

D'un point de vue général, les méthodes supervisées se différencient selon qu'elles adoptent une approche [2.5] :

- discriminative qui consiste à localiser les classes dans l'espace des paramètres (ou son espace transformé) en estimant leurs frontières. C'est le cas par exemple, de la règle des  $k$  plus proches voisins ( $k$ -ppv), des réseaux de neurones ou encore des Séparateurs à Vaste Marge (SVM, Support Vector Machine).
- générative, pour laquelle on cherche à estimer les paramètres de la distribution caractérisant chacune des classes considérées dans l'espace des paramètres retenu. Ces distributions sont très souvent caractérisées par une loi gaussienne ou un mélange de plusieurs de ces lois (GMM pour Gaussian Model Mixture).

Dans les paragraphes suivants, nous décrivons succinctement le principe d'un certain nombre d'entre elles.

*Règle des  $k$  plus proches voisins*

La règle des  $k$  plus proches voisins consiste à attribuer à un élément la classe à laquelle appartiennent les  $k$  éléments de référence (dont la classe est connue) d'une même classe dont il est le plus proche en termes de distance euclidienne. Par exemple, sur la figure 2.2 les points noirs et blancs sont les éléments de classe connue qui constituent l'ensemble d'apprentissage. L'élément inconnu est matérialisé par le point vert. Pour déterminer sa classe d'appartenance, il faut calculer toutes les distances entre ce point et chacun des points de la base de référence, de manière à déterminer les  $k$  voisins les plus proches parmi les points noirs et les points blancs. Sur cet exemple, en considérant que  $k$  vaut 5, les 5 voisins, d'une même classe, les plus proches du point vert sont des points noirs. Donc, l'élément inconnu appartient à la même classe que les éléments matérialisés par les points noirs.

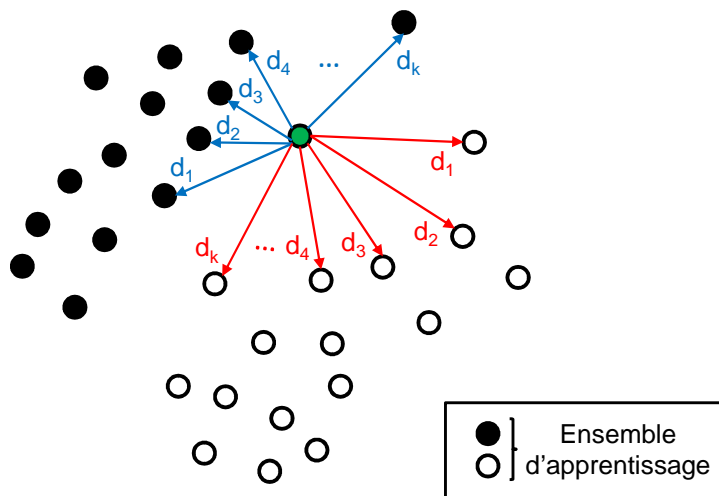


Figure 2.2. Règle des  $k$  plus proches voisins dans une espace de dimension 2

*Réseaux de neurones*

Cette méthode est inspirée par le fonctionnement des neurones biologiques. Le réseau de neurones est une structure composée d'une succession de couches de neurones, où chaque neurone d'une couche est connecté par des poids modifiables aux neurones de la précédente. La valeur de sortie d'un neurone sert ainsi au calcul de la valeur d'entrée des neurones de la couche suivante. La valeur d'entrée d'un neurone de la couche  $k$  est égale à la somme pondérée des sorties des neurones de la couche  $k-1$ , comme le montre la figure 2.3. Un seuil est ensuite appliqué à cette valeur d'entrée ( $e_{k,l}$  sur la figure 2.3) pour déterminer si le neurone est actif ou non et ainsi définir sa valeur de sortie  $s_{k,l}$ .

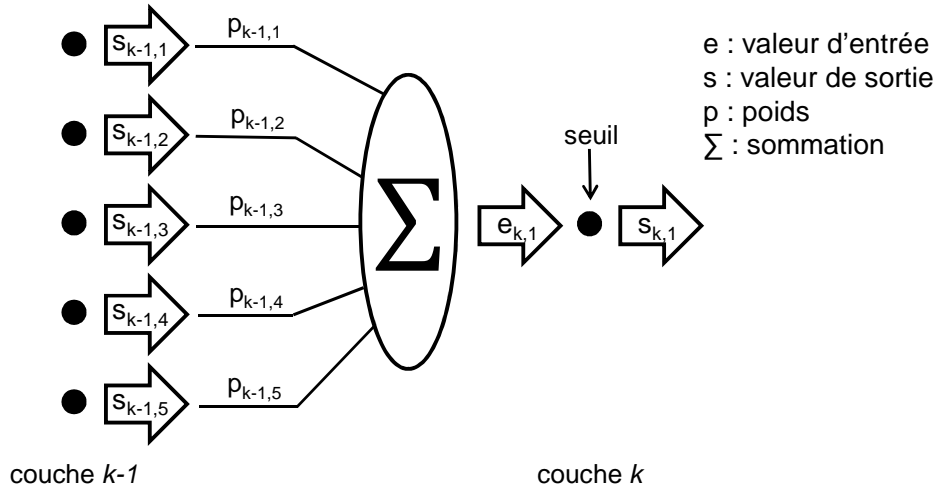


Figure 2.3. Relation entre les couches successives de neurones

Durant la phase d'apprentissage, on applique en entrée du réseau de neurones des données pour lesquelles on connaît, à l'avance, les valeurs de sortie (classes), comme sur la figure 2.4. Il s'agit ensuite de définir à l'intérieur du réseau, pour chaque neurone, les poids optimaux à appliquer pour retrouver les valeurs de sortie que l'on est supposé obtenir à partir de ces données d'entrée. Une fois la phase d'apprentissage terminée, le réseau est capable de classer automatiquement les nouveaux éléments de classe inconnue qui lui sont fournis en entrée.

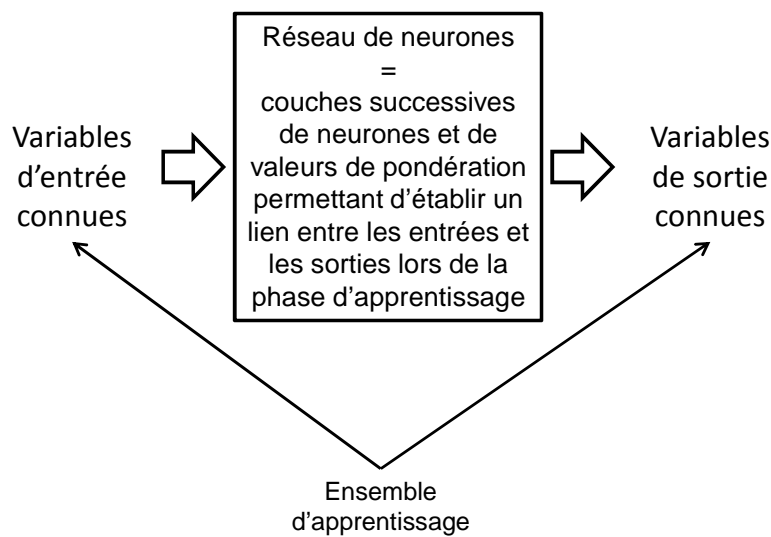


Figure 2.4. Principe de la phase d'apprentissage sur un réseau de neurones

### Séparateurs à Vaste Marge (SVM)

Le principe est de trouver une frontière séparatrice optimale, à partir d'un ensemble d'apprentissage. Cette frontière est constituée par un hyperplan. L'hyperplan optimal est celui qui maximise la distance (appelée marge) entre l'hyperplan et les échantillons les plus proches, appelés vecteurs supports. La figure 2.5 propose une illustration du principe des SVM. Comme précédemment sur la figure 2.2, l'ensemble d'apprentissage est représenté par des points noirs et des points blancs.

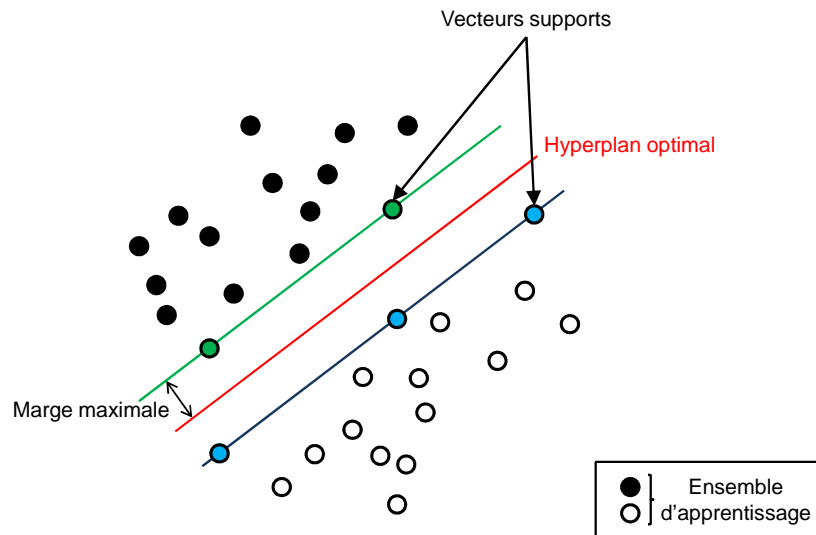


Figure 2.5. Principe des séparateurs à vaste marge

L'avantage de cette méthode réside dans le fait qu'une fois l'hyperplan déterminé, la classification des futurs éléments inconnus est réalisée uniquement à partir de leurs positions respectives par rapport aux vecteurs supports. Cela permet de réduire considérablement le nombre d'éléments à mémoriser dans la base de référence. Ainsi, sur la figure 2.6, dans le cas n°1 l'élément inconnu sera identifié comme appartenant à la classe 1 et dans le cas n°2, à la classe 2. Dans le cas n°3, il existe une indétermination étant donné que le point rouge est situé entre les deux lignes définies par les vecteurs supports.

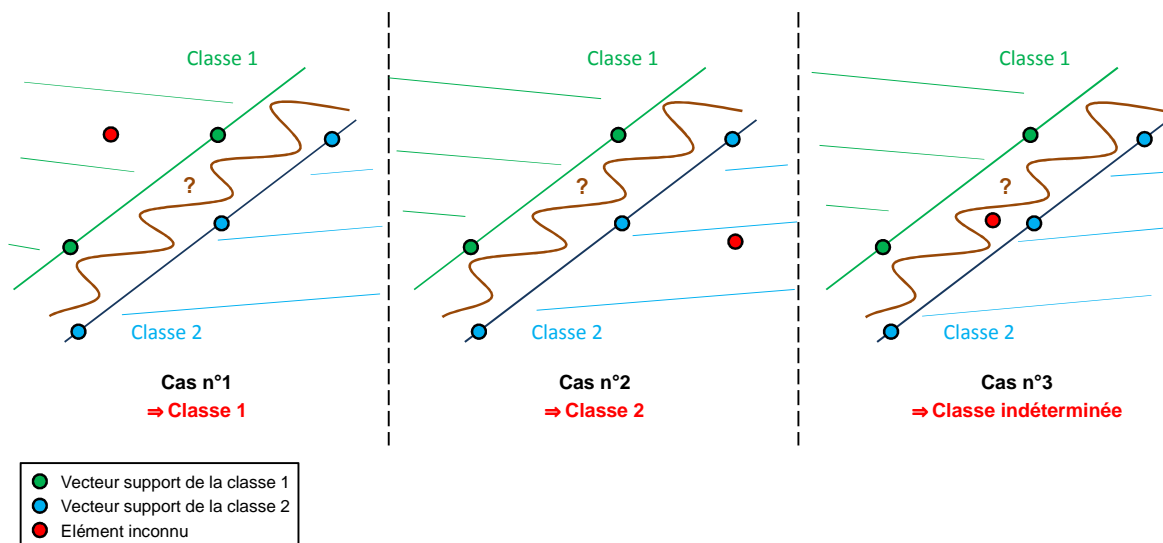


Figure 2.6. Classification d'un élément inconnu avec les séparateurs à vaste marge

### Gaussian Mixture Model (GMM)

Il s'agit de considérer qu'à l'intérieur d'une classe, la distribution des données de l'ensemble d'apprentissage suit une loi de probabilité dont la fonction de densité est un mélange de lois normales. Ainsi, la distribution des éléments d'une même classe est modélisée telle une somme de plusieurs gaussiennes. Pour chaque gaussienne, on commence par déterminer la variance et la moyenne. Ces paramètres sont ensuite optimisés de manière itérative, selon un critère de maximum de

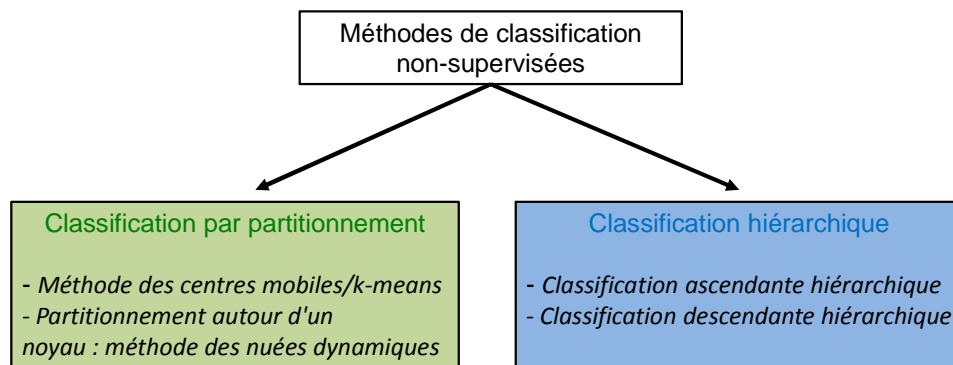
vraisemblance, afin de se rapprocher au mieux de la distribution réelle des éléments de référence de cette classe.

Une fois la distribution de chacune des classes ainsi approximée, on estime la probabilité d'appartenance d'un nouvel élément inconnu à chacune de ces distributions, et donc à la classe qui leur est respectivement associée. Finalement, on attribue à cet élément la classe à laquelle il a le plus de chances d'appartenir, c'est-à-dire la classe associée à la distribution qui a conduit à la probabilité d'appartenance la plus élevée.

### 1.2.b - Méthodes de classification non-supervisées

L'expression "non-supervisée" exprime le fait qu'il n'existe aucun "superviseur" pour identifier la classe à laquelle appartiennent les observations. La répartition d'un ensemble de données en plusieurs classes est réalisée sans connaissance a priori sur les classes recherchées. Comme les méthodes supervisées, ces techniques requièrent la définition d'un espace de paramètres dans lequel chaque élément à classer est projeté. Il s'agit donc de réaliser une classification "à l'aveugle" pour "deviner" comment sont distribuées les classes dans l'espace de classification choisi.

Les techniques de classification non-supervisées peuvent être regroupées en deux grandes familles : la classification par partitionnement et la classification hiérarchique, comme le montre la figure 2.7.



**Figure 2.7. Principales méthodes de classification non-supervisées**

Le principe général de chacune des méthodes citées en figure 2.7 est présenté rapidement dans les paragraphes suivants.

#### *Méthode des centres mobiles/k-means*

La méthode des centres mobiles est une méthode itérative qui nécessite une représentation dans l'espace des éléments à classer. Le nombre de classes  $k$  est déterminé a priori. On commence par sélectionner au hasard  $k$  éléments (un pour chaque classe). Ces éléments sont appelés centres. Une fois cette phase d'initialisation terminée, on répète les deux étapes suivantes, sur les autres éléments à classer, jusqu'à avoir atteint la variance interclasses maximale (cette valeur est atteinte lorsque la variance interclasses ne croît plus de manière significative entre deux itérations successives) :

- 1- Affecter chaque individu au centre (c'est-à-dire à la classe correspondante) le plus proche au sens de la distance euclidienne. A chaque itération, les éléments sont donc répartis en  $k$  classes, ou moins si, finalement, une des classes est devenue vide.
- 2- Calculer le centre de gravité, de chaque classe, qui devient le nouveau centre.



La méthode k-means reprend le même algorithme à la seule différence que les centres des classes sont recalculés à chaque affectation d'un individu à une classe. Cette modification rend l'algorithme plus efficace. En revanche la solution obtenue dépend de l'ordre dans lequel sont considérés les éléments à classer.

*Partitionnement autour d'un noyau : méthode des nuées dynamiques*

Il s'agit d'une variante de la méthode des centres mobiles. Elle consiste à remplacer chaque centre de classe par un noyau constitué d'éléments représentatifs de cette classe. Cette modification permet de corriger l'influence des valeurs extrêmes sur le calcul du centre de gravité.

*Classification ascendante hiérarchique*

Cette technique consiste à construire progressivement, et en commençant par le bas, un arbre (ou dendrogramme) avec les éléments à classer. Pour cela, on définit au départ autant de classes qu'il y a d'éléments. Ensuite, les éléments les plus proches sont regroupés deux par deux et ainsi de suite jusqu'à n'avoir plus qu'une seule classe à la racine de l'arbre. La figure 2.8 présente un exemple de dendrogramme et la méthode de détermination du nombre de classes [2.6]. Dans cet exemple, il s'agit de classer 47 grandes villes de France et de pays limitrophes selon leurs proximités relatives au sens de la distance routière. Le dendrogramme montre les regroupements successifs opérés (deux à deux, entre les deux villes ou groupements de villes les plus proches) jusqu'à la réunion, en une unique classe, de toutes les villes. La hauteur d'une branche est proportionnelle à la distance entre deux villes ou groupes de villes agrégés.

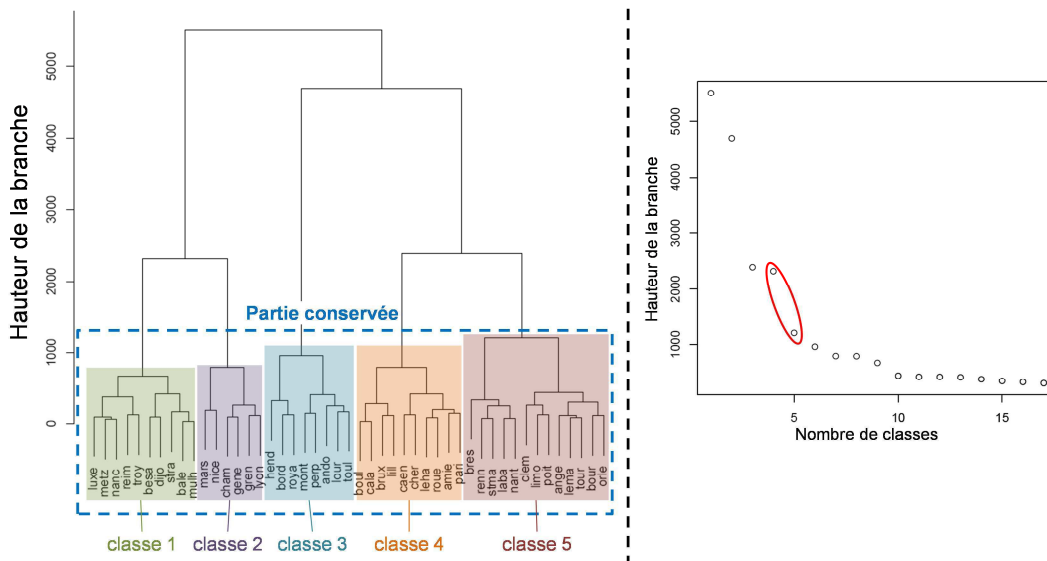


Figure 2.8. Dendrogramme (à gauche) et méthode de détermination du nombre de classes (à droite)

Le nombre de classes peut être déterminé de deux façons :

- directement à la vue du dendrogramme,
- à l'aide d'une représentation de la décroissance de la hauteur de la branche (soit de l'écart de distance) à l'issue de chaque regroupement en fonction du nombre de classes (à droite sur la figure 2.8). Le nombre de classes est défini en parcourant le graphique à partir de la droite. On choisit celui qui précède le premier saut significatif de la décroissance, soit dans notre exemple 5 d'après le graphe de droite de la figure 2.8.

Une fois le nombre de classes défini, on ne conserve, sur le dendrogramme, que la partie allant du bas de l'arbre (avec les 47 villes) jusqu'au regroupement en 5 paquets correspondant aux 5 classes.

### *Classification descendante hiérarchique*

Cette méthode reprend le même principe que celui de la classification ascendante hiérarchique mais il s'agit ici de fractionnement et non de regroupement. Initialement, tous les éléments sont rassemblés au sein d'une seule et même classe, par exemple A. L'élément qui se distingue le plus des autres (qui possède la plus grande dissimilarité) est affecté à une nouvelle classe, par exemple B, que l'on crée. Pour chaque élément de A, on détermine ensuite la dissimilarité moyenne à A, notée  $d(A)$ , et à B, notée  $d(B)$ . L'élément de A pour lequel la différence  $d(A)-d(B)$  est la plus grande est affecté à B. L'algorithme se poursuit sur les autres éléments de A tant que la plus grande différence  $d(A)-d(B)$  calculée est positive. Dès que ce n'est plus le cas, on passe à la classe B sur laquelle on applique le même traitement, en "sortant" les éléments les plus dissemblables, et ainsi de suite.

Pour la méthodologie de classification que nous proposons dans cette thèse, nous nous tournons vers une méthode supervisée employant une base de référence (ou d'apprentissage). Ce choix est motivé par le fait que nous disposons d'un banc de tests représentatif des conditions d'environnement EM observées en situations réelles, nous permettant de réaliser les différentes mesures nécessaires à sa construction. Cette thèse n'a pas vocation à comparer les performances que nous pourrions obtenir en appliquant toutes les méthodes présentées précédemment. Nous choisissons d'appliquer la règle des k plus proches voisins (avec différentes valeurs de k : 1, 3, 5 et 7) car, nous le verrons dans les chapitres suivants, cette règle s'avère des plus pratiques dans le cadre d'un problème à 2 dimensions comme le nôtre.

## II - Principe général d'une transformation temps-fréquence

L'analyse temps-fréquence fournit une représentation à la fois en temps et en fréquence des signaux non-stationnaires que l'on cherche à étudier. Elle constitue un outil de traitement du signal très puissant employé dans de nombreux domaines tels que l'acoustique, le biomédical ou encore le traitement d'images. L'intérêt de la transformation temps-fréquence est qu'elle permet de pallier l'inconvénient majeur de la transformée de Fourier qui est de représenter sous forme uniquement fréquentielle l'information contenue dans un signal temporel, entraînant ainsi la perte de la chronologie des événements (instant et fréquence d'apparition).

Le principe est de parcourir le signal temporel étudié avec une fenêtre de temps que l'on fait glisser sur la totalité du fichier. Cette fenêtre (de durée finie) permet de sélectionner une partie du signal sur laquelle on calcule un spectre. Une fois le premier spectre calculé, on fait glisser la fenêtre et l'on calcule les spectres suivants jusqu'à avoir balayé toutes les données du signal. Signalons que la façon dont est calculé chacun des spectres dépend de la transformation temps-fréquence utilisée. Finalement, la juxtaposition des différents spectres calculés permet d'obtenir l'évolution spectrale du signal au cours du temps. Cette méthode est illustrée par la figure 2.9.

Il est à noter que le résultat de la transformation temps-fréquence dépend de la fréquence d'échantillonnage ( $f_{ech}$ ) du signal étudié ainsi que de la largeur de la fenêtre de temps ( $d$ ) employée pour la sélection des données lors du calcul des spectres successifs. Les valeurs  $d$  et  $f_{ech}$  doivent ainsi être définies en fonction des bandes de fréquences à observer et de la résolution temps-fréquence souhaitée.

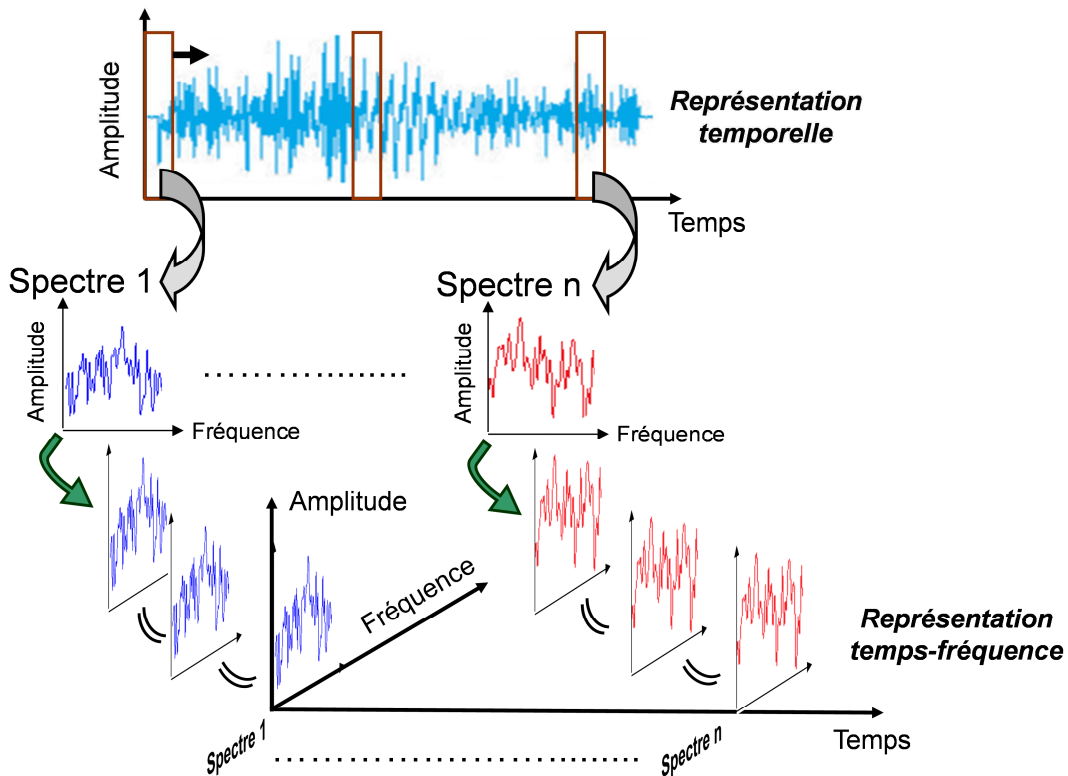


Figure 2.9. Principe de la représentation temps-fréquence

### III - Présentation de différentes transformations TF existantes

Il existe principalement deux grandes familles de représentations temps-fréquence (RTF) : linéaire et quadratique, comme illustré sur la figure 2.10 [2.7]. Dans la première, un signal quelconque est considéré comme une superposition linéaire de signaux élémentaires appelés atomes [2.8]. C'est le principe des décompositions atomiques (famille des représentations linéaires). Avec la seconde approche, l'énergie est répartie selon les deux variables temps et fréquence. C'est le principe des distributions d'énergie (famille des représentations quadratiques) [2.8].

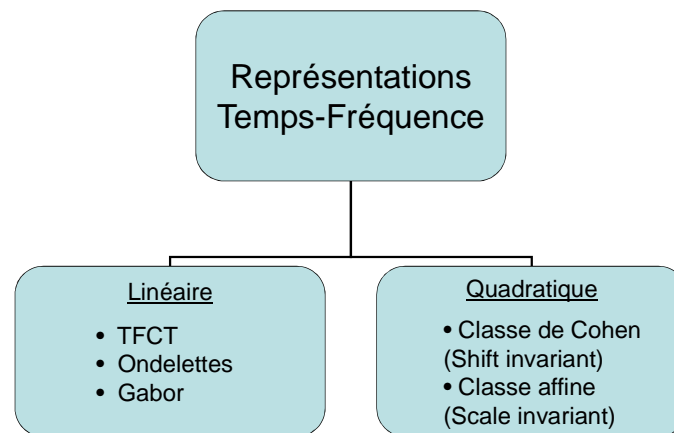


Figure 2.10. Familles de représentations temps-fréquence

Chacune de ces deux familles possède ses avantages et ses inconvénients. Le tableau 1 en dresse un bilan succinct [2.7].

	Représentations linéaires	Représentations quadratiques
<b>+</b>	<ul style="list-style-type: none"> <li>▪ Superposition linéaire : une RTF linéaire d'une combinaison linéaire de signaux est égale à la même combinaison linéaire des RTFs de ces signaux.</li> <li>▪ Absence de termes d'interférence pour des signaux multi-composantes</li> </ul>	<ul style="list-style-type: none"> <li>▪ Meilleure résolution temporelle et fréquentielle</li> <li>▪ Permettent d'obtenir des distributions d'énergie</li> </ul>
<b>-</b>	<ul style="list-style-type: none"> <li>▪ Dépendance entre résolutions temporelle (<math>\sigma_t</math>) et fréquentielle (<math>\sigma_f</math>) : <math display="block">\sigma_t \sigma_f \geq \frac{1}{4\pi}</math> (inégalité de Heisenberg)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Introduisent des termes d'interférence ("cross-terms") pour des signaux multi-composantes</li> </ul>

Tableau 1. Bilan comparatif des deux familles de représentations temps-fréquence

La section suivante présente et décrit successivement les principales transformées temps-fréquence existantes (définition, formule, exemples d'application) selon le regroupement présenté.

### III.1 - Représentations temps-fréquence linéaires

#### III.1.a - Transformée de Fourier à court terme (TFCT)

Egalement appelée Short-Time Fourier Transform (STFT) en anglais, il s'agit en fait d'une FFT glissante.

Pour une fonction  $x(t)$ , elle est définie à l'aide d'une fenêtre  $g(t)$  par la formule [2.9] :

$$TFCT_x^g(t, f) = \int_{-\infty}^{+\infty} x(t')g * (t'-t)e^{-2j\pi ft'} dt' \quad 2.1$$

Cette formule possède une forme équivalente dans le domaine discret [2.9] :

$$TFCT_x^g[k, l] = \sum_{k'} x[k']g * [k'-k]e^{-2j\pi k'l/L} \quad 2.2$$

avec un signal discret  $x[k]$ ,  $k \in \mathbf{Z}$  (ensemble des entiers relatifs)  
une fenêtre d'analyse  $g[k]$  de taille  $L$  et  $l=0, 1, \dots, L-1$

Dans les formules 2.1 et 2.2, \* désigne le complexe conjugué.

Que le signal étudié soit continu ou discret, cette transformation est linéaire et dépend de la fenêtre  $g$  employée. Le choix de la fenêtre d'analyse (type, largeur) est très important et dépend des caractéristiques attendues du signal à analyser. En ce qui concerne la largeur de la fenêtre, il faut garder à l'esprit qu'une fenêtre de petite taille entraîne une bonne résolution temporelle alors qu'une grande fenêtre entraîne, elle, une bonne résolution fréquentielle. Pour ce qui est maintenant du type de fenêtre utilisée, nous pouvons citer les fenêtres rectangle, de Hamming ou encore celle de Hanning. Notons cependant qu'afin d'éviter les fausses détections de pics dues aux lobes secondaires de la transformée de Fourier de la fenêtre, on évite généralement d'utiliser une fenêtre rectangulaire. De même, si le lobe principal de la transformée de Fourier de la fenêtre utilisée pour calculer la TFCT est trop large, il sera difficile de détecter deux composantes fréquentielles proches car leurs pics risquent de fusionner [2.9]. A titre d'illustration, une étude comparative entre plusieurs fenêtres est présentée dans la section III.1.c de ce chapitre. L'inconvénient majeur de la transformée de Fourier à court terme est que la taille de la fenêtre d'analyse est fixe, ce qui a pour effet de figer la résolution, comme on peut le voir sur la figure 2.11 [2.10].

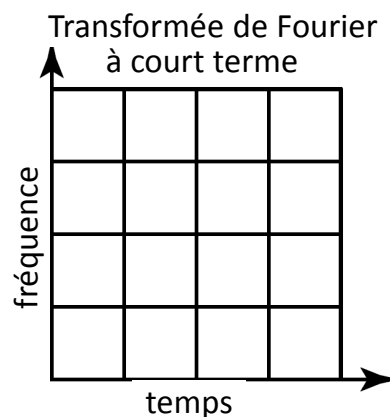


Figure 2.11. Résolutions en temps et en fréquence de la transformée de Fourier à court terme

Parmi les applications courantes de la TFCT, nous pouvons citer l'analyse de signal, l'identification de système, l'estimation spectrale ou encore la détection de signal [2.11].

### III.1.b - Transformée de Gabor

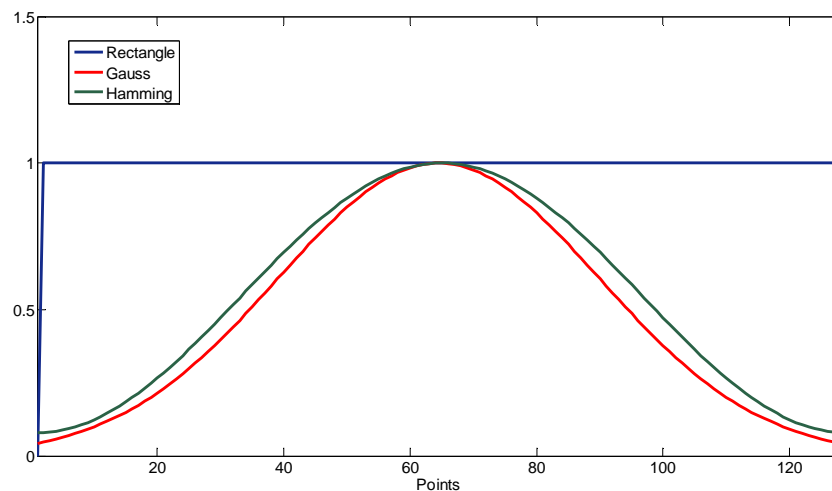
La transformée de Gabor constitue un cas particulier de la TFCT puisqu'il s'agit d'une transformée de Fourier à court terme avec une fenêtre gaussienne [2.12].

Cette transformée possède de nombreux avantages, notamment celui d'être localisée en temps et en fréquence de par l'utilisation d'une fenêtre gaussienne [2.13]. La transformée de Fourier d'une fenêtre gaussienne ne comporte pas de lobe secondaire car elle est, elle aussi, gaussienne [2.12], ce qui limite les problèmes de fausses détections de pics [2.9]. En fait, on récupère un coefficient non nul pour un instant d'apparition et une fréquence (ou bande de fréquences si le phénomène étudié est large bande) particulière. La précision de localisation (en temps) peut être contrôlée par l'ajustement de la largeur de la fenêtre associée [2.13]. Par conséquent, la transformée de Gabor s'avère particulièrement adaptée à la modélisation de transitoires [2.13].

La transformée de Gabor trouve son application dans des domaines aussi nombreux que variés : analyse de signaux radar, analyse de sons musicaux, analyse de la parole, détection de transitoires [2.13], [2.14].

### III.1.c - Comparaison de fenêtres

Nous nous intéressons à trois types de fenêtre : rectangle, Gauss et Hamming. Leurs représentations temporelles sont fournies par la figure 2.12 pour des fenêtres d'une largeur de 128 points.



**Figure 2.12. Représentations temporelles des trois fenêtres considérées**

Nous cherchons à comparer les performances qu'elles permettent d'obtenir en termes de calcul de spectre. Pour ce faire, nous commençons par calculer leurs transformées de Fourier respectives et les tracer, après normalisation, sur le même graphe, présenté en figure 2.13.

La fenêtre rectangle présente, comme nous l'avions précédemment expliqué, des lobes secondaires dont l'amplitude est importante (-13 dB pour le premier). D'ailleurs, malgré une décroissance conséquente, les lobes secondaires les plus extrêmes présentent des amplitudes encore supérieures à celles atteintes par les premiers lobes secondaires des fenêtres de Gauss et Hamming. En effet, avec ces dernières, l'amplitude des premiers lobes est de l'ordre de -44 dB. Notons également que pour la fenêtre de Gauss, mis à part le premier lobe secondaire pour lequel il y a égalité, tous les suivants sont d'amplitude inférieure à ceux de la fenêtre de Hamming. Tout cela a bien évidemment des conséquences lors du calcul du spectre d'un signal, comme l'illustre la figure 2.14 où nous considérons un signal dans la bande de fréquences GSM.

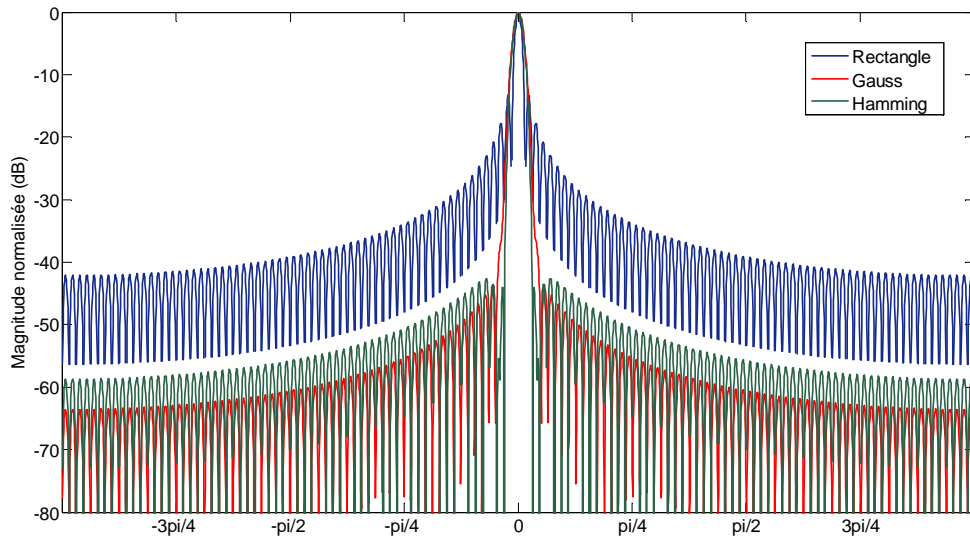


Figure 2.13. Transformées de Fourier des trois fenêtres considérées

Il s'agit d'un enregistrement de signal GSM public émis sur le canal 925.2 MHz avec une puissance de -40 dBm. Le spectre de ce signal temporel est calculé sur la bande 920.8-925.6 MHz en appliquant une TFCT avec une fenêtre de Gauss, rectangle et de Hamming

Nous constatons que chacune des trois fenêtres permet de retrouver la fréquence d'émission du signal puisque les trois maxima sont atteints à 925.2 MHz. Pour ces calculs, les fenêtres de Gauss et de Hamming n'ont pas été normalisées ; leurs gains respectifs sont de 0.51 et 0.54 alors que la fenêtre rectangle possède un gain de 1. C'est la raison pour laquelle seule cette dernière permet une estimation correcte du niveau de puissance du signal GSM. Toutefois les puissances estimées sur les canaux adjacents sont supérieures à ce qu'elles devraient à cause des lobes secondaires dont nous venons de mettre en évidence les fortes amplitudes. Ce phénomène est également visible avec les fenêtres de Gauss et de Hamming mais dans une moindre mesure. En effet, si les niveaux estimés sur le canal immédiatement adjacent demeurent quelque peu surestimés en raison de la largeur du lobe principal, le phénomène s'estompe rapidement grâce à la décroissance rapide de la puissance des lobes secondaires. Au final, dans ce cas d'étude, les fenêtres de Gauss et de Hamming offrent des performances similaires et satisfaisantes tant sur la localisation en fréquence que sur l'estimation de la puissance.

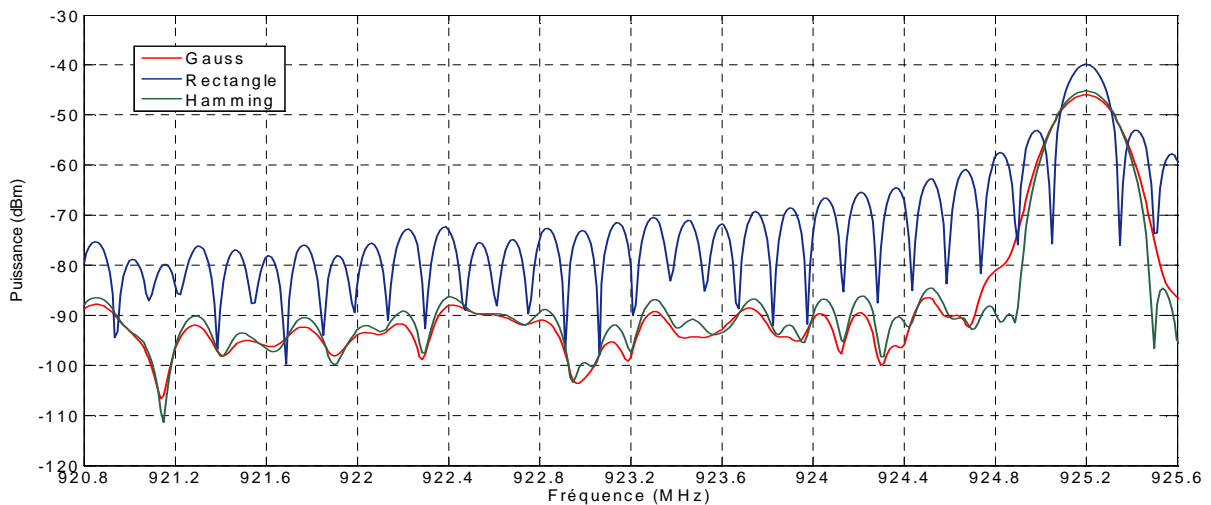


Figure 2.14. Comparaison des résultats obtenus lors de l'utilisation de ces trois fenêtres pour le calcul du spectre d'un même signal temporel

III.1.d - Transformée en ondelettes (WT : Wavelet Transform)

La transformée en ondelettes est une transformée temps-échelle. Il en existe trois types :

- continue,
- discrète,
- décomposition par paquets.

*Transformée en ondelettes continue (CWT : Continuous Wavelet Transform)*

Le principe est de décomposer un signal  $x(t)$  sur une base de fonctions translatées (en temps) et dilatées (par un facteur d'échelle 's', pour "scale") que l'on appelle ondelettes [2.9],

$$\psi_{t',s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-t'}{s}\right) \tag{2.3}$$

$\psi(t)$  est appelée l'ondelette mère. C'est par translation et dilatation de cette dernière que l'on obtient la base d'ondelettes  $\{\psi_{t',s}\}$ . Une ondelette est une fonction qui satisfait les conditions suivantes [2.15] :

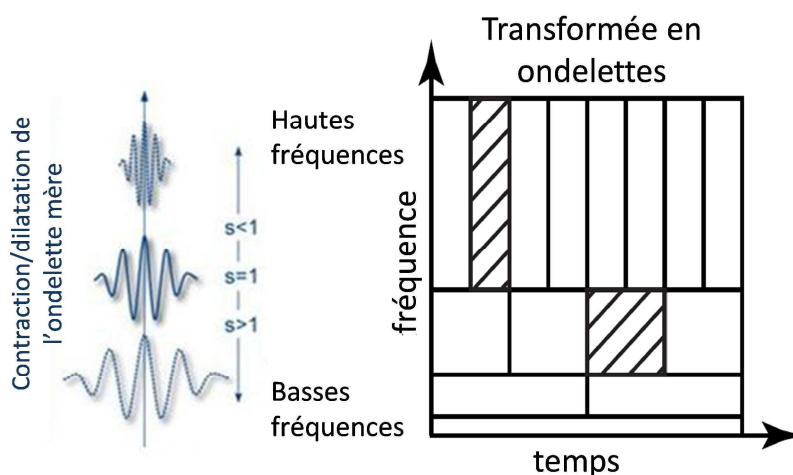
- Elle possède une structure courte, d'énergie finie, dans le domaine temporel,
- Elle présente quelques oscillations dans le même domaine,
- Elle doit être de moyenne nulle afin que la transformée en ondelettes soit inversible.

Le critère le plus important pour le choix d'une ondelette est de présenter, pour elle et sa transformée de Fourier, des oscillations les plus faibles possibles ; c'est ce qui permettra d'assurer une bonne résolution temporelle et fréquentielle [2.16].

La dilatation en amplitude, apportée par les termes  $s$  et  $\sqrt{s}$ , permet de conserver une norme constante pour tous les éléments de la base (énergie de l'ondelette) [2.16]. L'échelle utilisée,  $s$ , définit le niveau de compression de l'ondelette [2.17] :

- Petite échelle ( $s < 1$ ) → ondelette compressée → changements rapides → hautes fréquences
- Grande échelle ( $s > 1$ ) → ondelette dilatée → changements lents, caractéristiques grossières du signal mises en évidence → basses fréquences

La résolution temporelle varie proportionnellement à  $1/s$  et la résolution fréquentielle, proportionnellement à  $s$  [2.9], comme illustré sur la figure 2.15.



**Figure 2.15. Résolutions en temps et en fréquence de la transformée en ondelettes**



L'expression de la transformée en ondelettes est donnée par [2.9] :

$$WT_x^\psi(t',s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t) \psi * \left(\frac{t-t'}{s}\right) dt \quad 2.4$$

Le calcul de la transformée en ondelettes continue se déroule selon les étapes suivantes [2.18] :

1. L'ondelette est placée au début du signal à étudier (soit à  $t'=0$ ) et on fixe  $s=\alpha$ . En principe, il faudrait calculer la transformée pour toutes les valeurs de  $s$  supérieures et inférieures à  $\alpha$  (et ce jusqu'à  $\pm\infty$ ). Ici nous considérerons qu'il existe un nombre fini de valeurs de  $s$  (soit  $N_s$ ) et que la plus petite de ces valeurs (= niveau de compression le plus élevé) est égale à 1. On fera donc varier  $s$  de 1 à  $N_s$  et l'analyse sera donc réalisée des hautes fréquences vers les basses fréquences.

2. La fonction ondelette à l'échelle  $s=1$  est multipliée par le signal  $x(t)$  et intégrée sur toute la durée du signal (de  $t'=0$  à  $t'=D_x$ , la durée totale du signal étudié).

Le résultat est ensuite multiplié par  $1/\sqrt{s}$ . Cette multiplication a pour but d'effectuer une normalisation de l'énergie ; ainsi la transformée possèdera la même énergie à toutes les échelles.

3. On translate l'ondelette jusqu'à  $t'=T$  et on récupère la valeur de la transformée à  $t'=T$  et  $s=1$ .

4. On répète la procédure jusqu'à atteindre la fin du signal ( $t'=D_x$ ). Une "rangée" de points de la matrice de la transformée vient alors d'être calculée.

5. On incrémente  $s$  avec un pas suffisamment petit et on recommence les étapes précédentes pour toutes les valeurs suivantes de  $s$ , soit jusqu'à avoir atteint la valeur de  $s=N_s$ . On récupère alors la transformée sous la forme d'une matrice à  $N_s$  lignes et  $D_x + 1$  colonnes :

$$\begin{array}{c}
 t' = 0 \quad \dots \quad t' = D_x \\
 s = 1 \quad \left( \begin{array}{ccc} c_{1,1} & \dots & c_{1,D_x+1} \\ \vdots & \ddots & \vdots \\ c_{N_s,1} & \dots & c_{N_s,D_x+1} \end{array} \right) \\
 \vdots \\
 s = N_s
 \end{array}$$

Figure 2.16. Matrice des coefficients de la transformée en ondelettes

A chaque point  $(t',s)$ , dans le plan temps-échelle, le module du coefficient de la transformée fournit une information sur le degré de ressemblance entre le signal étudié et la version de l'ondelette  $\psi$  décalée de  $t'$ , à l'échelle  $s$  [2.15].

*Transformée en ondelettes discrète (DWT : Discrete Wavelet Transform)*

Il est important de ne pas confondre transformée en ondelettes discrète et discrétisation de la transformée en ondelettes continue. En effet, le fait d'utiliser un signal discret (ou de le discrétiser), comme nous l'avons supposé dans le paragraphe précédent, ne suffit pas. Il faut également employer un autre algorithme de calcul. La DWT est plus simple à implémenter que la CWT et nécessite un temps de calcul plus court.

Le principe est d'obtenir une représentation temps-fréquence d'un signal en utilisant une technique de filtrage numérique. On utilise des filtres avec différentes fréquences de coupure pour analyser le signal à différentes échelles. Pour analyser le signal en hautes fréquences, on le fait passer à travers une série de filtres passe-haut et inversement, pour analyser les basses fréquences, on le fait passer à travers une série de filtres passe-bas [2.19].

On suppose que l'on dispose d'un signal discret  $x[n]$  avec  $n=1, \dots, N$  entiers et échantillonné à la fréquence  $F_e$ . La DWT utilise deux fonctions appelées fonction d'échelle et fonction d'ondelette qui sont respectivement associées aux filtres passe-bas et passe-haut.

La transformée en ondelettes discrète est alors calculée en suivant la procédure suivante [2.19] :

1. Le signal original passe au travers d'un filtre passe-haut "halfband"  $g[n]$  et un filtre passe-bas "halfband"  $h[n]$ . C'est-à-dire qu'en sortie du filtre passe-bas, le signal contient les fréquences de 0 à  $F_{\max}/2 = F_e/4$ <sup>2</sup> et en sortie du filtre passe-haut, les fréquences de  $F_{\max}/2 = F_e/4$  à  $F_{\max} = F_e/2$ .

2. Toujours d'après le théorème de Nyquist-Shannon (voir note 2), la moitié des  $N$  échantillons peut être éliminée en sortie de chaque filtre. En effet, la fréquence maximale étant divisée par deux on peut en faire de même avec la fréquence d'échantillonnage et donc prendre  $F_e' = F_e/2$ , ce qui revient à ôter  $N/2$  échantillons.

Cette opération qui consiste à ne conserver qu'un échantillon sur deux en sortie de chaque filtre s'appelle la décimation. Le nombre d'échantillons est alors divisé par deux, ce qui fait que la résolution temporelle est, elle aussi, divisée par deux puisque l'on dispose de deux fois moins d'échantillons pour "décrire" la durée totale du signal. Quant à la résolution fréquentielle, elle est multipliée par deux puisque la bande de fréquences couverte a été divisée par deux et que l'on conserve le même nombre de points de fréquence.

Mathématiquement, l'opération de filtrage suivie de la décimation s'exprime par :

$$y_h[k] = \sum_n x[n]g[2k - n]$$

2.5

$$y_b[k] = \sum_n x[n]h[2k - n]$$

où  $y_h$  et  $y_b$  désignent respectivement les signaux en sortie des filtres passe-haut et passe-bas après décimation.

Cela constitue un premier niveau de décomposition de la transformée en ondelettes discrète.

3. Le signal en sortie du passe-bas (fréquences de 0 à  $F_e/4$ ) passe à nouveau au travers de deux filtres passe-haut "halfband"  $g[n]$  et passe-bas "halfband"  $h[n]$ . En sortie du filtre passe-bas, le signal contient alors les fréquences de 0 à  $F_{\max}/2 = F_e/8$  et en sortie du filtre passe-haut, les fréquences de  $F_{\max}/2 = F_e/8$  à  $F_{\max} = F_e/4$ . Après décimation, le nombre d'échantillons est là encore divisé par deux, ce qui a pour conséquence de diminuer la résolution temporelle (division par deux). Dans le même temps, la résolution fréquentielle a de nouveau doublé.

4. Le processus se poursuit jusqu'à ce qu'il ne reste plus qu'un seul échantillon en sortie des filtres après décimation, c'est-à-dire lorsque l'on a atteint le niveau  $L$  tel que  $2^L = N$ .

---

<sup>1</sup> Un filtre "halfband" est un filtre qui "coupe" les fréquences qui sont supérieures (cas d'un filtre passe-bas) ou inférieures (cas du filtre passe-haut) à la moitié de la fréquence maximale contenue dans un signal. Par exemple pour un signal dont la fréquence maximale est 100 kHz, la fréquence de coupure est égale à 50 kHz, et par conséquent un filtre passe-bas éliminera toutes les fréquences qui excèdent cette valeur et inversement avec un passe-haut.

<sup>2</sup> D'après le théorème de Nyquist-Shannon ( $F_e \geq 2F_{\max}$ ), si le signal est échantillonné à la fréquence  $F_e$ , alors la fréquence maximale ( $F_{\max}$ ) qu'il contient est égale à  $F_e/2$ .

La figure 2.17 illustre le principe décrit précédemment et en donne une représentation sous la forme usuelle d'un arbre :

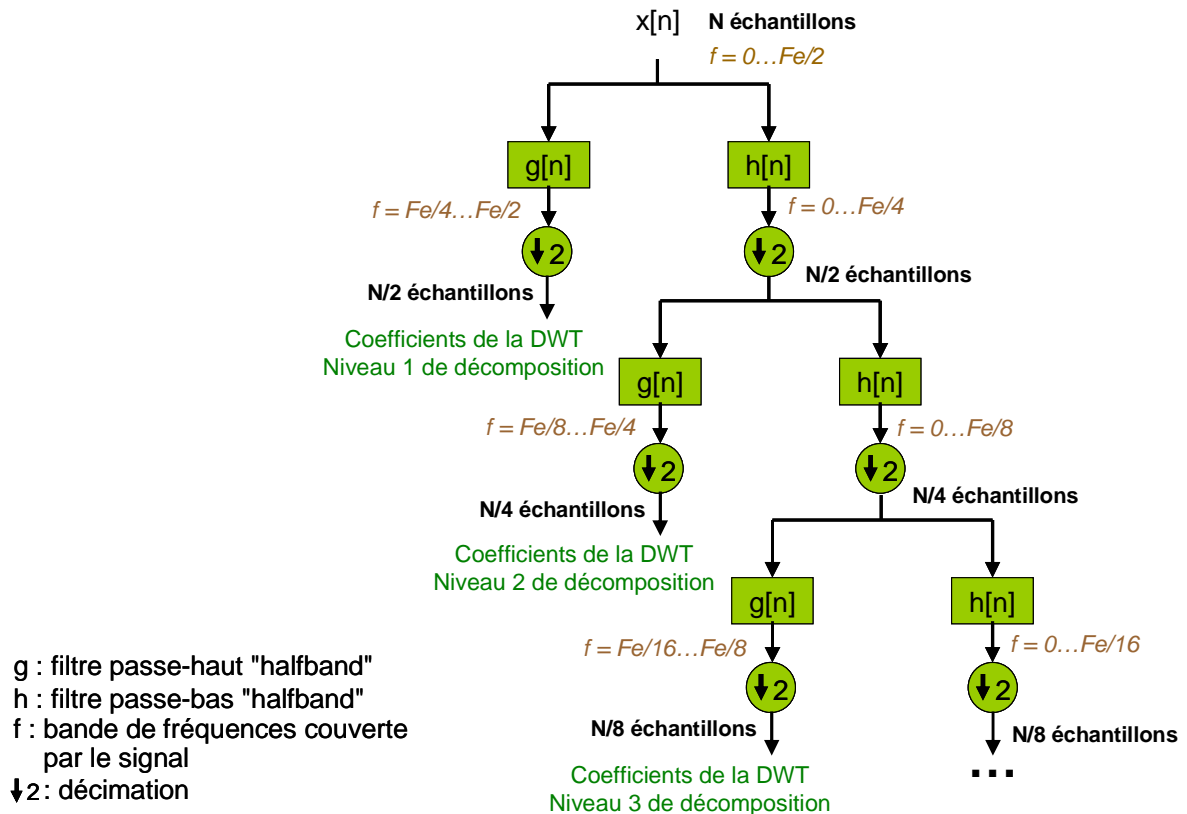


Figure 2.17. Schéma de principe de la transformée en ondelettes discrète

Une fois la décomposition terminée, on rassemble tous les coefficients qui constituent la DWT, c'est-à-dire tous les coefficients de chaque niveau en sortie des filtres passe-haut (appelés coefficients de détail) et le coefficient du dernier niveau de décomposition en sortie du filtre passe-bas (appelé coefficient d'approximation). Ce dernier correspond aux fréquences les plus basses ( $f = 0 \dots Fe/2^{(L+1)}$ ). En "remontant" l'arbre de décomposition jusqu'à la racine, on peut ainsi concaténer N coefficients, ce qui correspond bien au nombre N d'échantillons du signal initial.

Il est à noter qu'au fur et à mesure que l'on descend dans l'arbre, la résolution temporelle diminue (car il y a de moins en moins d'échantillons pour décrire le signal) alors que la résolution fréquentielle augmente (car la plage de fréquences couverte diminue à chaque niveau). Nous aurons donc une bonne résolution fréquentielle et une mauvaise résolution temporelle en basses fréquences et inversement pour les hautes fréquences.

*Décomposition en paquets d'ondelettes (WPT : Wavelet Packet Transform)*

Il s'agit également d'une transformation discrète. Elle reprend le même principe que la DWT, à savoir des filtrages successifs des échantillons du signal suivis de décimations. En revanche, à chaque niveau de décomposition, tous les coefficients sont à nouveau filtrés ; on ne se contente plus de filtrer les coefficients issus uniquement du filtre passe-bas [2.20].

Schématiquement, cela se traduit par :

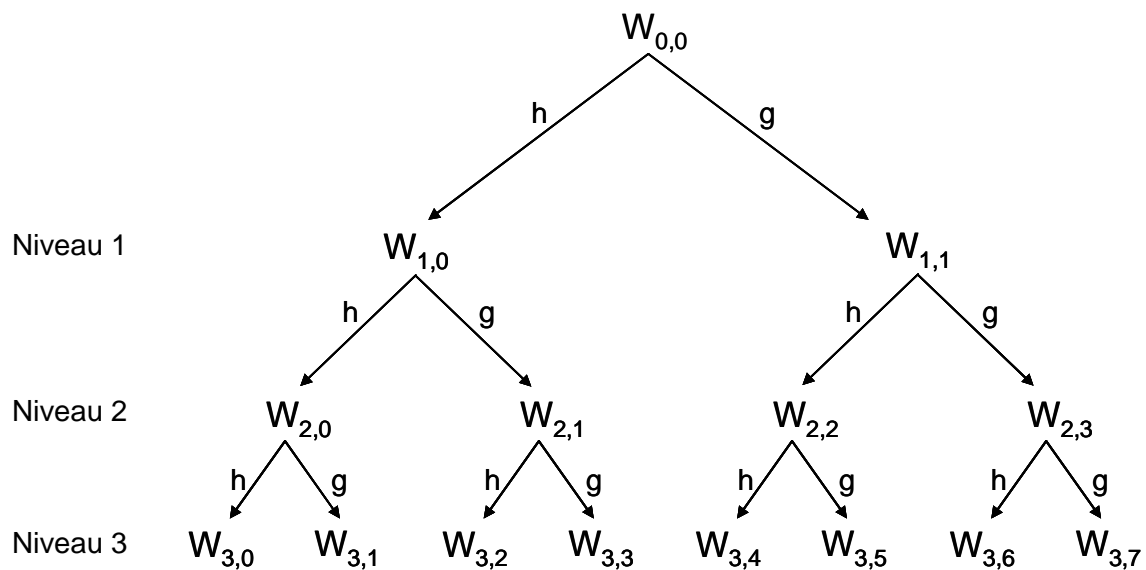


Figure 2.18. Arbre de décomposition en paquets d'ondelettes à l'ordre 3

Les "W" sont les coefficients de la WPT,  $h$  et  $g$  désignent respectivement les filtres passe-bas "halfband" et passe-haut "halfband". A chaque niveau et après filtrage, est réalisée une décimation qui n'est pas représentée sur le schéma de la figure 2.18.

Comme pour la transformée en ondelettes discrète, on constate une augmentation de la résolution fréquentielle et une diminution de la résolution temporelle à mesure que l'on descend l'arbre de décomposition. Nous obtenons à nouveau une bonne résolution fréquentielle et une mauvaise résolution temporelle en basses fréquences.

De plus, tous les coefficients de la décomposition ne sont pas "intéressants", dans le sens où certains ne contiennent que très peu (ou pas) d'information sur le signal et, par conséquent, ils peuvent être supprimés. Ce principe est notamment utilisé dans l'algorithme de recherche de la meilleure base ou du meilleur arbre ("best tree" de la Toolbox Matlab®).

Les applications de la transformée en ondelettes, qu'elle soit continue, discrète ou en décomposition par paquets, sont nombreuses : codage d'image et de signal, analyse de système, détection, traitement de signaux acoustiques ou sismiques, analyse de la parole, compression d'images et de signaux, débruitage.

### III.2 - Représentations temps-fréquence quadratiques

Les représentations temps-fréquence quadratiques regroupent les méthodes qui permettent d'obtenir des distributions d'énergie dans le plan temps-fréquence. Les transformées qui font l'objet de cette section, à savoir le spectrogramme et les distributions de Wigner-Ville et de Choi-Williams, font toutes les trois partie des éléments de la classe de Cohen. Le groupe de Cohen est constitué de toutes les représentations bilinéaires d'un signal distribuant son énergie dans le plan temps-fréquence et vérifiant les propriétés d'invariance par translation (temporelle et fréquentielle) et par modulation [2.9], [2.21].

Les éléments de la classe de Cohen sont ceux dont la représentation temps-fréquence peut s'écrire sous la forme [2.9] :

$$C_x(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi_{t-r}(t-v, \tau) x(v + \frac{\tau}{2}) x^*(v - \frac{\tau}{2}) e^{-2j\pi f \tau} dv d\tau \quad 2.6$$

où  $\phi_{t-r}(t-v, \tau)$  est appelé le noyau et \* désigne le complexe conjugué.

Le choix du noyau détermine la nature de la représentation temps-fréquence [2.22] :

- si  $\phi_{t-r}(u, \tau) = 1 \rightarrow$  distribution de Wigner-Ville
- si  $\phi_{t-r}(u, \tau) = e^{-u^2 \tau^2 / \sigma} \rightarrow$  distribution de Choi-Williams

### III.2.a - Spectrogramme (SPEC)

Le spectrogramme d'un signal  $x(t)$  est calculé à partir de la transformée de Fourier à court terme de ce signal :

$$\text{SPEC}_x(t, f) = | \text{STFT}_x(t, f) |^2 \quad 2.7$$

Il permet d'obtenir une forme de cartographie de l'énergie d'un signal donné à la fois en temps et en fréquence. De plus, il s'avère simple à construire et à interpréter, c'est pourquoi il est encore très utilisé aujourd'hui, notamment pour analyser des signaux lentement non-stationnaires (c'est-à-dire les signaux dont les caractéristiques spectrales varient lentement dans le temps) ou comportant un grand nombre de composantes [2.9]. Il est notamment très utilisé pour l'analyse de la parole [2.11].

### III.2.b - Distribution de Wigner-Ville (WD : Wigner-Ville Distribution)

La distribution de Wigner-Ville d'un signal  $x(t)$  s'exprime par [2.9] :

$$WD_x(t, f) = \int_{-\infty}^{+\infty} x(t + \frac{\tau}{2}) x^*(t - \frac{\tau}{2}) e^{-2j\pi f \tau} d\tau \quad 2.8$$

où  $x^*(t)$  est le complexe conjugué de  $x(t)$ .

Cette formule possède une forme équivalente dans le domaine discret [2.23] :

$$WD_x(n, f) = \sum_{k=-N}^{k=N} x(n + \frac{k}{2}) x^*(n - \frac{k}{2}) e^{-2j\pi f k / N} \quad 2.9$$

pour  $x(n)$  un signal discret défini pour  $0 \leq n \leq N-1$

Cette transformée est largement utilisée et sa simplicité d'implémentation fait qu'elle convient bien aux applications temps-réel. Elle produit des performances optimales pour un signal unidimensionnel alors que pour un signal comportant de multiples composantes, des termes d'interférence (ou termes croisés : "cross-terms") apparaissent. Ce phénomène est illustré sur la figure 2.19 qui présente la

distribution de Wigner-Ville d'un signal de saut de fréquence (frequency hopping) [2.24]. On y distingue clairement les deux fréquences couvertes (provenant des signaux "autoterms") ainsi que les termes d'interférence.

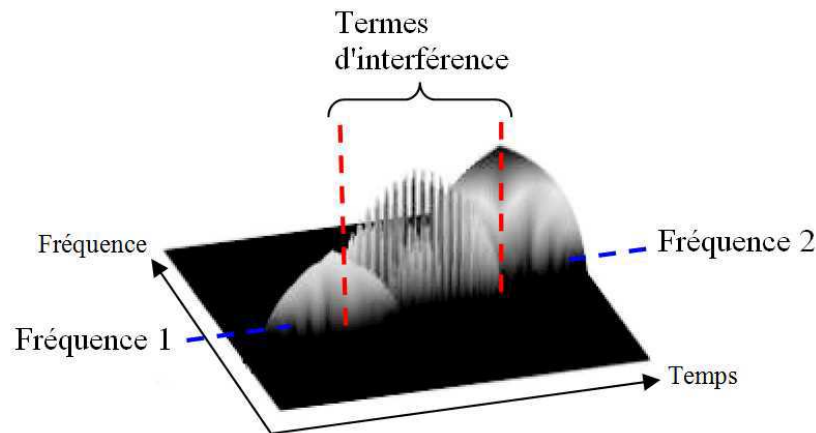


Figure 2.19. Distribution de Wigner-Ville d'un signal de saut de fréquence

Si l'on considère le profil de la distribution dans le cas d'un signal de longueur fixe (figure 2.19) et pour un déplacement sur l'axe du temps, la distribution suit une forme typique : elle augmente jusqu'au milieu de la fenêtre de temps où elle atteint son amplitude maximale, puis elle diminue [2.24]. Cette spécificité permet d'identifier le type de distribution utilisée.

### III.2.c - Distribution de Choi-Williams (CWD : Choi-Williams Distribution)

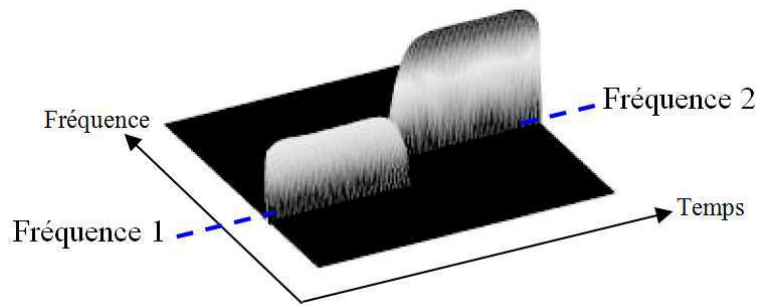
La distribution de Choi-Williams est une modification de la distribution de Wigner-Ville [2.25]. Sa formulation mathématique est la suivante [2.24], [2.26] :

$$CWD_x(t, \omega) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi\tau^2/\sigma}} e^{-\frac{(u-t)^2}{4\tau^2/\sigma}} x\left(u + \frac{\tau}{2}\right) x^*\left(u - \frac{\tau}{2}\right) e^{-j\omega\tau} du d\tau \quad 2.10$$

où  $x^*(t)$  est le complexe conjugué de  $x(t)$ .

Cette distribution réduit les effets d'interférence et de ce fait procure une meilleure visualisation du signal étudié dans le plan temps-fréquence. Cela est dû au noyau utilisé qui atténue les termes croisés. Ces derniers, bien que toujours présents, sont ainsi suffisamment atténués pour ne plus perturber l'analyse du signal [2.25]. Cependant, cela se traduit par une augmentation de la complexité d'implémentation.

Il est également à noter que le profil de la distribution est différent de celui obtenu avec la distribution de Wigner-Ville puisque dans ce cas, l'amplitude ne subit pas d'augmentation suivie d'une diminution sur la durée de la fenêtre de temps mais reste plate comme on peut le voir sur la figure 2.20 [2.24]. De ce fait les frontières de la distribution de Choi-Williams sont plus nettes que celles de la distribution de Wigner-Ville.



**Figure 2.20. Distribution de Choi-Williams du même signal de saut de fréquence que pour la distribution de Wigner-Ville**

La distribution de Choi-Williams présente plusieurs avantages :

- Elle présente une meilleure résolution temps-fréquence que le spectrogramme ou le scalogramme [2.27].
- Le paramètre  $\sigma$  permet de contrôler les termes d'interférence : plus  $\sigma$  est petit, plus le nombre de termes d'interférence qui sont adoucis et atténués est grand [2.26]. Cependant, si les termes d'interférence sont trop atténués, cela se traduit par une perte de résolution des signaux "autoterms" dans le plan temps-fréquence [2.25].

Si le paramètre  $\sigma$  est grand, alors la distribution se rapproche de la distribution de Wigner [2.26].

- En l'absence de bruit (seul est présent le signal que l'on cherche à analyser), les termes d'interférence génèrent des éléments caractéristiques de la signature du signal sur la distribution temps-fréquence. Alors que pour un signal bruité, ces termes ne font que produire des éléments indésirables [2.26].

### III.3 - Bilan

Tâchons maintenant d'établir un bilan succinct de ces transformées pour déterminer celle(s) qui serai(en)t la (les) mieux adaptée(s) à notre application. Rappelons que nous nous intéressons ici aux perturbations EM transitoires issues du contact glissant entre la caténaire et le pantographe.

En raison du caractère large bande des transitoires, une transformée de Wigner-Ville de ces signaux comporterait de nombreux termes d'interférence, ce qui rendrait son exploitation difficile.

Pour pallier ce problème, nous pourrions envisager d'employer une transformée de Choi-Williams, qui, nous venons de le voir, permet d'atténuer les termes d'interférences. Seulement, son implémentation est plus complexe, ce qui peut constituer un frein à une application temps-réel.

Le spectrogramme et la transformée de Fourier à court terme dont il découle peuvent être mis en place en utilisant différents types de fenêtres pour sélectionner les données sur lesquelles sont calculés les spectres. Il est notamment possible d'employer une fenêtre rectangulaire Or, nous l'avons montré, la transformée de Fourier d'une fenêtre rectangle donne un sinus cardinal qui, en plus de son lobe principal, contient plusieurs lobes secondaires dont l'amplitude peut-être suffisamment élevée pour permettre l'apparition de "rebonds" dans les spectres successifs du signal et donc sur la représentation temps-fréquence. De ce fait, les événements transitoires ne seraient pas correctement localisés sur la représentation et les puissances présentes sur les canaux adjacents pourraient être surestimées. L'utilisation, d'une fenêtre de Gauss (transformée de Gabor) ou de Hamming permet de résoudre ce problème. Comme la transformée de Gabor présente également l'avantage d'être bien adaptée à la modélisation et la détection de transitoires, c'est finalement la fenêtre de Gauss que nous retenons.

Restent donc finalement la transformée de Gabor et la transformée en ondelettes qui, comme nous l'avons également mis en évidence, convient, elle aussi, bien à la détection de signaux.

Dans le chapitre 3, nous décrivons deux exemples d'applications que nous avons développées et qui sont basées sur l'utilisation des transformées de Gabor et en ondelettes.

## Conclusion

Les deux principaux éléments théoriques sur lesquels repose ce travail de thèse ont été présentés. Une étude préliminaire des principales transformées temps-fréquence/temps-échelle a permis de mettre en évidence les apports et limitations de chacune. Au final, nous en avons retenu deux : la transformée de Gabor et la transformée en ondelettes qui, nous l'avons vu, conviennent bien à la détection de signaux et semblent les mieux adaptées à notre travail. Nous avons également décrit le principe général de la classification et présenté les méthodes de classification supervisées et non-supervisées à l'aide d'exemples d'application pratique. Nous avons choisi d'employer la règle des  $k$  plus proches voisins pour réaliser la classification des bruits EM transitoires. Cette règle fait partie des méthodes de classification supervisées qui nécessitent la construction d'une base de référence. Le banc de test mis en place au laboratoire (et qui sera introduit dans le chapitre 3) nous permet la construction de cette base. De plus, nous le verrons dans les prochains chapitres, notre espace de classification sera de dimension 2, ce qui fait de la règle des  $k$  plus proches voisins la méthode la plus simple et la plus rapide à implémenter.

Dans la bibliographie, il existe peu ou pas d'exemples d'utilisation des transformées temps-fréquence/temps-échelle et/ou de la classification en CEM. C'est pourquoi dans le chapitre 3 nous procédons à une mise en œuvre préliminaire de ces méthodes dans un contexte CEM à travers deux applications que nous avons développées.

La première concerne la détection de signaux transitoires noyés dans du bruit blanc gaussien. L'objectif est de détecter des événements transitoires masqués par le bruit EM contenu dans des fichiers de test que nous avons construits et déterminer leurs instants d'apparition respectifs. La méthode de détection proposée repose sur l'utilisation d'une transformée TF. Deux transformées différentes sont employées : Gabor et les ondelettes, avec pour objectif de comparer les performances obtenues par chacune d'entre elles en termes de détection.

Pour la deuxième application, il s'agit d'identifier le type de signal (bruit EM ou transitoire) contenu dans des enregistrements, sans visualisation préalable. Dans ce cas, l'ensemble de données initial est constitué d'enregistrements de signaux et les classes sont au nombre de deux : "transitoires" et "bruit". Pour réaliser la classification des signaux, nous utilisons, notamment, une décomposition en paquets d'ondelettes et la règle des  $k$  plus proches voisins. Cette méthode de classification supervisée nécessite de passer par une phase dite "d'apprentissage" durant laquelle est construite une base de référence, qui, in fine, sera le seul élément sur lequel reposera la classification des signaux inconnus.



# Chapitre 3

## Mises en œuvre préliminaires dans un contexte CEM

---

SOMMAIRE

---

<i>I - Application 1 : détection de signaux transitoires noyés dans du bruit blanc gaussien</i>	<b>52</b>
<i>II - Application 2 : classification de signaux</i>	<b>61</b>
<i>III - Bilan des deux applications préliminaires</i>	<b>72</b>
<i>IV - Principe de la méthodologie de classification proposée</i>	<b>73</b>
<i>Conclusion</i>	<b>77</b>

---

Nous rappelons que l'objectif de cette thèse est de proposer une méthodologie indépendante du système GSM-R à protéger qui permette de prédire l'impact des bruits EM transitoires (issus du contact caténaire-pantographe) sur la qualité d'une communication GSM-R et de les classer dans la catégorie correspondante (dégradation ou maintien d'une qualité suffisante). Les éléments théoriques intervenant dans ce travail de thèse ont été présentés au chapitre 2. Or, comme énoncé dans ce précédent chapitre, ces outils sont habituellement peu ou pas utilisés en CEM. Il est difficile d'en trouver des exemples d'application dans la littérature. Il devient donc nécessaire de procéder à des essais préalables de ces techniques afin de savoir si elles conviennent efficacement à notre problématique CEM. Dans ce but, ce chapitre présente deux applications mettant en œuvre les transformées temps-fréquence/temps-échelle et la classification dans un contexte CEM. La première est basée sur l'utilisation des transformées de Gabor et en ondelettes et la seconde, sur la décomposition en paquets d'ondelettes associée à la règle des k plus proches voisins. A l'issue de ces essais, nous établissons un bilan au cours duquel nous expliquons ce que nous retenons de ces deux applications pour développer la méthodologie de classification que nous proposons pour classer les bruits EM transitoires en fonction de leur impact sur la qualité des transmissions GSM-R.

## I - Application 1 : détection de signaux transitoires noyés dans du bruit blanc gaussien

### I.1 - Construction des fichiers de test

#### I.1.a - Principe

Le principe consiste à mélanger un fichier composé de plusieurs signaux transitoires (tous identiques) à du bruit blanc gaussien de façon à masquer les transitoires. Ainsi, les transitoires ne sont pas directement "visibles" sur le fichier temporel généré (taille totale : 131 072 points) et il est nécessaire de procéder à un traitement afin d'être capable de les détecter. La figure 3.1 donne un exemple de mélange « signal + bruit ». Nous constatons qu'aucun transitoire n'est visible directement cependant que le fichier en contient 5.

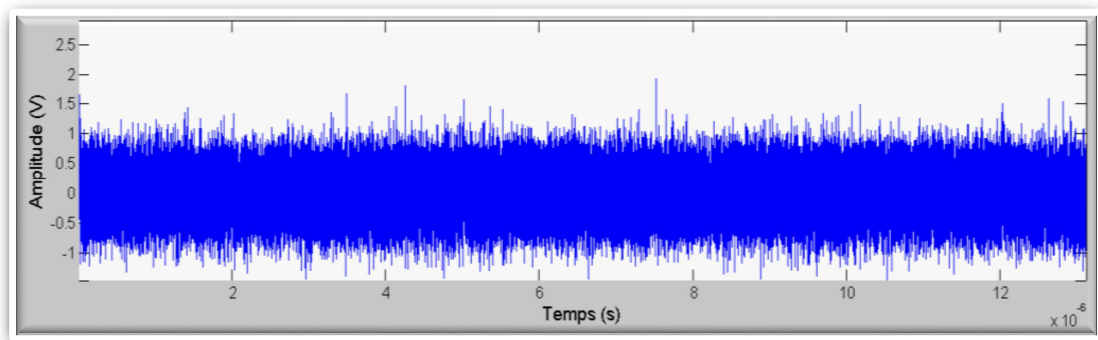


Figure 3.1. Mélange « signal + bruit » généré

Les signaux transitoires et le bruit sont tous les deux générés sous Matlab®. La figure 3.2 présente l'équation et la représentation temporelle du transitoire employé pour construire les fichiers de test.

$$S(t) = A \times (e^{-\frac{1}{D}t} - e^{-\frac{1}{RT}t}) \times u(t) \quad 3.1$$

Durée (Duration) :  $D = 5 \text{ ns}$

Temps de montée (Rise Time) :  $RT = 0.4 \text{ ns}$

$u$  : fonction échelon unitaire

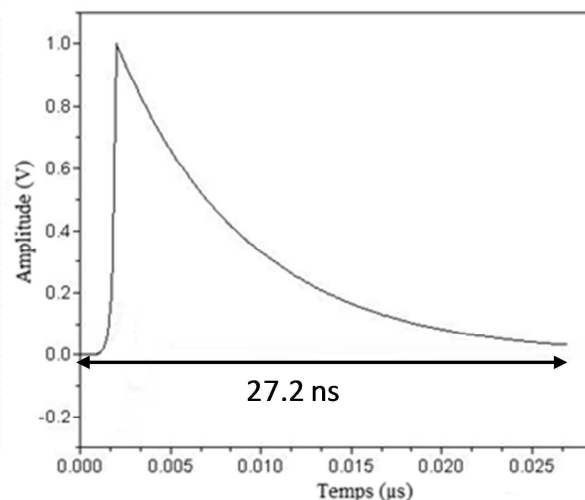


Figure 3.2. Signal transitoire employé

Précisons que les valeurs des paramètres  $D$  et  $RT$  proviennent des résultats présentés dans [3.1].

*1.1.b - Variation des paramètres*

Nous faisons varier trois paramètres lors de la construction des fichiers de test : le rapport signal à bruit, la durée totale du signal transitoire et sa récurrence. L'ensemble de variation de ces paramètres figure dans le tableau 2.

SNR (dB)	-9	-6	-3	-1.2	0	3
Récurrence	Rec1 fichier test	Rec2 1 µs	Rec3 0.5 µs	Rec4 0.1 µs	Rec5 27.2 ns	
Durée totale	D1 = 27.2 ns			D2 = 54.4 ns		

**Tableau 2. Ensemble de variation des paramètres des signaux de test**

*Remarque*

Le SNR est calculé sur la durée totale d'un transitoire (27.2 ns ou 54.4 ns) selon la formule :  
 $SNR = [ V_{RMS}(transitoire) / V_{RMS}(bruit) ]^2$ .

L'objectif est maintenant de détecter les signaux transitoires masqués par le bruit contenus dans nos fichiers de test. Cela se rapproche de ce qui se fait habituellement en traitement des signaux audio, lorsque l'on cherche à détecter, dans un enregistrement, par exemple un bruit inhabituel soudain parmi le bruit ambiant [3.2].

Ici, nous mettons en œuvre une méthode de détection basée sur l'utilisation d'une transformation temps-fréquence suivie de l'application d'un seuil sur la matrice des coefficients de la transformée. En fonction de ce que nous avons vu au chapitre 2, nous avons choisi d'utiliser une transformée de Gabor et une transformée en ondelettes. Nous cherchons, d'une part, à étudier l'impact des trois paramètres du tableau 2 sur notre algorithme de détection de signaux transitoires noyés dans du bruit blanc gaussien et, d'autre part, à comparer les résultats obtenus avec ces deux transformées.

*1.2 - Méthode de détection utilisant la transformée de Gabor ou les ondelettes*

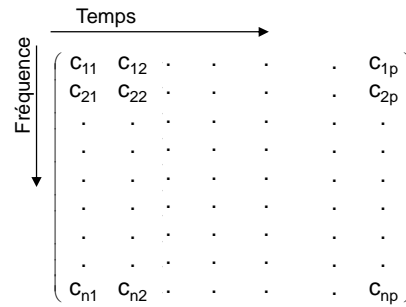
*1.2.a - Seuillage des coefficients de la matrice*

La première étape consiste à procéder au calcul de la transformée temps-fréquence de l'un des signaux de test que nous venons de créer.

Pour la transformée de Gabor, nous employons une fenêtre gaussienne de 256 points, ce qui, compte tenu de la fréquence d'échantillonnage (10 GHz), correspond à une durée de 25.6 ns. Le calcul est réalisé avec 4096 pas en fréquence.

La transformée en ondelettes est calculée avec l'ondelette de Haar en utilisant 128 pas d'échelle, valeur maximale acceptée par l'algorithme de calcul pour un fichier de 131 072 points. Il s'agit donc d'une transformée en ondelettes très performante.

Une fois la transformée temps-fréquence calculée, nous récupérons une matrice de coefficients dans laquelle chaque colonne correspond à une fenêtre (ou un point) de temps et chaque ligne est associée à une fréquence, comme illustré sur la figure 3.3.



**Figure 3.3. Matrice des coefficients de la transformée**

Dans cette matrice, il existe des coefficients qui sont représentatifs des signaux transitoires bruités et d'autres qui sont représentatifs du bruit seul. Pour notre application, nous avons besoin d'éliminer les coefficients qui sont caractéristiques du bruit uniquement. Cette opération est réalisée par l'application d'un seuil sur l'ensemble des coefficients de la matrice.

Le seuillage des coefficients issus d'une transformée temps-fréquence est une technique classique de débruitage des signaux. Le seuil que nous avons choisi d'appliquer est un seuil couramment cité dans la littérature et qui est défini par [3.3] :

$$th = \sigma \sqrt{2 \log_{10}(N)} \tag{3.2}$$

où N représente le nombre d'échantillons constituant le signal temporel de départ et  $\sigma$ , l'écart-type de ces mêmes échantillons.

Deux types de seuillage peuvent être mis en place : le seuillage dur et le seuillage doux [3.4]. Avec le seuillage dur (ou en anglais, hard thresholding), on ne conserve que les coefficients qui dépassent le seuil, sans les modifier, et on met les autres à 0 :

$$\text{Autrement dit, } \begin{cases} \text{si } |c_{ij}| \geq th \text{ alors } c_{\text{seuillé}_{ij}} = c_{ij} \\ \text{sinon } c_{\text{seuillé}_{ij}} = 0 \end{cases}$$

Avec le seuillage doux (ou en anglais, soft thresholding), on met à 0 les coefficients qui sont inférieurs au seuil et on recentre la valeur des coefficients conservés par rapport à la valeur du seuil :

$$\text{Autrement dit, } \begin{cases} \text{si } |c_{ij}| \geq th \text{ alors } c_{\text{seuillé}_{ij}} = \text{signe}(c_{ij}) (|c_{ij}| - th) \\ \text{sinon } c_{\text{seuillé}_{ij}} = 0 \end{cases}$$

Nous avons décidé de considérer un seuillage doux particulier et avons appliqué la condition suivante sur l'ensemble des coefficients  $c_{ij}$  de notre matrice :

$$\begin{cases} \text{si } |c_{ij}| \geq th \Rightarrow c_{ij} = 1 \\ \text{sinon } c_{ij} = 0 \end{cases}$$

Nous obtenons alors une nouvelle matrice (figure 3.4) composée uniquement de 0 et de 1, plus simple à exploiter. La perte d'information consécutive à ce seuillage est sans conséquence dans notre application, dans la mesure où nous souhaitons une simple détection de présence des transitoires.

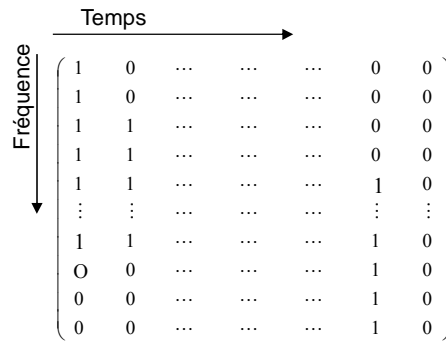


Figure 3.4. Matrice construite après seuillage des coefficients de la transformée

*1.2.b - Détection des positions des transitoires potentiels*

Pour identifier les colonnes susceptibles de contenir un transitoire, nous nous basons sur les deux observations suivantes :

- si une colonne de la matrice contient un 1, cela signifie que la fréquence correspondante est "couverte" avec un niveau suffisamment élevé pour dépasser le niveau de bruit défini par le seuil *th*.
- un transitoire est un évènement large bande, il couvre une large bande de fréquences.

Dès lors, un certain nombre de 1 consécutifs sur une colonne (le choix de ce nombre minimal, "delta" est expliqué ci-après), signifie qu'elle peut contenir un signal transitoire, comme le montre la figure 3.5. Le numéro de la colonne est rangé dans un tableau regroupant toutes les colonnes de détection potentielle et on passe à la colonne suivante jusqu'à avoir entièrement parcouru la matrice. Dans ce tableau est associé à chaque numéro de colonne remplissant les conditions de détection le nombre maximal de fréquences consécutives couvertes sur cette même colonne.

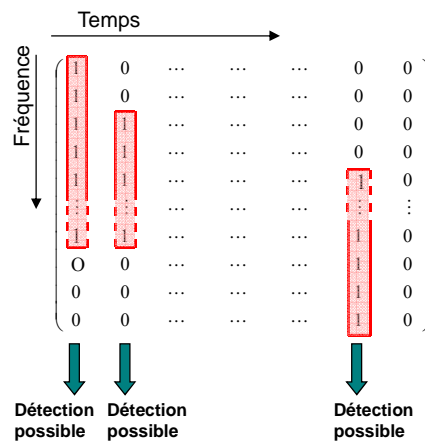


Figure 3.5. Identification des colonnes susceptibles de contenir un transitoire

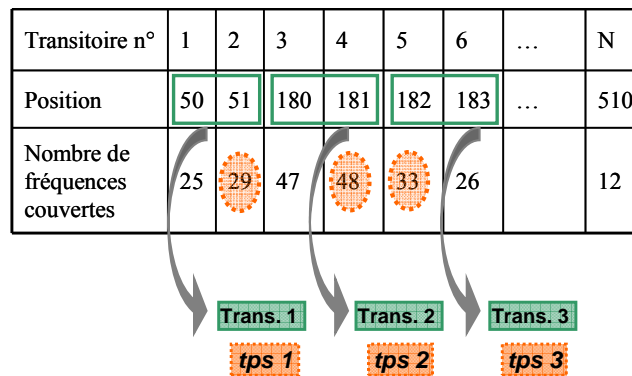
*Choix du nombre minimal de 1 consécutifs ("delta")*

Cette valeur n'est pas fixée complètement arbitrairement mais est évaluée pour chaque fichier suivant le principe suivant :

- on compte sur chaque colonne, le nombre maximal ("delta\_max") de fréquences consécutives couvertes.
- on calcule la moyenne des termes non nuls des "delta\_max".
- si cette valeur moyenne est  $\geq 10$ , alors on la conserve, sinon on prend la valeur delta = 10. La valeur 10 a été choisie de façon empirique, après plusieurs essais.

I.2.c - Regroupement des colonnes de détection successives et détermination des instants d'apparition associés

Comme nous l'avons noté précédemment, un transitoire présente la propriété de couvrir de larges bandes de fréquences et donc d'être observé à différentes fréquences. Puisqu'il possède également une certaine durée, il est aussi étalé dans le temps et par conséquent, apparaît dans notre cas, sur plusieurs colonnes successives. Ainsi, un même événement transitoire pourra être détecté sur plusieurs colonnes successives, entraînant alors des détections redondantes. Pour éviter cela, il convient de procéder à un regroupement des colonnes de détection, comme le montre la figure 3.6.



**Figure 3.6. Principe de regroupement des colonnes de détection et détermination des instants d'apparition associés**

On commence par choisir le nombre maximal ("nb\_fen") que l'on utilise pour le regroupement. Ce nombre est déterminé comme suit :

- on parcourt l'ensemble du tableau et on calcule le plus petit nombre de fenêtres successives que contient le tableau. Ainsi, on est sûr de pouvoir procéder au regroupement de toutes les fenêtres.
- si cette valeur est  $\leq 4$  alors on la conserve sinon, on choisit la valeur 4 (on procèdera alors au regroupement des fenêtres 4 par 4).

Ensuite, on parcourt une nouvelle fois le tableau avec la largeur de fenêtre ainsi choisie pour procéder au regroupement des colonnes de détection et déterminer les instants d'apparition des transitoires qui se sont produits. C'est en fait, à l'intérieur de chaque regroupement, la colonne sur laquelle on compte le plus grand nombre de fréquences consécutives couvertes qui détermine l'instant d'apparition.

*Cas particulier 1*

Si le nombre initial de transitoires détectés (N) n'est pas un multiple de nb\_fen, il reste des détections (au plus 3) qui ne sont pas traitées par l'algorithme, alors on les traite à part en essayant de regrouper les colonnes de détection avec la fenêtre la plus large possible.

*Cas particulier 2*

Si nb\_fen=1, alors toutes les détections supposées (ou potentielles) sont considérées, dans un premier temps, comme réelles aux instants d'apparition initialement identifiés.

I.2.d - Elimination des doublons

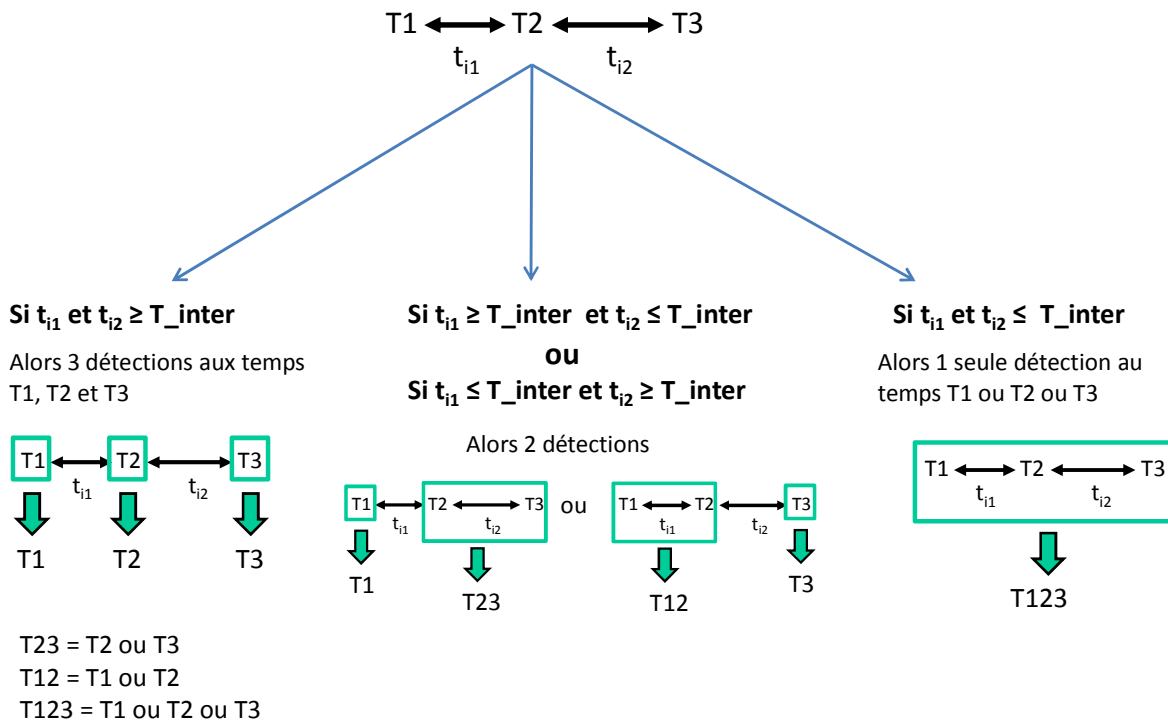
Pour éviter les fausses détections, il convient de s'assurer que l'intervalle de temps qui sépare chacun des transitoires ainsi détectés est supérieur à une valeur-limite soigneusement définie et ajustable si nécessaire. Les détections qui ne seraient pas suffisamment espacées sont "fusionnées" en une seule et leur instant d'apparition est déterminé suivant le même critère que précédemment (colonne qui contient le plus grand nombre de fréquences consécutives couvertes).

L'élimination des doublons éventuels a lieu en deux temps :

- par « fenêtre de trois »
- « un par un »

*Elimination par « fenêtre de trois »*

L'idée est de définir une valeur d'intervalle de temps minimum, "T\_inter", telle que pour trois transitoires successifs détectés respectivement aux instants T1, T2 et T3, on ait  $|T2 - T1| \geq T\_inter$  et  $|T3 - T2| \geq T\_inter$ . Autrement dit, cela revient à dire que pour tout transitoire détecté au temps T, il ne peut pas y avoir d'autre transitoire détecté dans l'intervalle  $I = [T - T\_inter ; T + T\_inter]$ . Si plusieurs transitoires étaient initialement détectés dans l'intervalle I, on procède à des fusions selon le principe illustré à la figure 3.7.



**Figure 3.7. Principe d'élimination des détections redondantes par « fenêtre de trois »**

*Elimination « un par un »*

Nous procédons à une deuxième élimination car, à l'issue de la première, il se peut que de fausses détections subsistent. Nous parcourons donc le tableau des détections en nous assurant que chaque transitoire est distant (en termes de temps) de celui qui le précède d'au moins T\_inter (on peut choisir d'utiliser une valeur différente de la précédente mais cela n'aurait que peu d'intérêt). Si ce n'est pas le cas, on "fusionne" les deux détections en une seule, selon le principe habituel. Ce processus de vérification de l'intervalle de temps entre deux transitoires successifs est appliqué en boucle jusqu'à ce que tous les transitoires détectés soient distants les uns des autres d'au moins T\_inter.

*1.2.e - Affichage des résultats*

La figure 3.8 illustre le résultat obtenu après la phase de détection. On y retrouve la représentation temporelle globale « signal + bruit » sur laquelle les transitoires sont "invisibles" et la représentation temps-fréquence employée (ici la transformée de Gabor) qui permet de faire apparaître les transitoires. L'interface fournit également le nombre de transitoires détectés ainsi que leurs instants d'apparition respectifs.

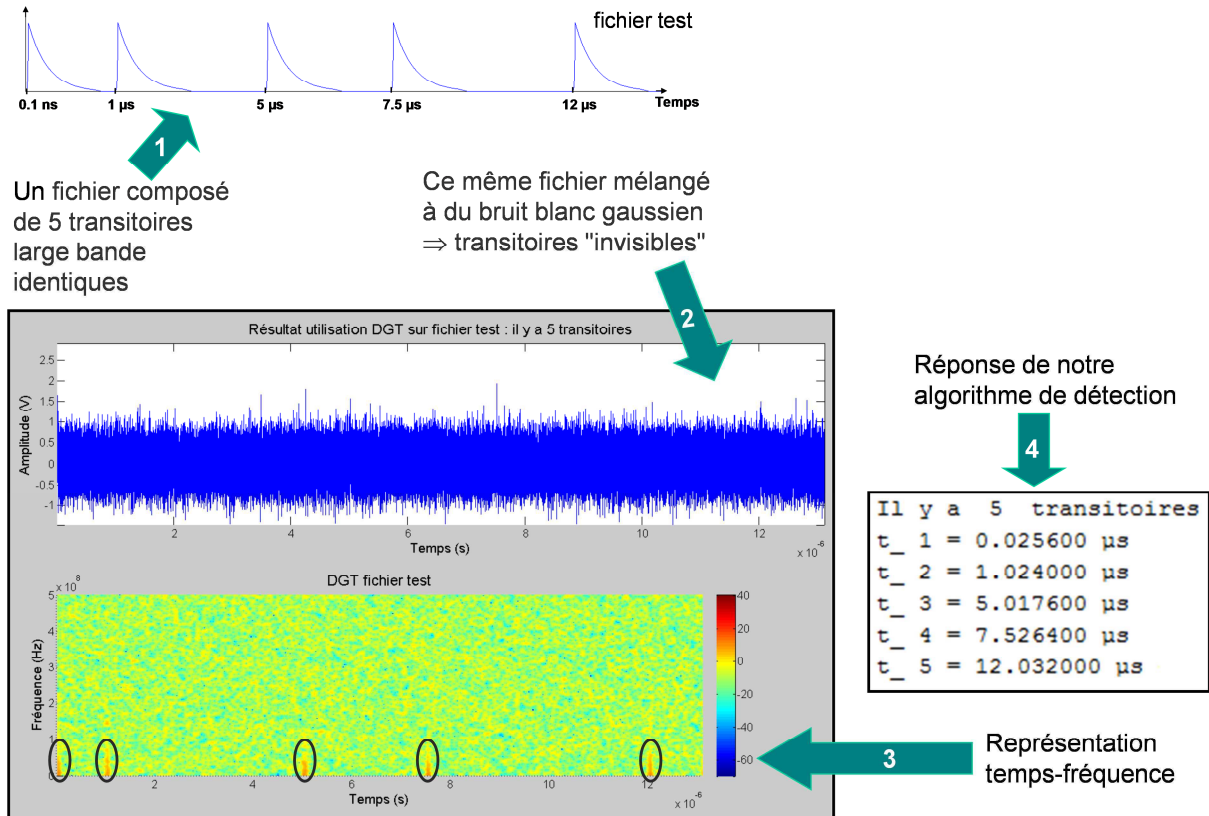


Figure 3.8. Affichage des résultats

*1.3 - Comparaison des résultats obtenus par les deux transformées*

La comparaison porte sur les résultats obtenus par notre méthode avec la transformée de Gabor d'une part, et la transformée en ondelettes d'autre part, dans diverses configurations de test. Afin de diversifier les configurations et tenter de généraliser nos résultats, nous faisons varier les trois paramètres précédemment introduits avec leur ensemble de variation dans le tableau 2 en page 53 :

- rapport signal à bruit (SNR),
- récurrence des signaux transitoires,
- durée (totale) des signaux transitoires.

L'objectif est de tester notre méthode pour différentes valeurs de ces trois paramètres afin d'étudier l'impact de chacun sur les performances de notre algorithme, en fonction de la transformée temps-fréquence employée.



I.3.a - Variations indépendantes des trois paramètres

Dans un premier temps, nous avons choisi de faire varier les paramètres un par un de manière à évaluer l'impact produit par chaque paramètre indépendamment des deux autres. Nous avons donc construit des fichiers de test pour lesquels un seul des trois paramètres varie à chaque fois et nous avons ensuite appliqué notre algorithme de détection avec la transformée de Gabor (DGT) puis, les ondelettes (CWT). A l'issue de chaque essai, nous avons noté le nombre de transitoires correctement détectés, ce qui nous a conduit aux observations présentées dans les trois paragraphes qui suivent [3.5].

*Variation du SNR*

Quand le SNR diminue, le bruit couvre davantage les transitoires, ce qui peut en affecter la détection. Nous constatons qu'avec notre méthodologie, la détection est réalisée sans problème jusqu'à atteindre un SNR de -3 dB avec les ondelettes et -1.2 dB avec Gabor. En-deçà de ces valeurs respectives, il n'est plus possible de détecter l'intégralité des transitoires.

*Variation de la récurrence des signaux transitoires*

Des problèmes de détections manquantes commencent à apparaître lorsque la récurrence des signaux atteint des valeurs élevées. En effet, avec la transformée de Gabor des détections sont manquantes à partir de Rec4 (respectivement Rec5 avec les ondelettes), c'est-à-dire lorsqu'un transitoire apparaît toutes les 100 ns, d'après les valeurs fournies dans le tableau 2 en page 53. On note également que la précision de la localisation temporelle de la détection diminue lorsque la récurrence augmente.

*Variation de la durée des signaux transitoires*

Nous avons choisi une durée D1 équivalente à la durée de la fenêtre utilisée pour la transformation de Gabor (25.6 ns) et une durée D2 deux fois plus longue. Que ce soit avec Gabor ou les ondelettes, aucun problème de détection n'a été causé par l'augmentation de la durée des signaux transitoires. Cela est dû au principe de regroupement des colonnes de détection qui permet d'éliminer les détections redondantes, comme nous l'avons expliqué précédemment.

I.3.b - Variations simultanées de deux des trois paramètres

Dans un deuxième temps, nous avons fait varier deux paramètres simultanément : le SNR et la récurrence des signaux transitoires. Nous ne nous sommes plus intéressés à la durée totale des signaux transitoires étant donnée qu'elle n'avait pas d'effet grâce à notre principe de regroupement des colonnes de détection. Nous nous sommes limités à 3 niveaux de récurrence (associés, par définition, chacun à un nombre connu de transitoires à détecter) et 5 valeurs de SNR. Pour chaque configuration de test, nous avons relevé le nombre de transitoires correctement détectés pour les deux transformées. Les résultats de ces essais sont présentés ci-après sous la forme d'un tableau comparatif (tableau 3).

SNR (dB) \ Rec	-6	-3	-1.2	0	3	Nbre de transitoires à détecter
1	0 3	3 5	5 5	5 5	5 5	5
3	1 12	22 26	26 27	27 27	27 27	27
4	13 47	77 100	94 128	113 131	114 131	131

Nombre de transitoires détectés en utilisant la DGT et les ondelettes

**Tableau 3. Bilan comparatif des résultats obtenus avec les deux transformées temps-fréquence**

Nous observons que, quel que soit le niveau de récurrence, à partir d'une valeur de SNR égale à 0 dB, les ondelettes permettent de retrouver le nombre exact de transitoires, ce qui n'est pas le cas avec la transformée de Gabor. Donc, dans le cas des ondelettes, c'est le rapport signal à bruit qui influe majoritairement sur les performances de la détection. En revanche, la récurrence (donc l'intervalle de temps entre deux transitoires successifs) influe nettement plus sur les résultats obtenus avec Gabor. Cela est dû à la largeur de la fenêtre utilisée pour procéder au regroupement des colonnes de détection. En effet, dans l'algorithme, nous les regroupons au maximum par 4 soit 104.2 ns (d'après la taille de la fenêtre gaussienne employée pour le calcul de la transformée : 256 points, soit 25.6 ns) et avec Rec 4 on obtient un intervalle de temps de 100 ns entre deux transitoires successifs. Il se peut donc que deux détections successives soient "fusionnées" en une seule du fait que la largeur de la fenêtre qui a servi à leur détection est plus importante que l'intervalle de temps qui sépare effectivement les deux signaux.

#### I.3.c - Bilan de la comparaison des résultats

A l'issue de ces simulations, nous extrayons quelques conclusions quant à cette méthode de détection de signaux transitoires noyés dans du bruit.

Premièrement, elle peut être mise en œuvre en utilisant l'une ou l'autre des deux transformations temps-fréquence retenues : la transformée de Gabor et les ondelettes.

Ensuite, l'étude comparative que nous avons réalisée permet de dégager les tendances suivantes :

- un léger avantage peut être concédé aux ondelettes qui semblent, en présence de bruit blanc gaussien, davantage robustes à la diminution du SNR et à l'augmentation de la récurrence des signaux transitoires.
- les ondelettes procurent une plus grande précision du temps de détection par rapport à Gabor dont la résolution en temps est, rappelons-le, limitée par la largeur de la fenêtre employée. Il serait donc possible d'améliorer la précision de notre algorithme en réduisant la largeur de la fenêtre gaussienne.
- les ondelettes ont toutefois l'inconvénient d'être plus "lourdes" à traiter, ce qui affecte la rapidité d'exécution de l'algorithme de détection.

Il faut cependant garder à l'esprit que les ondelettes sont "poussées" à l'extrême et qu'elles sont donc très performantes en termes de résolution temps-fréquence. Avec Gabor, en utilisant comme nous l'avons fait une fenêtre judicieusement définie par rapport à la bande de fréquences et aux signaux étudiés, nous obtenons des performances comparables avec un traitement nettement plus léger.

Au final, nous retenons que l'une ou l'autre des transformées peut être employée et procurer des performances comparables, à condition d'être correctement paramétrée.

## II - Application 2 : classification de signaux

Pour cette seconde application, il s'agit d'identifier la nature des signaux contenus dans des fichiers de mesures, sans les visualiser. Les travaux présentés dans cette section reprennent certains travaux présentés dans [3.6] et se basent sur l'utilisation de la décomposition en paquets d'ondelettes.

### II.1 - Principe de calcul

#### II.1.a - Calcul et représentation de la décomposition par paquets d'ondelettes

##### Fichiers utilisés et paramètres de la décomposition

Les fichiers utilisés pour cette application proviennent de mesures effectuées au-dessus d'une ligne de TGV, lors du passage des trains. Parmi l'ensemble des séries de mesures réalisées, seules deux sont sélectionnées pour cette étude (séries 1 et 2) car représentatives de l'ensemble. Ces dernières regroupent toutes deux des enregistrements de transitoires et de bruit. Chaque enregistrement comporte 4096 points. Pour les enregistrements de transitoires, ces 4096 points sont centrés autour de l'évènement transitoire contenu dans l'enregistrement.

Dans une première étape, une étude préalable à l'aide d'une transformation de Gabor a permis d'annoter "manuellement" la composition exacte des enregistrements des deux séries de mesures et de définir ainsi une vérité terrain :

##### Série 1

- 13 enregistrements de transitoires
- 37 enregistrements de bruit

##### Série 2

- 34 enregistrements de transitoires
- 18 enregistrements de bruit

Une transformation temps-fréquence est appliquée à chaque enregistrement. Il s'agit d'une décomposition en paquets d'ondelettes avec les paramètres suivants :

- ondelette : Daubechies 3 (Db3)
- niveau de décomposition (max) : 5 ce qui correspond à 63 nœuds

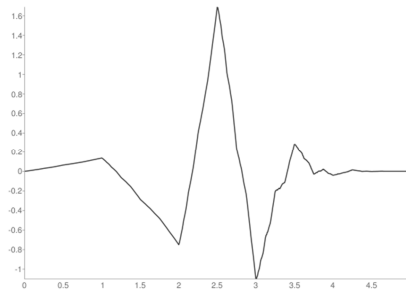


Figure 3.9. Fonction d'ondelette Db3

Nous décrivons maintenant pas à pas les différentes étapes suivies pour réaliser la classification des signaux. Les données dont nous disposons (enregistrements de 4096 points) sont séparées en 4 groupes, répartis en deux classes (bruit et transitoires) :

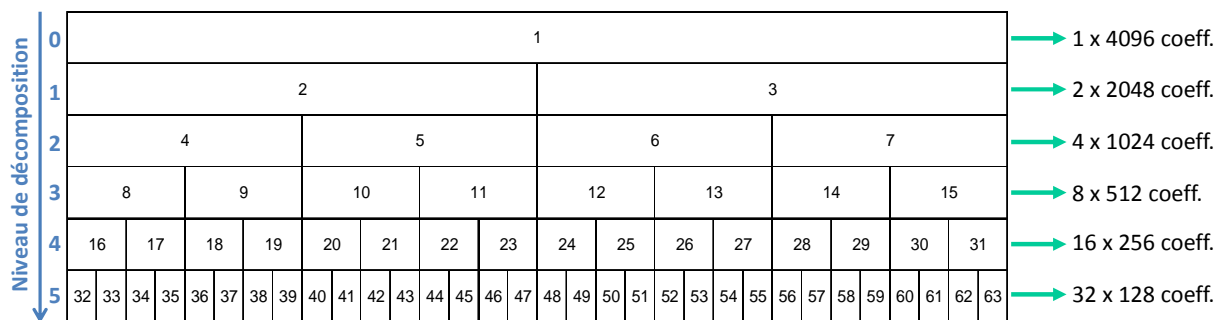
- bruit\_1 → contient les 37 enregistrements de bruit de série 1
  - bruit\_2 → contient les 18 enregistrements de bruit de série 2
- } classe du bruit
- transient\_1 → contient les 13 transitoires de série 1
  - transient\_2 → contient les 34 transitoires de série 2
- } classe des transitoires

L'objectif final est d'être capable d'identifier automatiquement la classe à laquelle appartient un enregistrement pris au hasard dans l'une des deux séries de mesures. Pour ce faire, un traitement identique est appliqué à chacun de ces 4 groupes séparément, comme nous le décrivons dans les paragraphes qui suivent.

*Représentation de la décomposition sous la forme d'un arbre*

Le même processus est appliqué aux 4 groupes introduits précédemment. Pour faciliter la compréhension, à chaque étape nous ne présentons que le traitement appliqué sur le groupe « transient\_1 » composé de 13 transitoires.

La décomposition par paquets d'ondelettes est appliquée à chaque enregistrement en employant les paramètres précédemment fournis. A l'issue de chaque transformation temps-fréquence, les coefficients issus de la décomposition sont rangés dans un tableau reprenant l'arbre de décomposition habituel, à savoir :

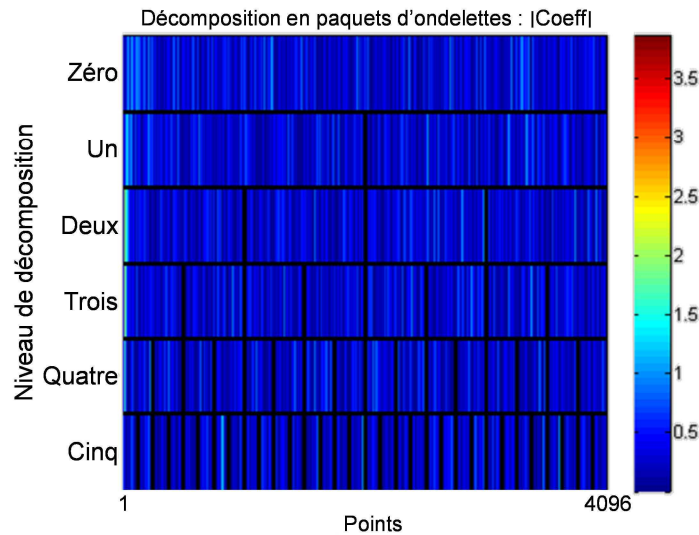


**Figure 3.10. Représentation de la décomposition sous la forme d'un arbre**

Ici nous avons 6 niveaux de décomposition (de 0 à 5), ce qui nous donne 63 nœuds (ou paquets de coefficients). A chaque niveau (soit encore à chaque ligne) correspond un nombre de nœuds. Or, comme le nombre de coefficients reste le même d'un niveau de décomposition à l'autre, chaque nœud contient en fait un nombre de coefficients qui dépend du numéro de ligne où il se trouve. Plus le numéro est élevé, plus il existe de nœuds et donc moins un nœud contient de coefficients.

Si la décomposition avait été menée jusqu'au bout, la dernière ligne aurait été composée de 4096 nœuds contenant chacun un coefficient. Nous n'avons pas jugé utile d'atteindre un tel niveau de décomposition pour le traitement que nous envisageons par la suite, c'est pourquoi nous avons opté pour un niveau égal à 5.

La figure 3.11 montre le résultat obtenu en représentant la valeur absolue des coefficients rangés par niveau de décomposition de la même manière que dans le tableau précédent (figure 3.10). Sur la figure 3.11, nous faisons également apparaître les lignes de séparation entre chaque paquet de coefficients ainsi que les lignes qui délimitent les niveaux de décomposition. Cette forme de représentation sera réemployée par la suite pour générer la cartographie de l'énergie, comme nous allons le montrer dans le prochain paragraphe.



**Figure 3.11. Représentation de la valeur absolue des coefficients de la décomposition sous la forme d'un arbre**

Ce procédé est ensuite appliqué sur les 12 autres enregistrements de transitoire et les tableaux de décomposition obtenus sont stockés.

### II.1.b - Calcul et représentation de la cartographie de l'énergie

#### *Calcul de l'énergie moyenne pour chaque nœud et représentation de la cartographie de l'énergie*

Il s'agit de repartir, pour chaque enregistrement, du tableau précédemment construit avec les paquets de coefficients et de déterminer la valeur de l'énergie moyenne de chaque paquet (ou nœud), en utilisant l'expression :

$$nrj(k) = \frac{1}{N_k} \sum_{i=1}^{N_k} y_k^2(i) \quad 3.3$$

avec  $y_k$  : vecteur des coefficients du nœud k  
 $N_k$  : nombre d'éléments du nœud k

Au final, nous récupérons 63 valeurs d'énergie moyenne par enregistrement :  $\{nrj(1), nrj(2) \dots nrj(63)\}$  dont une représentation, reprenant la disposition sous la forme d'un arbre, est donnée sur la figure 3.12. Ce procédé de calcul (et éventuellement de représentation) des valeurs d'énergie moyenne est répété autant de fois que le groupe contient d'enregistrements.

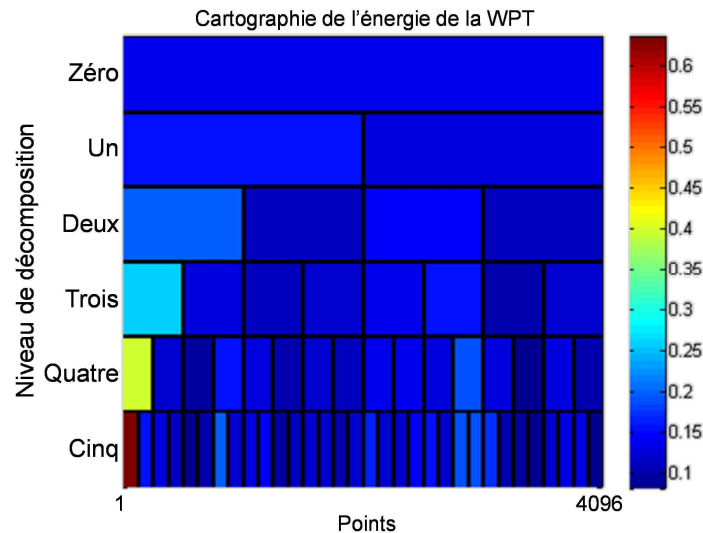


Figure 3.12. Cartographie de l'énergie

*Création de la matrice d'énergie*

Chaque valeur moyenne est ensuite rangée dans une matrice que nous nommerons matrice d'énergie  $E$  et qui contient 63 lignes (= le nombre total de nœuds) et un nombre de colonnes égal au nombre d'enregistrements de 4096 points contenus dans le groupe considéré (soit dans notre exemple 13). Le principe de construction de cette matrice est décrit par la figure 3.13, pour un nombre  $p$  d'enregistrements.

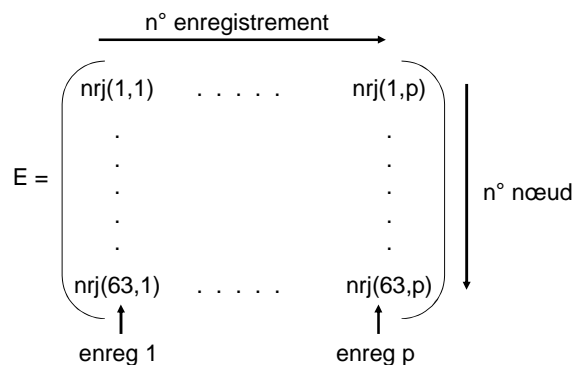


Figure 3.13. Matrice d'énergie  $E$

*Normalisation de l'énergie*

Cette dernière étape consiste à normaliser l'énergie par rapport à l'énergie moyenne du bruit contenu dans l'ensemble des enregistrements de la série de mesures. Par exemple, dans le cas de la série 1, la normalisation (aussi bien pour « transient\_1 » que pour « bruit\_1 ») est effectuée à partir de l'énergie moyenne de « bruit\_1 ». Pour cela, pour chaque enregistrement, on divise la valeur de l'énergie calculée en chaque nœud par la valeur moyenne du bruit en ce même nœud sur l'ensemble des enregistrements de bruit de la série de mesures considérée. Le principe de construction de la matrice d'énergie normalisée est décrit par la figure 3.14. Nous partons de la matrice  $E$  construite pour les enregistrements de bruit d'une série de mesures (par exemple série 1) et nous commençons par calculer l'énergie moyenne de chaque nœud sur l'ensemble des enregistrements (donc sur tous les éléments d'une même ligne). Ces valeurs moyennes sont notées  $\{E_{\text{bruit\_moy}}(1) \dots E_{\text{bruit\_moy}}(63)\}$  sur la figure 3.14. Ensuite nous prenons la matrice  $E$  des enregistrements de transitoire de cette même série de mesures (série 1) et nous divisons chaque élément d'une même ligne par la valeur moyenne

$E_{\text{bruit\_moy}}$  correspondante ( $E_{\text{bruit\_moy}}(1)$  pour la ligne 1 et ainsi de suite). Nous obtenons alors la matrice d'énergie normalisée, notée  $E_{\text{norm}}$  sur la figure 3.14, du groupe choisi (dans notre exemple « transient\_1 »).

Ensuite, il faut construire successivement les 3 autres matrices d'énergie normalisée (une pour chaque groupe) en suivant à chaque fois la même procédure.

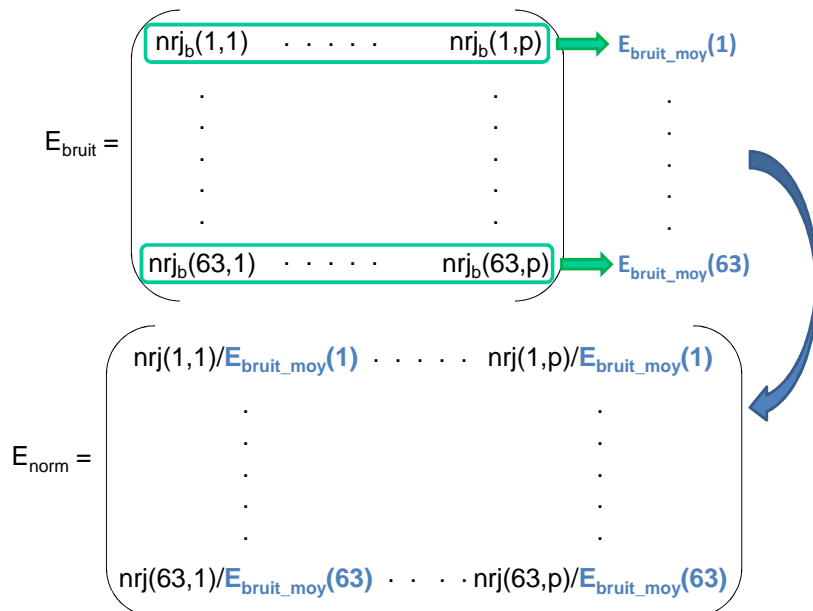


Figure 3.14. Construction de la matrice d'énergie normalisée

Ce principe de normalisation n'est pas applicable en l'état sur des signaux dont nous ne connaissons pas à l'avance la nature. Dès lors, comment sélectionner les enregistrements sur lesquels sera calculée l'énergie moyenne du bruit ?

Trois solutions sont envisageables. La première consiste à réaliser des enregistrements de l'environnement de bruit EM en l'absence de train lors des mesures. Après avoir vérifié que ces enregistrements ne contiennent pas de transitoires, ils serviront à calculer l'énergie moyenne du bruit.

Une deuxième solution pourrait être de sélectionner, a posteriori, manuellement quelques enregistrements de bruit parmi l'ensemble des mesures pour déterminer l'énergie moyenne du bruit de l'ensemble des mesures. Cette sélection devrait être réalisée par un "expert" capable de déterminer avec certitude le contenu (bruit ou transitoire) d'un fichier de mesures.

Enfin, nous pourrions, utiliser, pour chaque fichier de mesures les quelques centaines de points enregistrés avant le trig de déclenchement de l'oscilloscope et qui ne contiennent que du bruit. Ainsi, la valeur moyenne de l'énergie serait calculée grâce ces points, séparément sur chaque enregistrement. Chaque enregistrement (qu'il s'agisse de bruit ou de transitoire) se verrait alors appliquer une normalisation par rapport à l'énergie moyenne du bruit qu'il contient.





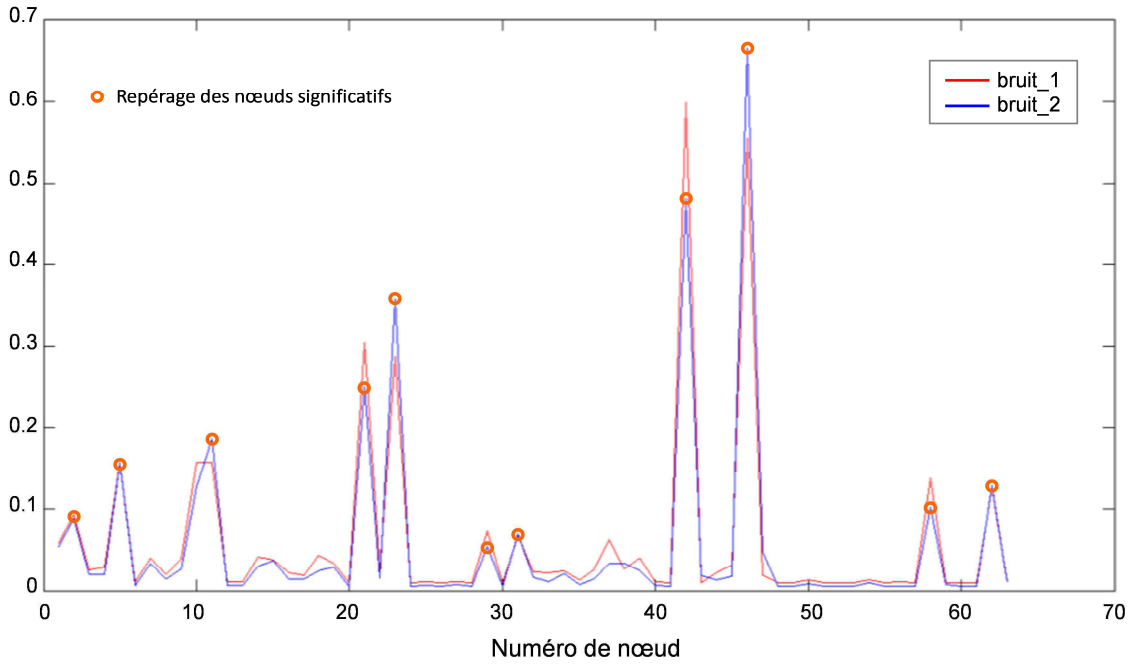


Figure 3.15. Représentation de la première colonne de U pour les enregistrements de bruit des séries de mesures 1 et 2

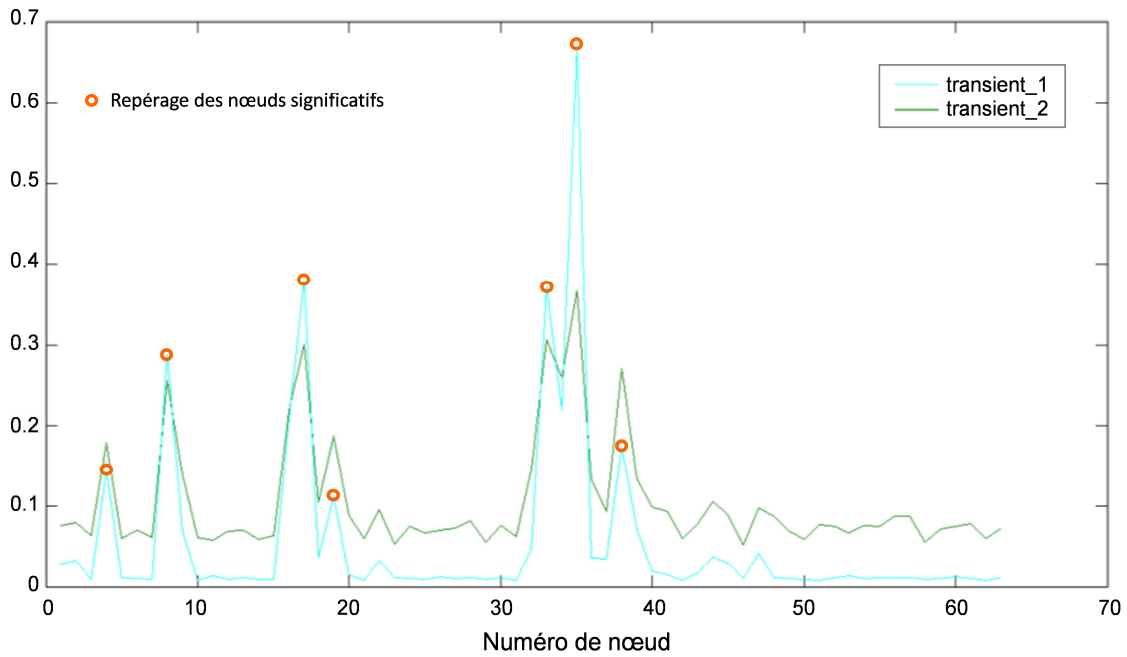
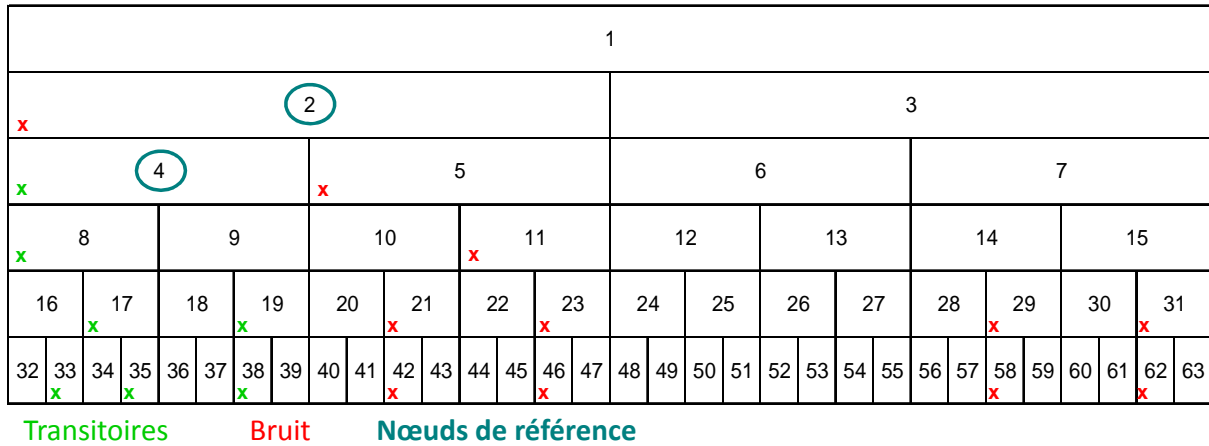


Figure 3.16. Représentation de la première colonne de U pour les transitoires des séries de mesures 1 et 2

Pour commencer, il faut repérer les maxima principaux (cercles orange sur les deux graphes) sur les deux représentations des vecteurs singuliers et associer chaque maximum au nœud qui lui correspond sur l'arbre de décomposition. Nous comptabilisons ainsi 11 nœuds pour le bruit et 7 pour les transitoires. Bien évidemment, nous n'allons pas retenir tous ces nœuds car tous ne sont pas réellement significatifs. De plus, une sélection plus fine peut être réalisée en observant leur position dans l'arbre de décomposition, comme nous allons l'expliquer dans le paragraphe suivant.

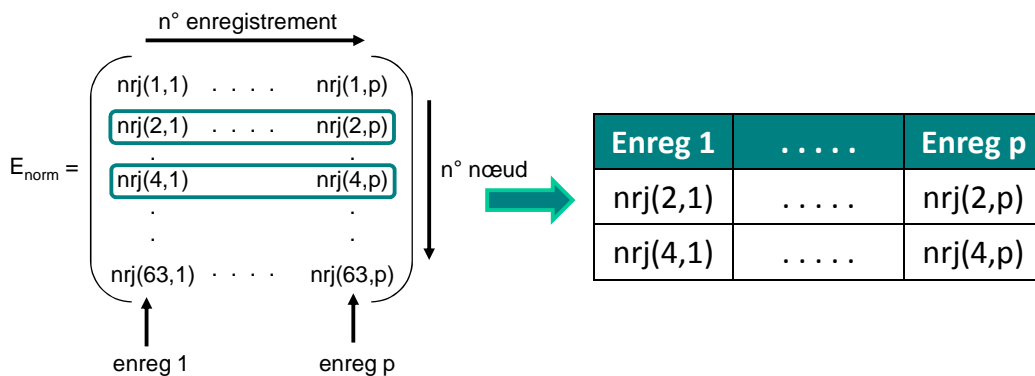
II.1.d - Sélection des nœuds significatifs sur la cartographie

Les positions des nœuds repérés en section II.1.c ont été reportées sur l'arbre de décomposition (figure 3.17) en rouge pour le bruit et en vert pour les transitoires. L'idée est de considérer la dernière ligne de l'arbre et de le remonter en éliminant au fur et à mesure les nœuds qui possèdent "une parenté" avec des nœuds du niveau de décomposition précédent. Par exemple, on peut éliminer de la liste des nœuds significatifs les nœuds 42, 46, 58 et 62 puisque ce sont "les fils" respectifs des nœuds 21, 23, 29 et 31. En appliquant le même raisonnement à l'ensemble des nœuds présélectionnés, il devient possible de réduire la liste des nœuds significatifs à seulement deux nœuds : 2 (bruit) et 4 (transitoires).



**Figure 3.17. Repérage et sélection des nœuds significatifs sur l'arbre de décomposition**

Une fois l'élimination des nœuds apparentés terminée, il ne reste plus qu'à sélectionner, sur la matrice d'énergie normalisée de chaque fichier, les valeurs d'énergie normalisée associées aux nœuds 2 et 4 pour l'ensemble des enregistrements (ce qui correspond aux lignes 2 et 4 de la matrice). Ces valeurs sont stockées dans un tableau, comme le montre la figure 3.18. Au final, nous disposons donc de 4 tableaux contenant les valeurs respectivement associées aux groupes « bruit\_1 », « transient\_1 », « bruit\_2 » et « transient\_2 ».



**Figure 3.18. Sélection des nœuds 2 et 4 sur l'ensemble des enregistrements**

II.1.e - Construction du diagramme de dispersion

Il s'agit maintenant de construire le diagramme de dispersion qui nous permet de visualiser, en deux dimensions, la distribution des éléments d'un même fichier. La figure 3.19 présente le diagramme de dispersion complet des 4 groupes « bruit\_1 », « transient\_1 », « bruit\_2 » et « transient\_2 ».

Ce diagramme fait apparaître deux classes de signaux (bruit et transitoires) avec un niveau de séparation satisfaisant.

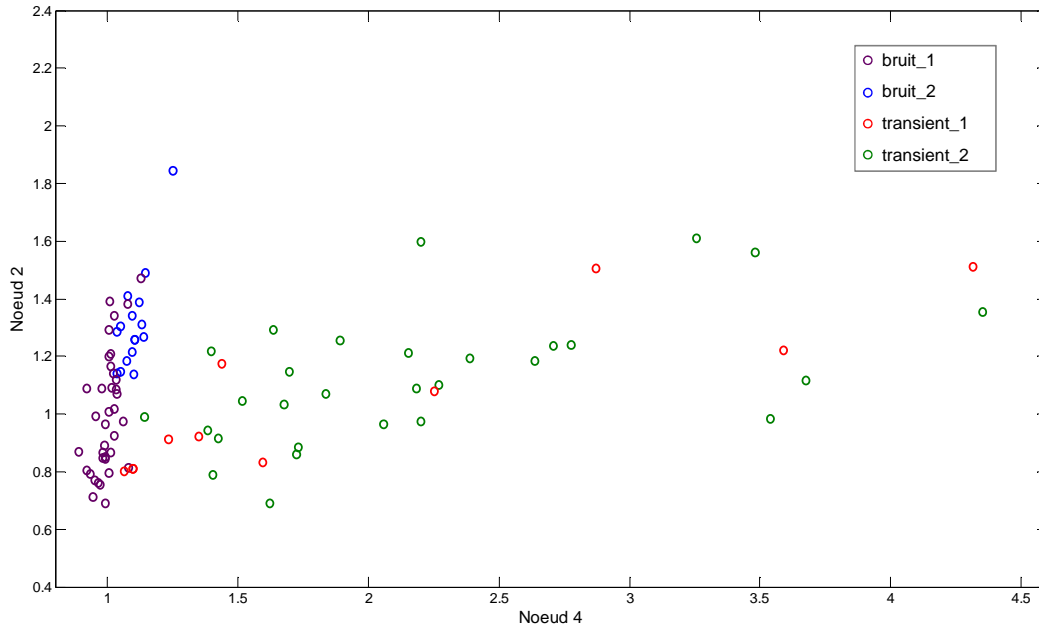


Figure 3.19. Diagramme de dispersion des 4 groupes

*II.1.f - Utilisation de la règle du plus proche voisin*

Le diagramme de dispersion est employé pour identifier la classe à laquelle appartient un signal que l'on suppose d'origine inconnue. L'idée est d'extraire, du signal à identifier, les informations nécessaires pour le placer sur le diagramme et d'utiliser ensuite la règle du plus proche voisin afin de déterminer sa classe. Si le point sur le diagramme est plus proche d'un élément de la classe des transitoires alors notre signal est identifié comme un signal transitoire et inversement avec le bruit. D'un point de vue mathématique, il s'agit en fait de calculer la distance euclidienne entre le point associé à notre signal de test et chaque point des deux classes de référence. Chacune de ces distances est stockée dans le tableau associé à sa classe et la valeur minimale de chaque tableau est notée  $D_{min}$  ( $D_{min\_trans}$  pour les transitoires et  $D_{min\_bruit}$  pour le bruit). Finalement, le point de test est attribué à la classe dont il est le plus proche, soit encore la classe qui a donné le plus petit  $D_{min}$ . Ainsi, sur l'exemple de la figure 3.20, le point du signal de test (en rouge) sera identifié comme appartenant à la classe des transitoires (en vert foncé).

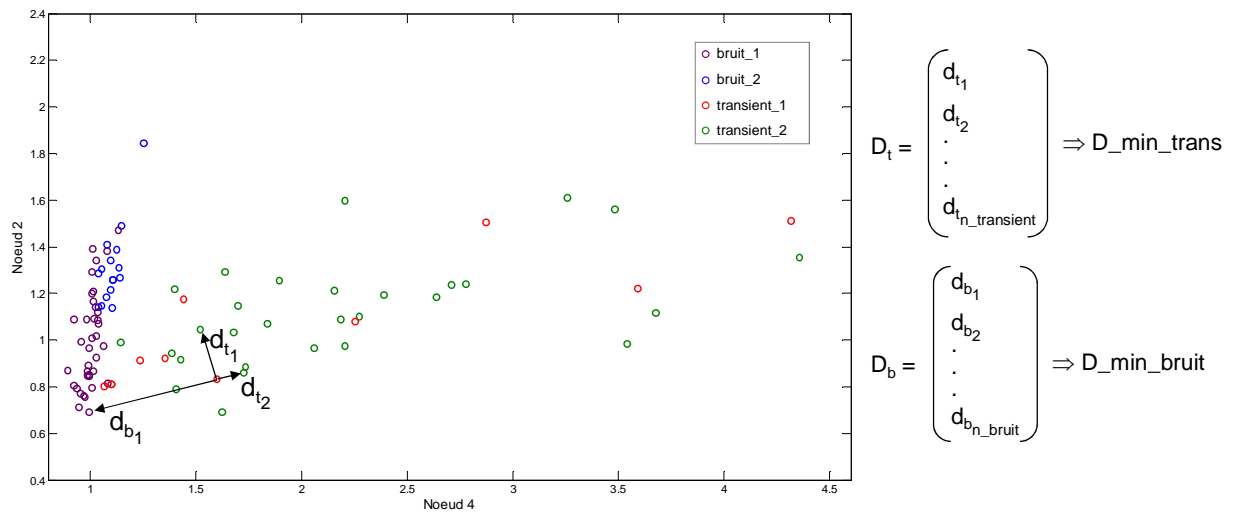
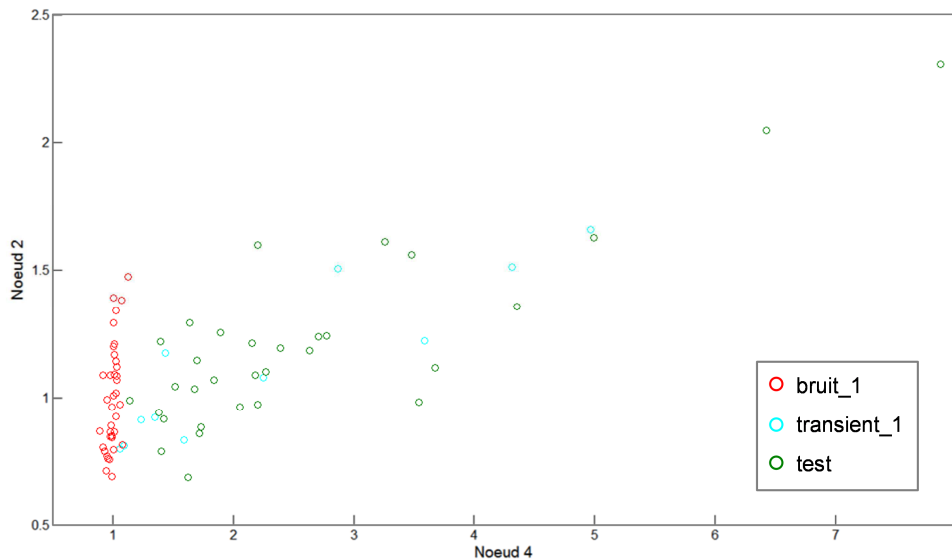


Figure 3.20. Calcul de toutes les distances entre l'élément à identifier et ses voisins

## II.2 - Résultats obtenus sur un jeu de données réelles

### II.2.a - Présentation du jeu de données

Le jeu de données qui a servi à ces essais se compose des 4 groupes présentés au début de cette section : « bruit\_1 », « transient\_1 », « bruit\_2 » et « transient\_2 ». Lors de chaque test, deux des 4 groupes (un pour les transitoires, un autre pour le bruit) servaient de référence lors de la construction du diagramme de dispersion. Ainsi, les signaux à identifier étaient classés en fonction de leur position respective sur le diagramme par rapport à ces points de référence. La figure 3.21 montre l'essai effectué afin de réaliser la classification des transitoires de la série de mesures 2 en utilisant comme référence les éléments associés aux transitoires et aux enregistrements de bruit de la série 1.



**Références :** transitoires → transient\_1    bruit → bruit\_1

fichier\_test : transient\_2 → 34 transitoires

**Résultat :**

- 33 transitoires

- 1 enregistrement de bruit

⇒ 1 erreur

Figure 3.21. Essai de la méthode de classification sur les transitoires de la série de mesures 2

Comme indiqué sur la figure 3.21, le groupe « transient\_2 » contient 34 transitoires. Notre méthode de classification permet d'identifier correctement 33 d'entre eux et commet une erreur de classification puisqu'il en attribue un à la classe du bruit. Cela correspond néanmoins à un pourcentage de bonne classification de 97.06%.

### II.2.b - Résultats

Plusieurs essais ont été menés en utilisant différentes combinaisons de groupes de référence et de groupes à classer. Les résultats obtenus sont rangés dans le tableau 4 qui regroupe le nombre d'identifications correctes, le nombre d'erreurs et les pourcentages correspondants sur l'ensemble des enregistrements testés (102 en tout) mais également par classe (47 transitoires et 55 enregistrements de bruit).

		Ensemble des enregistrements (102)	
Identifications correctes		78	
		76,47%	
Erreurs		24	
		23,53%	
		Transitoires (47)	Enreg. de bruit (55)
Identifications correctes		43	35
		91,49%	63,64%
Erreurs		4	20
		8,51%	36,36%

**Tableau 4. Résultats des essais de classification**

Le pourcentage global d'identifications correctes est satisfaisant, bien que perfectible. Si l'on considère séparément les classes, il apparaît que les transitoires sont correctement identifiés à 91.49 % contre 63.64% pour les enregistrements de bruit.

Afin d'améliorer la classification des signaux de bruit pour lesquels le taux de bonne classification s'avère le plus faible, nous avons tenté :

- d'augmenter le nombre de nœuds significatifs (de 2 à 5 puis de 5 à 7). Cela n'a toutefois pas permis d'améliorer le taux de bonne classification, ce qui signifie que les nœuds 2 et 4 sont vraiment les plus significatifs.
- d'utiliser d'autres règles que celle du plus proche voisin (3 plus proches voisins et plus proche centre de gravité). Il s'est avéré que la règle des trois plus proches voisins donnait des résultats légèrement meilleurs que ceux obtenus par la règle initiale.

### III - Bilan des deux applications préliminaires

Les deux applications qui viennent d'être présentées ont donné des résultats intéressants sur la détection et la classification de signaux dans notre contexte d'étude CEM.

L'application 1 montre notamment qu'une transformée TF permet de mettre en évidence l'apparition de signaux transitoires et de les localiser en temps avec une précision satisfaisante. Cette conclusion s'appliquant aussi bien à la transformée de Gabor qu'aux ondelettes. Dans le cadre de nos travaux, nous nous intéressons au système de communication GSM-R qui emploie des bandes de fréquences larges de 4 MHz pour les liens montant (876-880 MHz) et descendant (921-925 MHz). Compte tenu de la puissance d'émission d'un mobile GSM-R (8 W maximum), le lien montant n'est pas affecté par les perturbations EM transitoires présentes dans l'environnement ferroviaire. En revanche, la puissance du signal de réception GSM-R à bord des trains varie au cours du temps et peut atteindre des valeurs faibles (jusque -92 dBm) lorsque le train se situe entre deux stations de base successives. De ce fait, le lien descendant est le plus vulnérable à bord en présence de signaux transitoires. C'est donc le seul lien que nous considérons, limitant ainsi notre étude à la bande 921-925 MHz. Utiliser une transformée en ondelettes sur une bande réduite ne présente pas d'intérêt. Nous choisissons dès lors de réaliser une transformation de Gabor, en l'optimisant de manière à restreindre les calculs à la bande de fréquences qui nous intéresse. L'intérêt est double : limiter les calculs aux fréquences d'étude et gagner en rapidité d'exécution afin de se rapprocher d'une application temps-réel.

Avec l'application 2, nous comprenons qu'il est possible d'identifier la nature d'un signal, a priori inconnue, uniquement à partir de grandeurs caractéristiques extraites de celui-ci. Cela suppose d'avoir au préalable étudié les signaux en question afin de déterminer les valeurs significatives qui en caractérisent la nature et de définir la manière de les extraire. Pour ce qui est du GSM-R, ce travail a été réalisé lors d'une étude de sa susceptibilité en présence de signaux transitoires, issus du contact glissant entre la caténaire et le pantographe [3.8], dont les principaux résultats ont été précédemment présentés dans le chapitre 1.

De ces observations est née une idée : associer les deux applications pour aboutir à une méthode de classification de bruits EM transitoires en fonction de leur effet sur une transmission numérique (GSM-R) au cours d'un trajet à bord d'un train. Cette méthode de classification repose sur l'extraction de paramètres (cf. application 2) représentatifs de l'effet produit par les signaux transitoires sur la qualité d'une communication GSM-R. L'extraction est donc réalisée à partir des coefficients issus d'une transformation de Gabor (cf. application 1), sur la base d'une étude de la susceptibilité du GSM-R face aux signaux EM transitoires qui a été réalisée auparavant. Ces paramètres sont ensuite représentés sur un diagramme de dispersion de la même manière que dans l'application 2.

La section IV est maintenant consacrée à la description du principe de cette méthode de classification. En particulier, nous présentons les objectifs de notre méthodologie, décrivons le banc de tests/mesures mis en place au sein du laboratoire, ainsi que la configuration des équipements employée et le protocole de mesures suivi.

## IV - Principe de la méthodologie de classification proposée

### IV.1 - Objectif de notre méthodologie

Afin de disposer de conditions de travail plus souples qu'à bord de trains, nous avons reproduit en laboratoire les conditions de bruit EM dans lesquelles le système GSM-R doit opérer. Les seuls signaux de perturbation auxquels nous nous intéressons et que nous considérons dans notre étude sont les perturbations EM transitoires issues du contact glissant entre le fil caténaire et le pantographe. Comme précédemment expliqué dans le chapitre 1, ces dernières peuvent, en raison des larges bandes de fréquences (dont celles du GSM-R) qu'elles couvrent et des niveaux élevés de bruit EM qu'elles génèrent, perturber une communication GSM-R.

Cette situation est illustrée schématiquement par la figure 3.23 : durant une communication GSM-R (entre un mobile et un simulateur de réseau), des interférences EM transitoires se produisent et sont susceptibles de perturber la transmission.

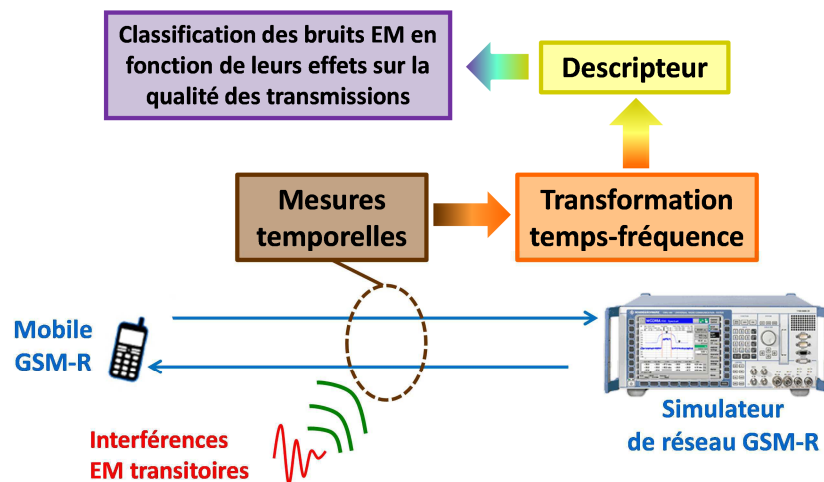


Figure 3.23. Objectif de notre approche

Notre objectif est de pouvoir, à l'aide d'une simple mesure temporelle, prédire l'impact que ces bruits EM transitoires auront sur la qualité de la communication GSM-R, puis de les classer dans la catégorie correspondante (dégradation ou maintien d'une qualité suffisante). Cette classification des bruits EM est réalisée grâce à un descripteur qui sera défini dans le chapitre 4. L'idée sous-jacente est en fait de proposer une méthodologie qui soit indépendante du système GSM-R que l'on cherche à protéger. Ainsi, au stade d'avancement où nous nous situons, il suffirait d'un oscilloscope couplé à l'antenne GSM-R du train pour procéder à la mesure des signaux et d'un ordinateur pour appliquer le traitement.

Une telle approche nécessite la génération d'un jeu de données représentatives des conditions réelles d'opération du GSM-R à bord des trains, d'un banc de mesures/tests entièrement configurable et d'un protocole de collecte des fichiers de mesures adapté.

Les scénarios de bruit EM transitoire sont construits à partir de modèles caractéristiques des transitoires issus du contact imparfait fil caténaire-pantographe. Ces modèles proviennent de l'analyse de signaux transitoires collectés à bord de trains lors de plusieurs campagnes de mesures [3.1], [3.8]

Ils sont au nombre de deux :

- la double exponentielle modulée par un sinus à la fréquence centrale du lien descendant. Il s'agit du "sinus-expo" dont l'équation est :

$$S(t) = A \times \left( e^{\frac{1}{D}t} - e^{-\frac{1}{RT}t} \right) \times u(t) \times \sin(2\pi Ft) \quad 3.5$$

où  $A = 1 \text{ V}$ ,  $D = 5 \text{ ns}$ ,  $RT = 0.4 \text{ ns}$ ,  $F = 923 \text{ MHz}$  et  $u$  est la fonction échelon unitaire. Ces valeurs ont été définies à partir des résultats présentés dans [3.1].

- la double exponentielle dont l'expression mathématique (cf. équation 3.1) est la même que celle du sinus-expo sans la multiplication par "sin(2πFt)".

Le banc de mesures et le protocole de collecte des fichiers de mesures font l'objet de la section IV.2.

## IV.2 - Banc de mesures, configuration employée et protocole suivi

### IV.2.a - Banc de mesures

Un banc complet de mesures a été développé en laboratoire. Il permet de simuler la présence de signaux transitoires durant une communication GSM-R et de réaliser l'ensemble des mesures nécessaires. Ce banc est composé :

- d'un mobile GSM-R et un simulateur de réseau (Rohde & Schwarz CMU 200) avec lequel sera établie puis évaluée la communication,
- d'un générateur de signaux arbitraires (AWG : Arbitrary Waveform Generator) qui permet la génération des signaux transitoires,
- d'un analyseur de spectre pour le calibrage des puissances des signaux de communication et d'interférence,
- d'un oscilloscope numérique pour la collecte des fichiers de mesures dans le domaine temporel.

Deux coupleurs (diviseurs de puissance) permettent d'assembler ces différents éléments comme le montre la figure 3.24.

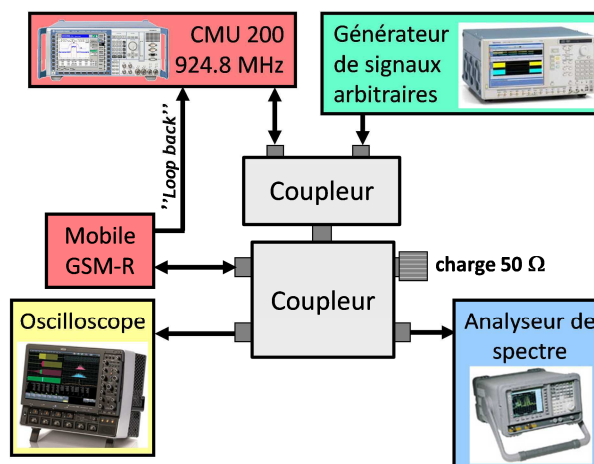


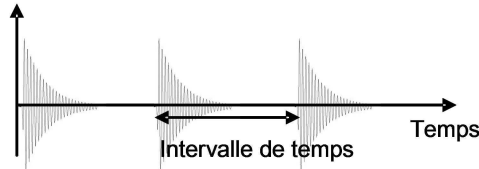
Figure 3.24. Banc de mesures/tests développé



IV.2.b - Configuration des équipements

Le canal utilisé pour la communication avec le CMU correspond au dernier canal GSM-R (924.8 MHz sur le lien descendant). La puissance du signal de communication est réglée en fonction du rapport signal-à-bruit que l'on souhaite obtenir. Dans le cadre de nos travaux, le SNR est défini comme le rapport entre la puissance du signal GSM-R et celle du bruit EM induit par les signaux transitoires sur le canal GSM-R à chaque apparition.

Le générateur de signaux arbitraires génère des signaux transitoires échantillonnés à 2.5 GHz en mode séquence, c'est-à-dire qu'il "répète" indéfiniment un même signal transitoire avec un intervalle de temps constant (et paramétrable) entre chaque occurrence. La figure 3.25 fournit une illustration de l'une de ces séquences.



**Figure 3.25. Exemple de séquence de transitoires générée par l'AWG**

L'oscilloscope collecte chaque fichier de mesures sur 2.5 millions de points avec une fréquence d'échantillonnage ( $f_{\text{échant}}$ ) de 2.5 GHz, ce qui correspond à une durée d'enregistrement de 1 ms. Cette configuration de mesure permet, d'une part de suivre l'évolution des signaux sur une durée relativement longue et surtout comparable à la durée d'un burst GSM-R (577  $\mu$ s). D'autre part, il est possible de mesurer des signaux transitoires dont la durée est extrêmement brève (de l'ordre d'une vingtaine de ns) avec un nombre de points satisfaisant en raison de la finesse de l'échantillonnage employé.

IV.2.c - Protocole de collecte des fichiers temporels

Une fois les équipements configurés comme décrit dans le paragraphe IV.2.b, nous pouvons procéder à la collecte des fichiers temporels. Celle-ci se déroule en 4 étapes :

- 1- Etablissement d'une communication entre le mobile GSM-R et le simulateur de réseau. C'est à ce moment-là qu'est appliquée la puissance du signal de communication émis par le CMU 200.
- 2- Génération du bruit EM transitoire à l'aide du générateur de signaux arbitraires.
- 3- Mesure du BER par le CMU 200 et, en fonction de la valeur obtenue, détermination de la classe à laquelle appartient le fichier de mesures. La classe ainsi identifiée sera intégrée dans le nom du fichier lors de son enregistrement.
- 4- Enregistrement du fichier de mesures à l'oscilloscope pour une durée de 1 ms.

Deux classes sont définies pour qualifier la qualité de la communication GSM-R en présence de bruit EM transitoire et ainsi labelliser les fichiers de mesures :

- la classe "bon" qui correspond à un  $BER \leq 1.13\%$  (bonne qualité de communication).
- la classe "mauv" qui correspond à un  $BER > 1.13\%$  (mauvaise qualité de communication).

Il s'agit ici du BER calculé avant toute correction d'erreur et il est défini comme [3.9] :

$$BER = \frac{\text{Nombre de bits erronés}}{\text{Nombre total de bits transmis}} \times 100\% \tag{3.6}$$

Ce BER est déterminé par le CMU 200 qui envoie une séquence de bits connue au mobile GSM-R. Le mobile est configuré en mode de fonctionnement "loop back", ce qui signifie qu'il renvoie au

simulateur les informations qu'il a reçues de ce dernier. Il lui renvoie donc la séquence qu'il a reçue et le CMU détermine le nombre de bits erronés par comparaison des deux séquences. La valeur limite de 1.13% utilisée pour définir les deux classes est la valeur de référence appliquée par les opérateurs ferroviaires pour juger de la qualité suffisante ou non d'une communication GSM-R [3.10].

## Conclusion

Ce chapitre s'est attaché à mettre en œuvre l'ensemble des outils théoriques et expérimentaux nécessaires à ce travail de Doctorat.

Les outils théoriques (transformées TF et une technique de classification) ont été mis en œuvre au travers de deux applications dédiées à la CEM. Les résultats obtenus ont montré que ces éléments théoriques, jusque-là peu ou pas employés en CEM, pouvaient être mis à profit dans ce domaine de recherche.

L'idée d'associer ces deux techniques a conduit à la construction d'une méthode de classification de bruits EM transitoires en fonction de leur effet sur une transmission GSM-R. Le principe de cette méthode ainsi que les outils expérimentaux qu'elle nécessite (banc de mesures, configuration des équipements, protocole de collecte des fichiers temporels) ont été décrits dans la dernière section de ce chapitre 3.

Cette méthode de classification repose sur l'extraction de paramètres représentatifs de l'effet produit par les signaux transitoires sur la qualité d'une communication GSM-R. L'extraction est réalisée à partir des coefficients issus d'une transformation de Gabor dont le calcul a été optimisé de manière à ne traiter que la bande de fréquences du lien descendant GSM-R. Ces paramètres sont ensuite représentés sur un diagramme de dispersion et la classification est réalisée grâce à la règle des k plus proches voisins. Le chapitre 4 est dédié à la description de cette méthode de classification, depuis la définition du descripteur jusqu'à la possible application à bord des trains. Nous y détaillons l'ensemble des calculs sur un exemple de scénario de bruit EM et présentons les résultats associés. L'interface de calcul et visualisation, spécialement développée pour notre méthode, y est également décrite.



# Chapitre 4

## Méthode de classification développée

---

### SOMMAIRE

---

<i>I - Critères de sélection et définition du descripteur</i>	<b>80</b>
<i>II - Description de la méthode</i>	<b>82</b>
<i>III - Présentation de l'interface de calcul et visualisation</i>	<b>95</b>
<i>Conclusion</i>	<b>102</b>

---

La méthode de classification que nous avons développée permet d'identifier l'effet produit par des bruits EM transitoires sur la qualité d'une communication GSM-R. Cette méthode repose sur l'utilisation d'un "descripteur", à savoir un paramètre mesuré ou estimé à partir de signaux acquis, représentatif de tous les facteurs influant sur la qualité de la communication. Dans le cadre de nos travaux, ce dernier permet d'une part, de suivre l'évolution des signaux perturbateurs et, d'autre part, d'estimer leurs effets sur la qualité des transmissions numériques. Il constitue un élément déterminant pour notre méthode car c'est à partir des seules informations qu'il contient qu'est réalisée la classification. Par conséquent, sa définition a fait l'objet de beaucoup de soins.

Ce chapitre décrit la méthodologie employée afin de définir un descripteur adapté à l'effet des bruits EM transitoires sur les communications sol-trains par GSM-R. Une fois définis les critères que doit remplir le descripteur, la définition et la méthode de recueil du descripteur au cours du temps, basée sur l'utilisation d'une transformée temps-fréquence, sont présentées. Nous décrivons ensuite la méthode de classification des conditions de bruits EM transitoires agissant sur une transmission GSM-R qui repose sur deux paramètres extraits du descripteur. Finalement, nous présentons l'interface qui a été développée pour paramétrer et réaliser l'ensemble des calculs et visualiser les résultats obtenus avec notre méthode de classification.

## I - Critères de sélection et définition du descripteur

### *1.1 - Critères de sélection du descripteur*

Notre méthode de classification repose entièrement sur l'utilisation d'un descripteur qui doit être représentatif de tous les facteurs influant sur la qualité de la communication GSM-R. En effet, c'est à partir des seules informations qu'il contient qu'est réalisée la classification.

Le descripteur doit permettre à la fois de suivre l'évolution des signaux perturbateurs et d'estimer leurs effets sur la qualité des transmissions numériques. Pour ce faire, il est nécessaire de suivre l'évolution, au cours du temps, à la fois du signal utile (signal de communication GSM-R) et du bruit EM transitoire. Nous avons décidé d'employer un descripteur qui met en relation les niveaux du signal utile et du bruit EM transitoire. Ce choix se justifie tout d'abord par le fait que leurs niveaux de puissance respectifs ne sont pas constants, notamment en situation de fonctionnement réelle (mobile GSM-R embarqué à bord d'un train). Par exemple, ils évoluent en fonction de la distance entre le mobile et la station de base pour les signaux GSM-R. Ensuite, l'impact du bruit EM transitoire dépend du niveau d'interférence qu'il induit sur les bandes de fréquences GSM-R, mais également du niveau des signaux de communication. Nous ne pouvons donc pas simplement fixer des valeurs-limites de niveau de bruit mais nous devons également observer et suivre le rapport entre les deux valeurs.

Une étude préalable portant sur la susceptibilité du système GSM-R en présence de bruits EM transitoires issus du contact glissant caténaire-pantographe a été menée [4.1]. La méthodologie employée et les principaux résultats ont été présentés au chapitre 1. Nous rappelons que cette étude a montré que l'effet produit par ce type d'interférences sur la qualité d'une communication GSM-R dépend de deux paramètres :

- la récurrence des signaux transitoires,
- la façon dont ces signaux dégradent le rapport signal à bruit à chaque apparition.

Ainsi, pour être pertinent notre descripteur doit permettre de suivre l'évolution de la récurrence et du SNR afin de rendre compte de l'impact de ces deux paramètres. De même, afin de relier ce descripteur à l'effet produit sur la qualité des communications GSM-R, il convient de suivre son évolution avec une résolution temporelle équivalente à l'échelle de la durée de transmission d'un bit GSM-R, soit 3.7  $\mu$ s. Cela nous permettra de mettre en évidence la perte ou la dégradation d'un bit et donc de faire le lien avec le BER qui constitue le critère de qualité utilisé pour déterminer la classe d'appartenance d'un fichier. L'ensemble de ces constatations nous conduit à la formulation du descripteur énoncée dans la section qui suit.

### *1.2 - Définition du descripteur*

Compte tenu des observations précédentes, le descripteur idéal remplissant l'ensemble de ces critères est défini par la formule :

$$D(t) = \frac{P_{\text{GSM-R}}(f_{\text{GSM-R}}, t)}{P_{\text{interférences}}(f_{\text{GSM-R}}, t)} \quad 4.1$$

où  $P_{\text{GSM-R}}(f_{\text{GSM-R}}, t)$  désigne la puissance du signal utile à l'instant  $t$  et  $P_{\text{interférences}}(f_{\text{GSM-R}}, t)$  correspond à la puissance ramenée, également à l'instant  $t$ , par les signaux transitoires sur le canal employé pour la communication GSM-R.

Et avec la condition que  $t$  s'incrémente avec un pas  $\leq 3.7 \mu\text{s}$  (= durée de transmission d'un bit GSM-R), de manière à obtenir au moins une valeur de  $D$  par bit GSM-R.

Cependant, il n'est pas possible d'évaluer le niveau du bruit transitoire sur le même canal que celui employé pour la communication GSM-R. En effet, les deux signaux étant superposés, nous ne pouvons pas distinguer le niveau de puissance produit respectivement par chacun. Or, nous savons que les transitoires sont des événements large bande, donc nous pouvons faire l'hypothèse qu'ils couvrent de manière uniforme l'ensemble des fréquences de fonctionnement du GSM-R (921-925 MHz). Ainsi, pour déterminer le niveau de bruit ramené par les transitoires, nous recueillons le niveau maximal de puissance sur la totalité de la bande 921-925 MHz à l'exception du canal de communication GSM-R utilisé. Nous faisons donc l'approximation suivante :

$$P_{\text{interférences}}(f_{\text{GSM-R}}, t) \approx P_{\text{max sur les autres canaux}}(t) \quad 4.2$$

La combinaison des équations 4.1 et 4.2 nous conduit finalement au descripteur défini par :

$$D(t) = \frac{P_{\text{GSM-R}}(f_{\text{GSM-R}}, t)}{P_{\text{max sur les autres canaux}}(t)} \quad 4.3$$

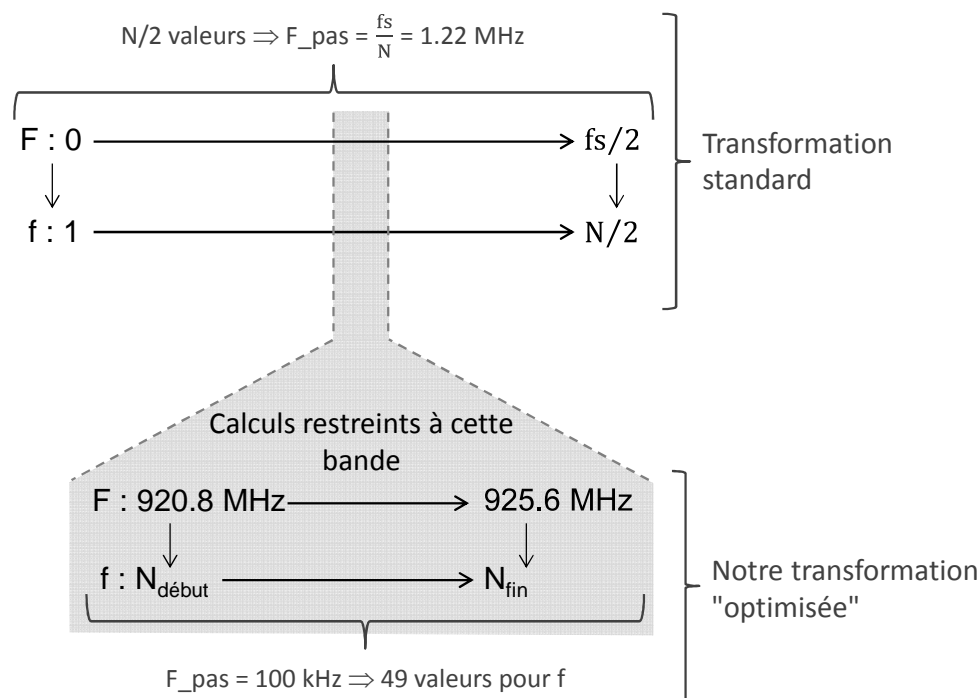
Les valeurs du signal utile,  $P_{\text{GSM-R}}$ , et du bruit EM transitoire,  $P_{\text{max sur les autres canaux}}$ , sont toutes deux déterminées à l'aide d'une transformation temps-fréquence que nous avons adaptée au contexte de notre application. Elle est décrite dans la section II.1.

## II - Description de la méthode

### II.1 - Transformée de Gabor optimisée

Comme énoncé à la fin du chapitre 3, nous choisissons d'employer une transformée de Gabor. Toutefois, nous l'optimisons de manière à effectuer les calculs uniquement sur la bande de fréquences étudiée (921-925 MHz). Rappelons que nous limitons notre étude au lien descendant du GSM-R car c'est le lien qui, à bord des trains, s'avère être le plus vulnérable.

Compte tenu de la configuration de mesure adoptée ( $f_{ech}=2.5$  GHz), une transformation de Gabor standard fournirait l'évolution spectrale du signal en opérant sur la totalité de la bande de fréquences allant de 0 à  $f_{ech}/2$  (soit ici de 0 à 1.25 GHz) avec un pas en fréquence qui dépendrait de la largeur de la fenêtre employée. Or, nous n'avons pas besoin d'étendre le calcul à toutes ces fréquences dans la mesure où seule la bande de fréquences du lien descendant est considérée. Nous limitons donc les calculs de notre transformée "optimisée" à cette bande particulière, tout en conservant une résolution fréquentielle suffisamment fine (100 kHz), de manière à pouvoir distinguer les canaux GSM-R espacés de 200 kHz. Pour cela, l'intervalle de variation du paramètre  $f$  est réduit selon le principe décrit par la figure 4.1.

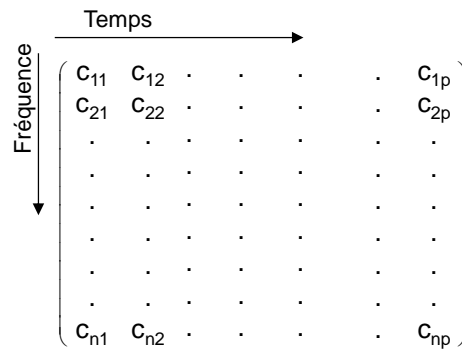


**Figure 4.1. Principe utilisé pour optimiser la transformée de Gabor**

Avec une transformation standard (partie supérieure de la figure 4.1),  $f$  serait incrémenté de 1 à  $N/2$ ,  $N$  étant la largeur de la fenêtre gaussienne. Le calcul serait ainsi effectué sur la bande de fréquences  $F = [0, f_{ech}/2]$ , avec  $N/2$  valeurs pour  $f$ . Compte tenu de la fréquence d'échantillonnage ( $f_{ech} = 2.5$  GHz) et de la largeur de la fenêtre gaussienne utilisée ( $N = 2048$  points), cela correspond à la bande de fréquences  $F = [0, 1.25$  GHz] avec un pas en fréquence égal à 1.22 MHz. Sachant que les canaux GSM-R sont espacés de 200 kHz, il serait impossible de les distinguer. Pour résoudre ce problème et également rendre le calcul plus rapide, nous réduisons la bande de calcul à l'intervalle  $F = [920.8$  MHz, 925.6 MHz] avec un pas en fréquence de 100 kHz. L'intervalle de variation correspondant de  $f$ ,  $[N_{start}, N_{stop}]$ , est déterminé grâce à un produit en croix et au final, le paramètre  $f$  s'incrémente sur 49 valeurs.

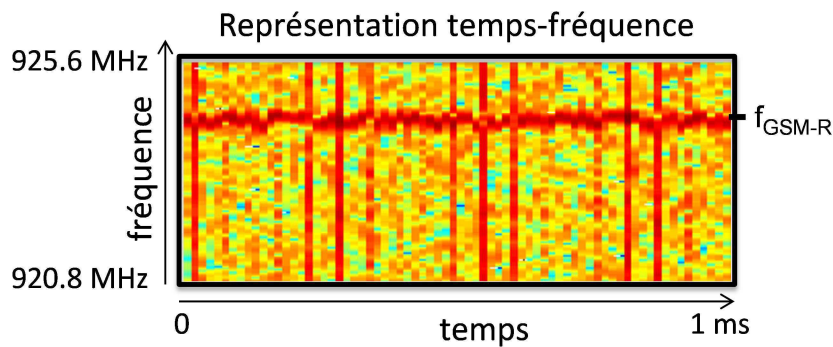


Comme illustré sur la figure 4.2, les coefficients sont rangés dans une matrice. Chaque ligne correspond à une fréquence de la bande 920.8-925.6 MHz et chaque colonne à une fenêtre de calcul de la transformée. Ainsi, la puissance émise sur un canal donné pendant toute la durée de la mesure est donnée par une ligne de la matrice. De même, le spectre, à un instant donné, du signal sur la bande 920.8-925.6 MHz est fourni par les coefficients d'une colonne. Comme expliqué dans le paragraphe précédent, notre matrice compte 49 pas en fréquences, donc  $n=49$  sur la figure 4.2. Ensuite, étant donné la largeur de la fenêtre gaussienne utilisée ( $N = 2048$  points), la durée de la mesure (1 ms, soit encore 2.5 millions de points) et le facteur de recouvrement appliqué (0.5), le calcul est réalisé sur 2439 fenêtres et donc  $p=2439$  sur la figure 4.2.



**Figure 4.2. Représentation de la matrice des coefficients de notre transformée de Gabor optimisée**

La figure 4.3 montre la représentation temps-fréquence typiquement obtenue avec notre transformée de Gabor optimisée. Sur cette représentation apparaît une ligne horizontale qui est présente sur la totalité du fichier de mesures (1 ms). Cette ligne correspond au signal de communication GSM-R dont la fréquence d'émission est constante sur la durée de la mesure. Nous notons également la présence de plusieurs lignes verticales qui couvrent toute la bande de fréquences 920.8-925.6 MHz. Ces dernières marquent l'apparition des signaux transitoires (qui, rappelons-le, sont des événements large bande).



**Figure 4.3. Exemple de représentation temps-fréquence obtenue avec notre transformée**

## II.2 - Recueil du descripteur

Le descripteur est calculé en parcourant la RTF pour déterminer  $P_{\text{GSM-R}}$ , et ensuite  $P_{\text{max}}$  sur les autres canaux à chaque instant  $t$ .

### II.2.a. Détermination de $P_{\text{GSM-R}}$ et $P_{\text{max}}$ sur les autres canaux à chaque instant $t$

La première étape consiste à détecter la fréquence d'émission du signal GSM-R. Le principe est de parcourir la RTF pour déterminer à chaque instant  $t$  le canal employé par la communication. Une fois la totalité des colonnes de la matrice parcourue, la fréquence GSM-R la plus fréquemment détectée est considérée comme celle utilisée par la communication. Par exemple, sur la figure 4.4, le canal le plus fréquemment détecté est 924.8 MHz. Précisons qu'il correspond bien au canal GSM-R utilisé lors de cette mesure.

Ensuite, la valeur de la puissance du signal GSM-R est estimée en se reportant à la ligne correspondante à ce canal dans la matrice de la transformée. Une règle supplémentaire a toutefois été introduite pour tenir compte de la présence des signaux transitoires. En effet, si un transitoire est présent à l'instant  $t$ , alors les deux signaux (utile et transitoire) sont mélangés sur le canal GSM-R et dans ce cas, nous ne pouvons pas distinguer leurs puissances respectives. Il nous faut alors appliquer notre nouvelle règle qui consiste à conserver la valeur calculée à l'instant précédent pour le signal GSM-R :  $P_{\text{GSM-R}}(t) = P_{\text{GSM-R}}(t-1)$ . Cette caractéristique est illustrée sur les courbes de la figure 4.5. La première courbe représente l'évolution de la puissance du signal sur le canal 924.8 MHz directement donnée par la transformée de Gabor. Nous constatons que des hausses soudaines de puissance se produisent à plusieurs reprises : celles-ci marquent l'apparition des signaux transitoires et de ce fait, le niveau mesuré à ces instants ne correspond pas à celui du signal de communication. En appliquant la nouvelle règle à notre algorithme, ces hausses ne sont plus prises en compte lors de l'estimation de  $P_{\text{GSM-R}}$ , comme on peut le voir sur la deuxième courbe qui représente l'évolution de la puissance du signal sur le canal détecté. De plus, nous notons une bonne correspondance entre les niveaux de signal calculés sur le canal effectivement employé et sur le canal détecté. Cela montre que notre algorithme permet bien de récupérer la puissance seule du signal de communication,  $P_{\text{GSM-R}}$ , à chaque instant  $t$ .

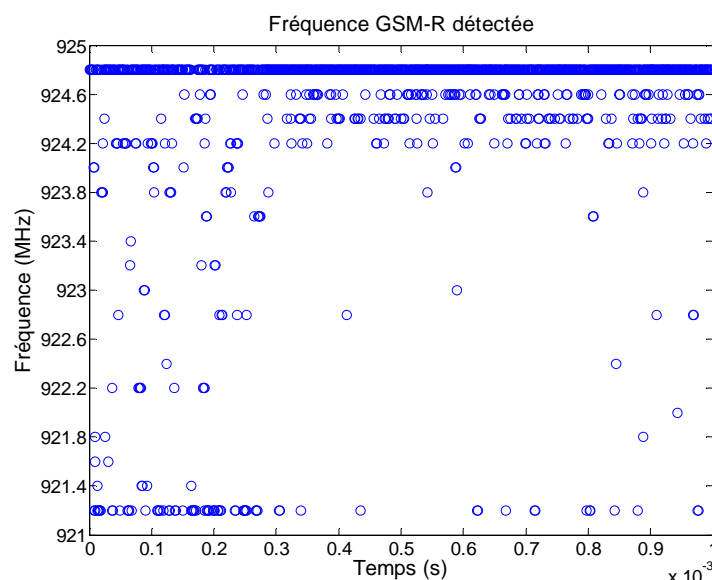


Figure 4.4. Détection de la fréquence GSM-R employée par la communication

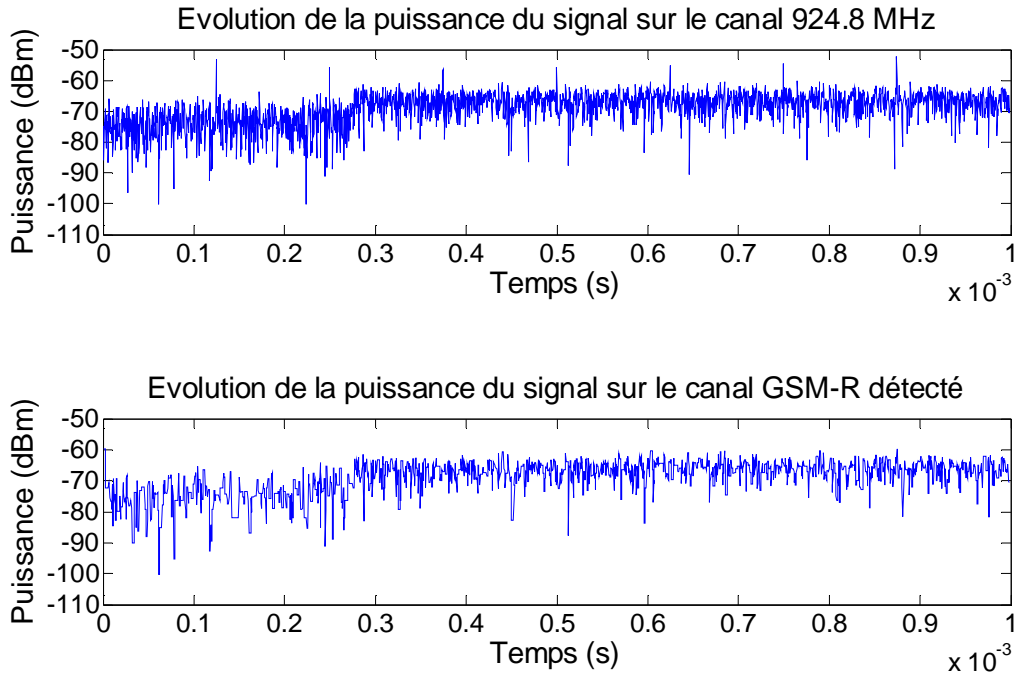


Figure 4.5. Puissances sur le canal GSM-R effectivement employé et celui détecté

Quant à  $P_{\max}$  sur les autres canaux, elle est déterminée en prenant la valeur maximale sur la bande de fréquences 921-925 MHz, à l'exception du canal utilisé pour la communication. Le rapport entre les deux niveaux de puissance,  $P_{\text{GSM-R}}$  et  $P_{\max}$  sur les autres canaux, en chaque instant de la TFR nous donne l'évolution du descripteur D au cours du temps.

### II.2.b. Calcul du descripteur

En procédant comme expliqué en II.2.a, nous obtenons autant de valeurs pour D que la transformée comporte de colonnes, c'est-à-dire 2439 (cf. section II.1). Or, le but est de relier notre descripteur à des valeurs de BER qui constitue le critère de qualité utilisé. Par conséquent, il est nécessaire de travailler à l'échelle d'un bit de transmission GSM-R (soit 3.7  $\mu\text{s}$ ).

Compte tenu du débit du signal GSM-R de 271 kbits/s, 271 bits GSM-R sont transmis durant les 1 ms que dure la mesure. Afin d'obtenir une valeur pour le descripteur D toute les 3.7  $\mu\text{s}$ , nous devons combiner le calcul du rapport  $P_{\text{GSM-R}}/P_{\max}$  sur les autres canaux sur plusieurs fenêtres successives de la transformée.

Pour ce faire, nous devons d'abord estimer les valeurs de  $P_{\text{GSM-R}}$  et  $P_{\max}$  sur les autres canaux pour plusieurs fenêtres successives de la représentation temps-fréquence. Le nombre de fenêtres de regroupement ( $N_f$ ) est fixé, en fonction de la durée de la fenêtre gaussienne utilisée lors du calcul de la transformée, de manière à couvrir 3.7  $\mu\text{s}$ . Ensuite, nous prenons les maxima respectifs de ces deux signaux sur les  $N_f$  fenêtres (notés  $\text{Max\_P\_GSM-R\_ech\_bit}$  et  $\text{Max\_P\_autres\_canaux\_ech\_bit}$ ). En traçant la courbe de la puissance du GSM-R ainsi déterminée (soit la courbe des  $\text{Max\_P\_GSM-R\_ech\_bit}$ ), nous constatons que d'importantes fluctuations très ponctuelles, telles que celles observables sur la figure 4.5, sont toujours présentes. Pour éviter que celles-ci ne faussent le calcul du descripteur par sous-estimation du signal de communication, nous calculons la moyenne des valeurs maximales  $\text{Max\_P\_GSM-R\_ech\_bit}$ . De cette manière, nous obtenons une unique valeur de puissance pour le signal de communication sur la durée de transmission des 271 bits.

Finalement, le ratio entre cette valeur moyenne et les maxima de  $P_{\max \text{ sur les autres canaux}}$  issus du regroupement des  $N_f$  fenêtres (notés  $\text{Max\_P\_autres\_canaux\_ech\_bit}$ ) nous donne le descripteur. Une formulation algorithmique de cette opération peut être écrite comme suit :

$$D(k) = \frac{\text{Moyenne}(\text{Max\_P\_GSM-R\_ech\_bit}(k))}{\text{Max\_P\_autres\_canaux\_ech\_bit}(k)} \quad 4.4$$

où

$$\text{Max\_P\_GSM-R\_ech\_bit}(k) = \max_j \left\{ P_{\text{GSM-R}} \left( T_b(k-1) + j \frac{d_f}{2} \right), \forall j \in [1, N_f] \right\}$$

et

$$\text{Max\_P\_autres\_canaux\_ech\_bit}(k) = \max_j \left\{ P_{\max \text{ sur les autres canaux}} \left( T_b(k-1) + j \frac{d_f}{2} \right), \forall j \in [1, N_f] \right\}$$

Avec  $k$ , le numéro du bit GSM-R et  $k \in [1, 271]$

$T_b$ , la durée de transmission d'un bit GSM-R (3.7  $\mu\text{s}$ )

$N_f$ , le nombre de fenêtres successives de la RTF sur lesquelles est effectué le regroupement

$d_f$ , la durée de la fenêtre gaussienne employée pour le calcul de la transformée

Avec une fréquence d'échantillonnage de 2.5 GHz, les 2048 points de la fenêtre gaussienne équivalent à une durée  $d_f = 819$  ns. En tenant compte du recouvrement de 50%, il faut donc regrouper  $N_f = 8$  fenêtres pour couvrir 3.7  $\mu\text{s}$ , comme l'illustre la figure 4.6.

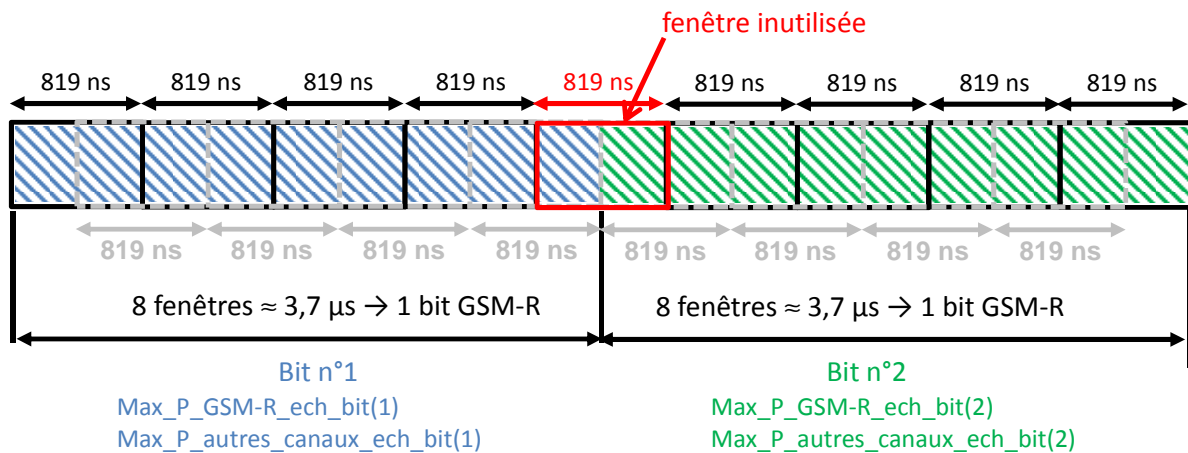


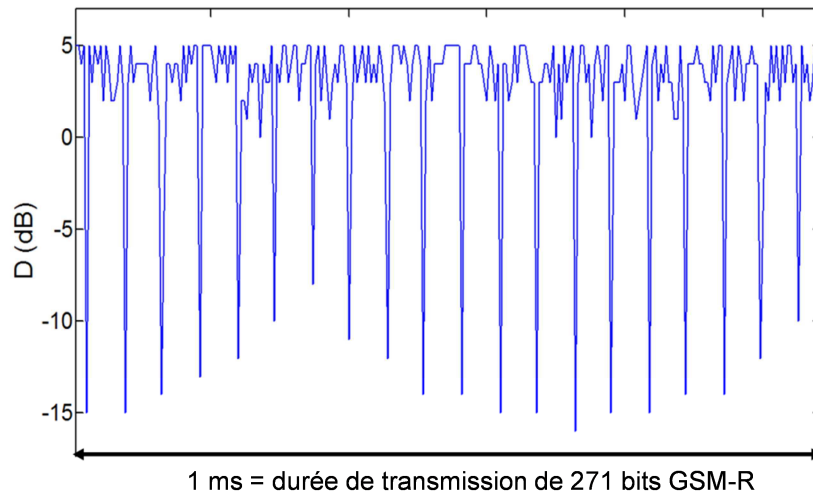
Figure 4.6. Principe de regroupement des fenêtres lors du calcul du descripteur

Une fois les calculs effectués selon ce principe, nous obtenons la courbe de l'évolution temporelle du descripteur sur le temps de transmission de 271 bits GSM-R telle que présentée sur la figure 4.7. Dans cet exemple, le descripteur a été recueilli en présence de signal GSM-R et d'un même signal transitoire répété toutes les 50  $\mu\text{s}$ .

Sur cette courbe sont présents des minima qui marquent les apparitions du signal transitoire. En fait, à chaque fois qu'un transitoire se produit, il génère du bruit EM qui couvre notamment les bandes du GSM-R avec un niveau de puissance supérieur à celui du signal utile dont le niveau reste constant tout au long de la mesure. De ce fait, on observe une diminution soudaine de la valeur prise par  $D$  au moment de l'apparition d'un transitoire. Cette caractéristique permet de recenser le nombre de transitoires qui se sont produits sur la durée de la mesure.

On constate que les valeurs des minima sont particulièrement faibles si l'on se réfère aux ordres de grandeur des SNR habituellement observés. Ceci provient de l'estimation du niveau d'interférence

induit par les bruits transitoires. En effet, en raison de la bande passante de l'oscilloscope, le signal transitoire, large bande, met en évidence un niveau plus élevé que ce que mesure un analyseur de spectre utilisant une bande de résolution (RBW) étroite. En revanche, le signal GSM-R étant à bande étroite, son niveau estimé est en accord avec une mesure à l'analyseur de spectre. De ce fait les minima du descripteur sont plus faibles que ceux attendus. Néanmoins, ce ne sont pas les valeurs du descripteur en elles-mêmes qui nous intéressent mais la façon dont celui-ci évolue au cours du temps. En effet, le principe est d'associer une évolution du descripteur à un état de la qualité de la communication.



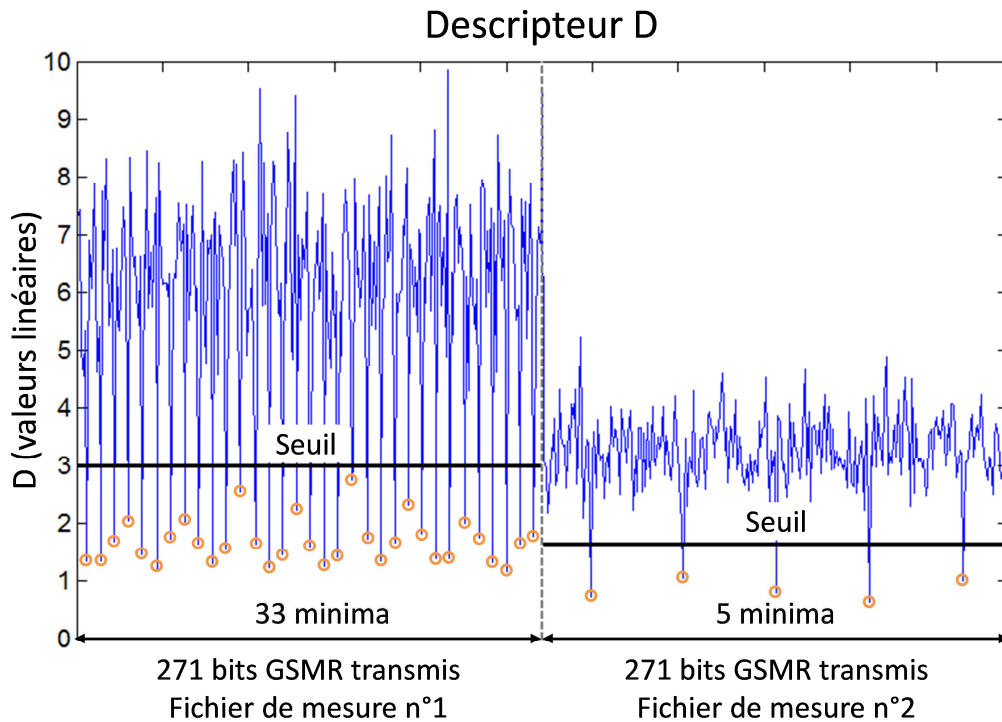
**Figure 4.7. Exemple de courbe du descripteur**

### *II.3 - Extraction des paramètres*

A partir de la courbe du descripteur ainsi obtenue, il est possible d'extraire certains paramètres. Nous rappelons que les deux paramètres influant sur la qualité d'une communication GSM-R sont la fréquence d'apparition des transitoires et le SNR produit à chaque occurrence. D'après ce qui a été vu dans la section II. 2, il est possible de connaître le nombre de transitoires qui se sont produits durant la mesure en comptant le nombre de minima. De plus, à chaque minimum, la "chute" de valeur du descripteur est représentative de la diminution du SNR provoquée par l'apparition du signal transitoire. En conséquence, le nombre de minima constitue le premier paramètre utilisé pour établir le lien entre la qualité des transmissions GSM-R et les conditions de bruit EM dans lesquelles ces dernières se déroulent. Le deuxième paramètre, qui doit rendre compte de l'évolution du SNR, correspond à la somme des valeurs prises par D à ces mêmes minima (soit encore à la somme des minima).

Compter le nombre de minima sur la courbe du descripteur nécessite de définir un seuil de détection. Toutes les valeurs en-deçà de ce dernier sont considérées comme des minima. La définition du seuil est un point crucial car elle conditionne la bonne comptabilisation des transitoires. En effet, si le seuil est trop faible, certains minima peuvent ne pas être détectés. A l'inverse, s'il est trop élevé, de fausses détections peuvent être introduites. Dans les deux cas, il existe des erreurs dans le comptage des minima. Ainsi le nombre de minima détectés et par conséquent la somme de ces mêmes minima sont erronés et ne reflètent donc pas la réalité de la mesure. Pour notre méthode, cela se traduit par des erreurs de classification. Par conséquent, le seuil doit être choisi avec soin.

Une fois le seuil défini, notre algorithme détecte automatiquement les valeurs de D inférieures à ce dernier et effectue le calcul de la somme des valeurs prises par D à ces minima. La figure 4.8 en montre l'application sur deux fichiers de mesures distincts.



**Figure 4.8.** Exemple d'extraction des paramètres à partir de la courbe du descripteur pour deux fichiers de mesures

Pour obtenir les résultats présentés ici, le seuil est défini comme un pourcentage de la valeur moyenne du descripteur sur chaque fenêtre d'une durée de 1 ms. C'est la raison pour laquelle deux seuils différents sont appliqués sur cette figure. De plus, le pourcentage appliqué n'est pas nécessairement le même pour toutes les mesures.

Nous reviendrons plus en détails sur la définition du seuil dans la section II.5 qui lui est consacrée. Nous y expliquerons, à partir de plusieurs exemples, son importance pour notre méthode de classification et décrivons la manière dont nous avons choisi de procéder pour déterminer au mieux sa valeur.

#### *II.4 - Construction de la base de référence et visualisation du diagramme de dispersion*

Une fois le nombre de minima (paramètre 1) et la somme des minima (paramètre 2) extraits de la courbe du descripteur pour un fichier de mesures, ces valeurs sont stockées dans un tableau. Ce tableau constitue notre base de référence. Il est séparé en deux parties : l'une contenant les éléments appartenant à la classe "bon" et l'autre, ceux de la classe "mauv". L'objectif est de construire une base de référence qui contienne le plus d'éléments possible, et surtout qui soit représentative de toutes les conditions de bruit EM transitoire que le GSM-R peut rencontrer dans l'environnement ferroviaire. Il faut donc diversifier au maximum les scénarios de bruit considérés en variant les signaux transitoires perturbateurs ainsi que les configurations de test (SNR, fréquence d'apparition des transitoires).

La construction de la base de référence complète sera décrite en détails dans le chapitre 5 dans lequel sont présentés et discutés les résultats de notre méthode. Nous ne présentons ici que la partie de la base associée au scénario de bruit EM transitoire n°1. Ce dernier se compose du modèle sinus-expo

(précédemment défini au chapitre 3) répété avec un intervalle de temps (TI) fixe entre deux transitoires successifs.

Pour ce qui est des configurations de test, nous les faisons varier comme suit :

- variation du niveau de puissance du signal utile entre -60 et -76 dBm sur 12 valeurs. Cela nous permet d'avoir une importante plage de variation du SNR.
- lors de la génération des interférences transitoires, variation de l'intervalle de temps entre deux transitoires successifs sur 32 valeurs allant de 20  $\mu$ s à 1 ms, comme présenté sur la figure 4.9.

Après avoir correctement réglé le niveau du signal GSM-R, les signaux transitoires sont générés en appliquant successivement les 32 valeurs de TI. On collecte un fichier de mesures à l'oscilloscope pour chaque valeur de TI appliquée. Chaque fichier est ainsi représentatif d'une valeur de SNR et d'une fréquence d'occurrence des transitoires. Ainsi, pour une valeur de SNR donnée, 32 fichiers sont collectés.

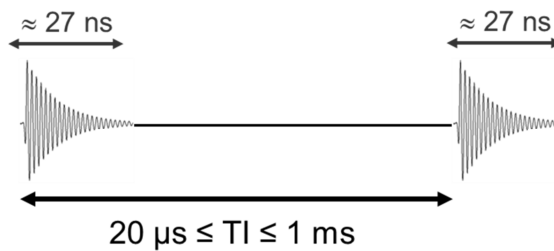


Figure 4.9. Scénario de bruit EM transitoire n°1

Ces premiers essais permettent de construire une partie de notre base de référence (notée base réf. 1) composée de 246 éléments associés à la classe "bon" et 101 à la classe "mauv". La figure 4.10 en fournit une représentation.

Eléments classe "bon" - BER $\leq$ 1.13 %		
Mesure n°	Nombre de minima	Somme des minima
1	1	5.47
2	1	5.65
.	.	.
.	.	.
.	.	.
246	24	52.38
Eléments classe "mauv" - BER $>$ 1.13 %		
Mesure n°	Nombre de minima	Somme des minima
1	25	51.05
2	9	14.45
.	.	.
.	.	.
.	.	.
101	5	4.02

Figure 4.10. Représentation de base réf. 1

Nous cherchons maintenant à savoir si l'espace défini par nos deux paramètres est adapté à notre problème de classification. Dans ce but, nous visualisons la répartition des éléments de notre base réf. 1 à l'aide d'un diagramme de dispersion. Cela nous permet de vérifier visuellement si nous obtenons une bonne séparation des classes et donc de juger de la pertinence des deux paramètres

employés pour la classification. Le diagramme de dispersion des éléments de base réf. 1 est présenté figure 4.11.

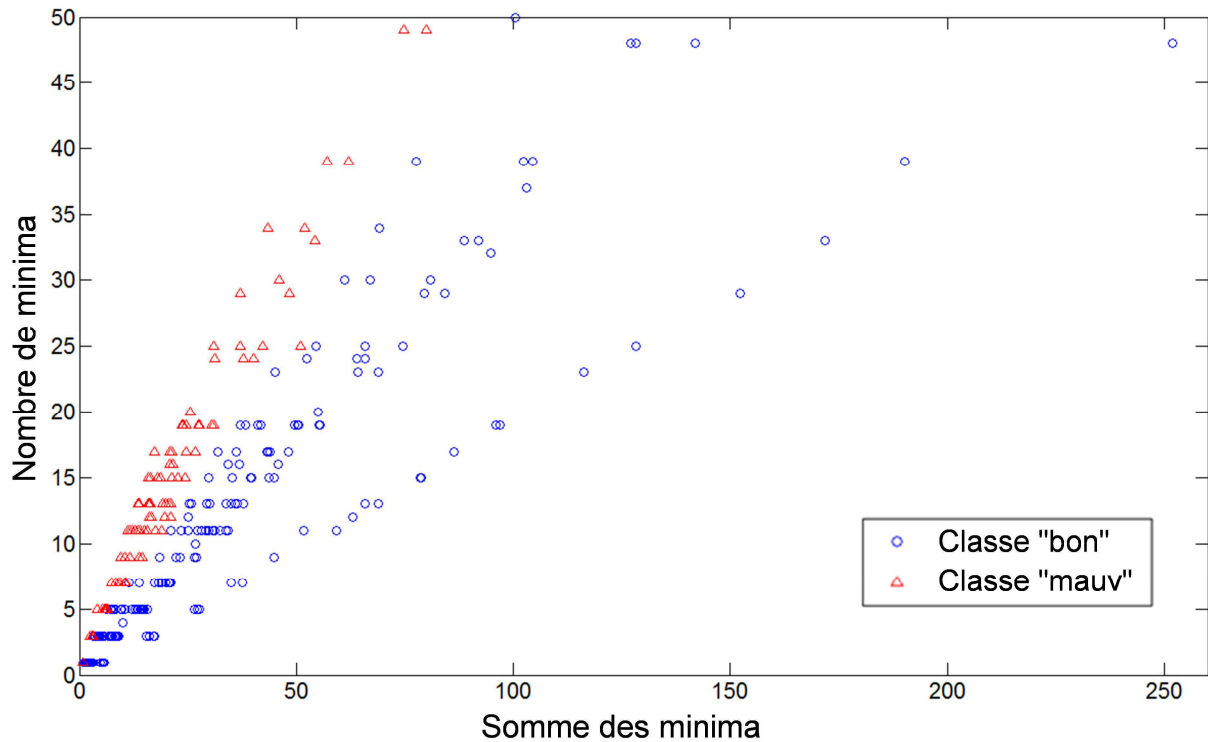


Figure 4.11. Diagramme de dispersion des éléments de base réf. 1

Comme nous pouvons le constater sur ce diagramme, nous parvenons à une séparation satisfaisante des deux classes : les éléments de la classe "mauv", matérialisés par des triangles rouges, sont sur la partie gauche du diagramme et ceux de la classe "bon", représentés par des cercles bleus, à droite. Cela signifie que notre descripteur et les deux paramètres qui en sont extraits sont pertinents pour la caractérisation de la qualité des signaux GSM-R. Toutefois, il apparaît que l'espace n'est pas uniformément couvert et que plusieurs zones des deux classes ne comportent pas ou peu de symboles. Cette absence de symboles peut en particulier nuire à la définition précise de la frontière entre les deux classes. Pour remédier à cela, il est nécessaire de considérer d'autres scénarios de bruit qui conduiront à de nouveaux couples (nombre de minima, somme des minima) à positionner sur le diagramme de dispersion.

En pratique, à bord des trains la qualité de la communication GSM-R est contrôlée par un paramètre estimé par le mobile : le Rxqual. Ce dernier évalue la qualité du signal. Il est obtenu en quantifiant le BER sur 8 niveaux suivant la correspondance définie dans le tableau suivant (plus la valeur de Rxqual est basse, meilleure est la qualité) [4.2].



RXQUAL	BER	
	Plage de valeurs	Valeur représentative
0	BER < 0.2 %	0.14 %
1	0.2 % < BER < 0.4 %	0.28 %
2	0.4 % < BER < 0.8 %	0.57 %
3	0.8 % < BER < 1.6 %	1.13 %
4	1.6 % < BER < 3.2 %	2.26 %
5	3.2 % < BER < 6.4 %	4.53 %
6	6.4 % < BER < 12.8 %	9.05 %
7	12.8 % < BER	18.10 %

**Figure 4.12. Tableau de correspondance entre le Rxqual et le BER**

La valeur représentative correspond à la moyenne géométrique des bornes de la plage et est calculée selon la formule :

$$\text{Valeur représentative} = \sqrt{2} \times 2^{\text{Rxqual}} \times 0.1 \% \quad 4.5$$

Ainsi ce ne sont pas des valeurs mais des plages de valeurs de BER qui servent à définir la qualité de la communication GSM-R. En particulier, les opérateurs ferroviaires considèrent que le Rxqual doit être inférieur ou égal à 3 pour assurer une qualité suffisante, ce qui correspond à un BER compris entre 0 et 1.6% [4.3].

Afin de nous rapprocher du critère appliqué à bord des trains, nous nous inspirons de ces intervalles pour définir nos classes. Cette fois-ci nous n'utilisons pas deux classes mais trois, afin de mettre en évidence sur le diagramme de dispersion la zone intermédiaire au-delà de laquelle on bascule de la partie "bonne qualité de communication" à la partie "mauvaise qualité de communication", et inversement. Dans ce cas, les trois classes ("bon", "mauv" et "inter" pour intermédiaire) sont choisies de la façon suivante :

- classe "bon" → BER ≤ 0.8%.
- classe "inter" → 0.8% < BER ≤ 1.6%
- classe "mauv" → BER > 1.6%.

En appliquant une telle répartition, nous obtenons le diagramme de dispersion de la figure 4.13. Sur ce graphe, les éléments appartenant à la troisième et nouvelle classe sont matérialisés par des carrés verts, les autres classes utilisent les symboles précédents.

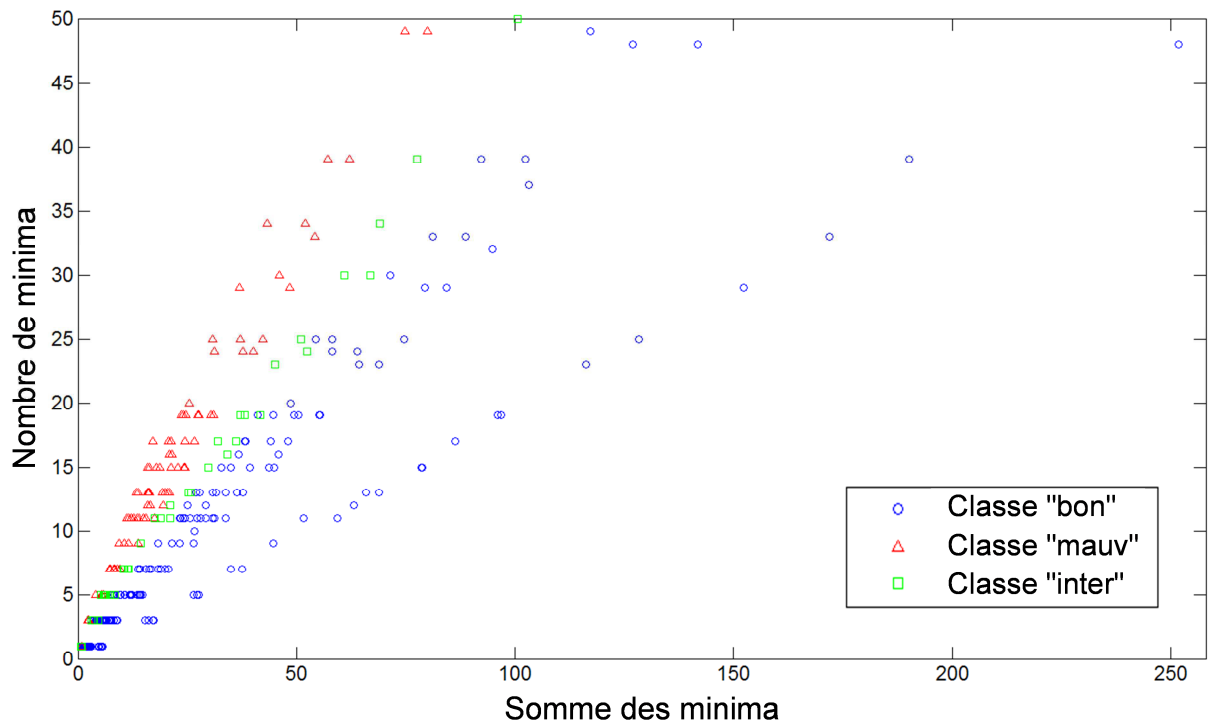


Figure 4.13. Utilisation de 3 classes pour le diagramme de dispersion

Sur ce nouveau diagramme il apparaît clairement que les éléments qui étaient précédemment proches de la frontière appartiennent à la classe "inter". De ce fait, la zone délimitée par ces éléments tient bien son rôle de zone de basculement entre les zones associées aux deux qualités initiales (bonne et mauvaise) de communication.

Dans l'optique d'une application à bord des trains, cette classe "inter" pourrait servir à alerter l'opérateur d'un risque de dégradation de la qualité de la communication GSM-R supérieure à ce que les spécifications tolèrent. Imaginons qu'au cours du déplacement du train, plusieurs points apparaissent successivement dans la partie du diagramme associée à cette troisième classe. Cela signifie que le train traverse une zone dans laquelle les conditions de bruit EM transitoire deviennent moins favorables au bon fonctionnement du GSM-R. A partir de là, nous pouvons envisager de mettre en place un processus d'alerte qui, dès lors qu'un certain nombre de points (à définir) successifs sont venus se positionner dans la zone "inter", en avertit le conducteur ou l'opérateur. Ainsi, ce dernier est alerté de la possibilité de traversée prochaine d'une "zone critique" et peut anticiper l'éventuelle dégradation à venir de la communication GSM-R.

### *II.5 - Définition du seuil de détection : importance pour la méthode de classification et choix retenu*

Une bonne définition du seuil de détection est indispensable au bon fonctionnement de notre méthode de classification. En effet, comme précédemment expliqué lors de la description de l'étape d'extraction des paramètres, c'est à partir de la valeur du seuil que les minima sont détectés et comptés. Ainsi, un seuil mal choisi entraîne des erreurs dans le comptage des minima. Or, si le nombre de minima est erroné, la somme de ces mêmes minima l'est également. La figure 4.14 illustre les trois cas possibles : seuil trop faible, seuil trop élevé et seuil correct.

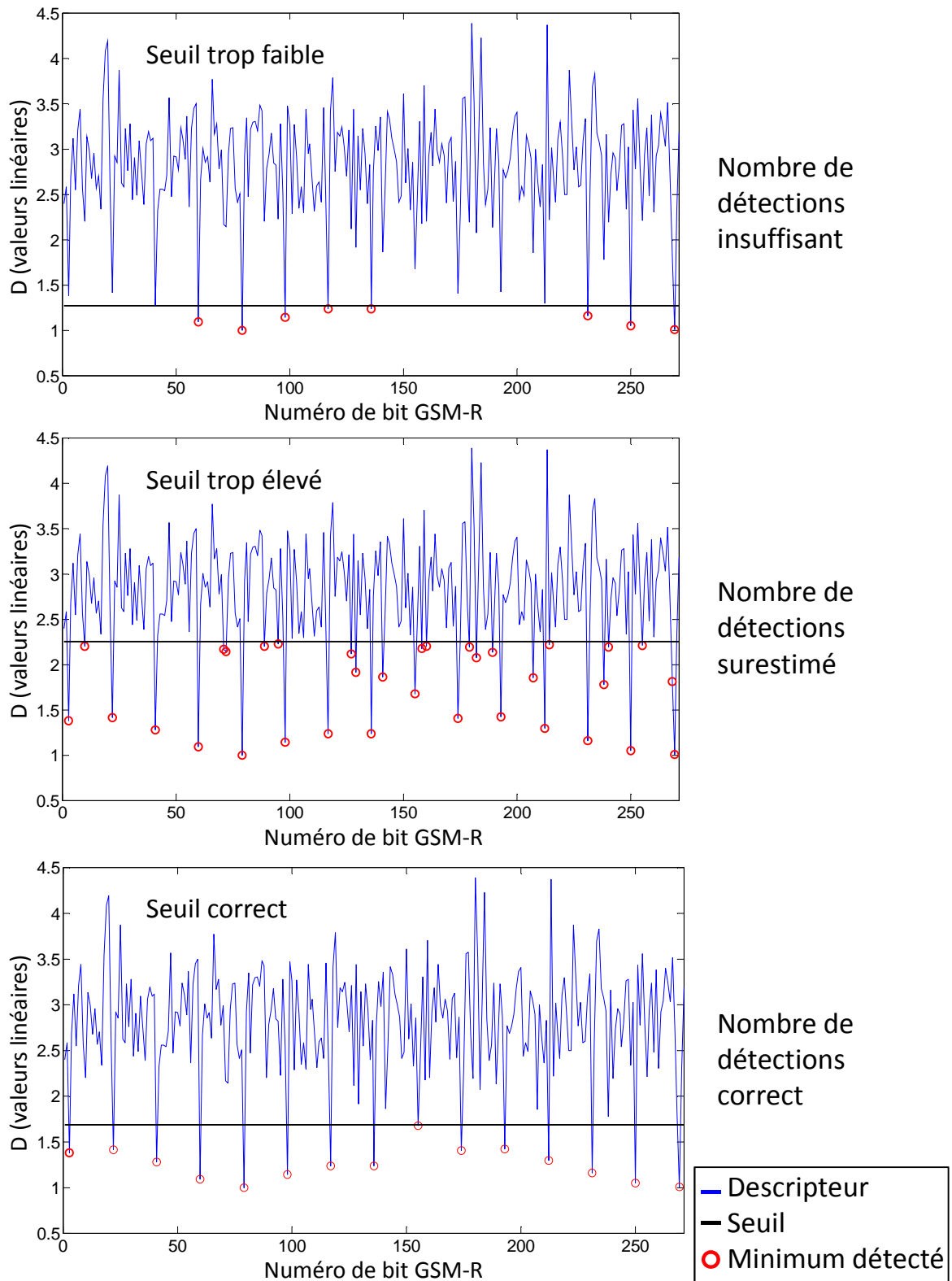
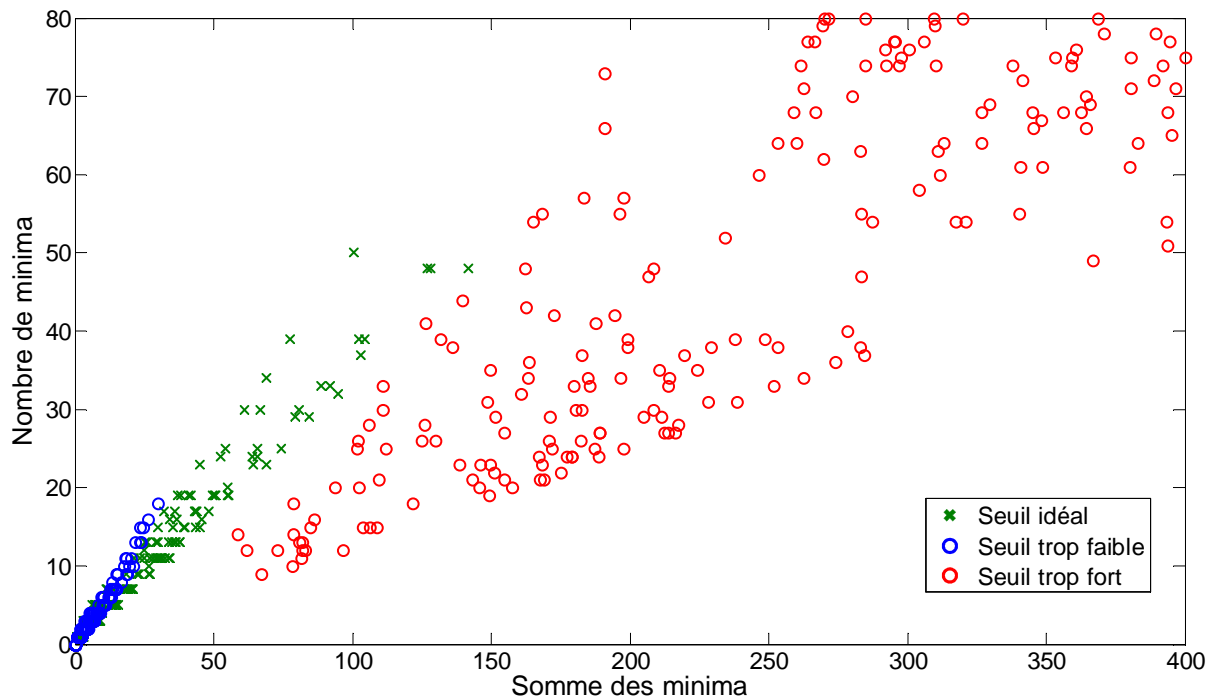


Figure 4.14. Détermination du meilleur seuil possible

Si le seuil est trop faible, certains minima ne sont pas détectés et le nombre de détections est insuffisant. Par conséquent les deux paramètres (nombre de minima et somme des minima) sont sous-estimés. Sur le diagramme de dispersion, les éléments sont alors décalés vers le coin inférieur gauche.

A l'inverse, s'il est trop élevé, de fausses détections sont introduites et, dans ce cas, il se produit une surestimation des deux paramètres et les éléments sont décalés vers le coin supérieur droit. Une illustration de ce phénomène est donnée en figure 4.15. Dans les deux cas, ces mauvais positionnements sur le diagramme de dispersion peuvent conduire à des erreurs de classification.



**Figure 4.15. Influence du seuil sur le diagramme de dispersion**

Pour traiter ce problème critique du seuil, nous avons choisi de le définir manuellement en visualisant la courbe du descripteur au cours du temps grâce à une interface graphique spécialement développée par Nabil Ech-Charraq au cours de son stage au sein de l'équipe CEM du LEOST [4.4]. Cette interface fait l'objet de la section III. Le principe est de visualiser la courbe du descripteur pour l'ensemble des fichiers de mesures. Nous déterminons par des essais de valeur successifs, de la même manière que précédemment présenté sur la figure 4.14, la valeur de seuil à appliquer pour correctement détecter et compter le nombre de minima avec le moins d'erreurs possible. Sur la figure 4.14, la visualisation du descripteur et du seuil est appliquée à un seul fichier, nous montrons dans la section III comment est réalisé le traitement de plusieurs fichiers de mesures successifs à l'aide de notre interface de calcul et visualisation.

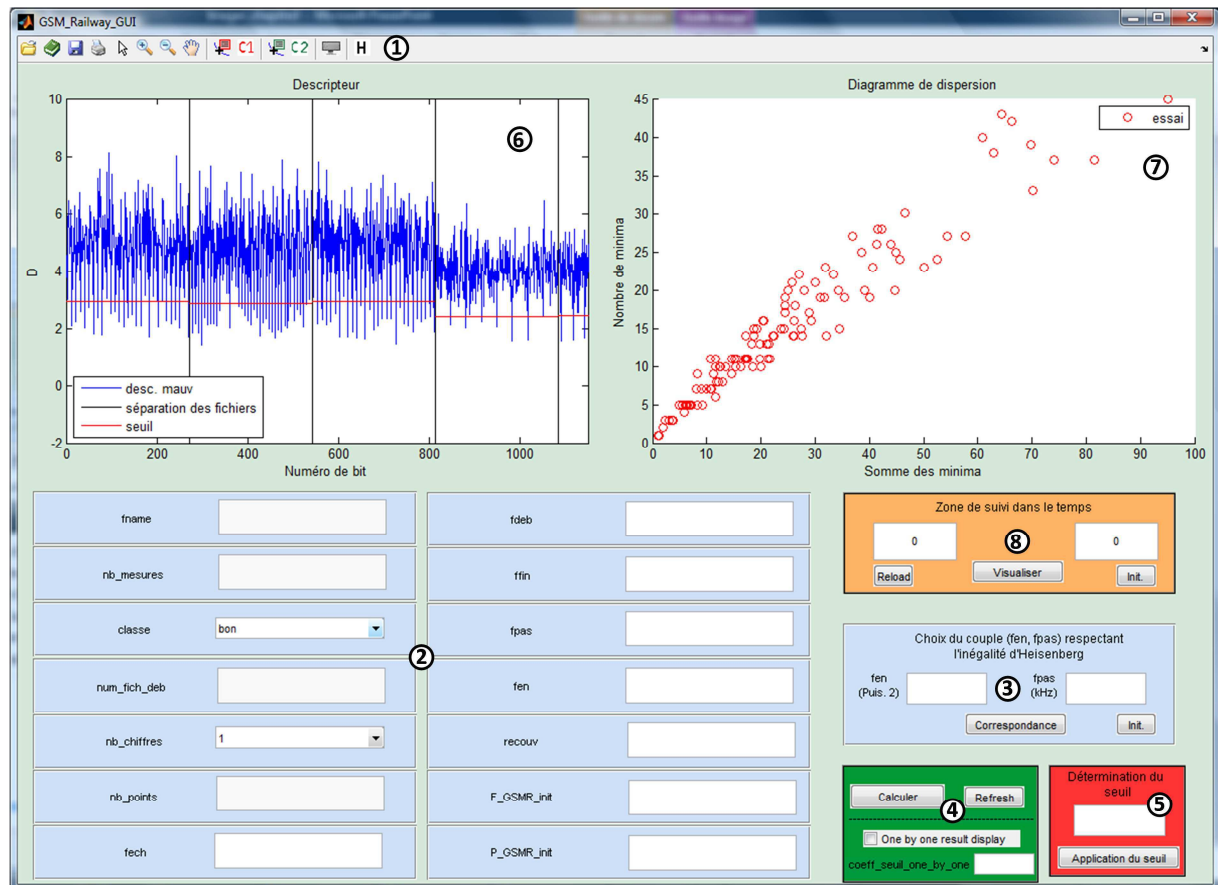
### III - Présentation de l'interface de calcul et visualisation

#### III.1 - Description générale de l'interface

Une interface graphique a été développée sous Matlab<sup>®</sup> afin de faciliter la mise en œuvre de notre méthode et l'exploitation des résultats [4.4]. Elle se présente sous la forme d'une unique fenêtre séparée en principalement trois parties :

- définition des paramètres de calcul (partie inférieure gauche de la figure 4.16),
- les 4 commandes développées : lancement du calcul, détermination et application du seuil, choix du couple (fen,fpas) et suivi dans le temps (partie inférieure droite de la figure 4.16),
- visualisation du descripteur et du diagramme de dispersion (partie supérieure de la figure 4.16).

Les différents outils, boutons, champs d'entrée et graphiques que comporte notre interface sont localisés et définis sur la figure 4.16. Nous revenons plus en détails sur chacun de ces éléments dans le paragraphe III.2.



- ① Barre d'outils
- ② Définition des paramètres de calcul
- ③ Choix du couple (fen,fpas) respectant l'inégalité d'Heisenberg
- ④ Commande de lancement du calcul
- ⑤ Détermination et application du seuil
- ⑥ Représentation du descripteur
- ⑦ Représentation du diagramme de dispersion
- ⑧ Définition de l'intervalle de suivi dans le temps

Figure 4.16. Vue d'ensemble de l'interface de calcul et visualisation

### III.2 - Description des différents champs, outils et commandes

#### III.2.a. Champs d'entrée des paramètres de calcul

Les paramètres de calcul servent d'entrée aux fonctions de notre algorithme : "lecture des fichiers de mesures", "calcul de la transformée de Gabor" et "détermination de la fréquence et de la puissance du signal GSM-R". Figure 4.17, un code couleur est utilisé pour les distinguer et mettre en évidence la fonction à laquelle ils sont associés. Notons que les valeurs "F\_GSMR\_init" et "P\_GSMR\_init" sont définies uniquement pour initier le calcul de la fréquence et de la puissance du signal GSM-R. En effet, pour appliquer la règle de conservation ( $P_{\text{GSM-R}}(t) = P_{\text{GSM-R}}(t-1)$ , si un transitoire est présent) à chaque instant  $t$ , il est nécessaire d'avoir des valeurs d'initialisation pour F\_GSMR et P\_GSMR. Ainsi, si un transitoire est détecté à l'instant  $t=0$  alors, nous imposons :  $F_{\text{GSM-R}}(0) = F\_GSMR\_init$  et  $P_{\text{GSM-R}}(0) = P\_GSMR\_init$ .

fname	C1mesure_2_classes_bon_	fdeb	920.8e6
nb_mesures	3	ffin	928.6e6
classe	bon	fpas	100e3
num_fich_deb	mauv	fen	2048
nb_chiffres	1	recouv	0.5
nb_points	2.5e6	F_GSMR_init	924.8e6
fech	2.5e9	P_GSMR_init	-59.9

Lecture des fichiers de mesures
  Calcul de la transformée de Gabor
  Détermination de la fréquence et de la puissance du signal GSM-R

Figure 4.17. Définition des paramètres de calcul

#### III.2.b. Choix du couple (fen, fpas) respectant l'inégalité d'Heisenberg

Une fonction a été développée pour faciliter le choix du couple (fen, fpas) utilisé pour le calcul de la transformation de Gabor et qui doit respecter l'inégalité d'Heisenberg :

$$\sigma_t \sigma_f \geq \frac{1}{4\pi} \quad 4.6$$

avec  $\sigma_t$  : résolution temporelle et  $\sigma_f$  : résolution fréquentielle

Après avoir défini la fréquence d'échantillonnage, il suffit de fixer, par exemple, la largeur de la fenêtre de Gabor (fen) pour que la fonction établisse automatiquement la correspondance avec le pas en fréquence (fpas). Cette situation correspond aux parties (a) et (b) de la figure 4.18. Inversement, nous pouvons choisir de fixer fpas et d'obtenir la valeur de fen correspondante, comme illustré sur la partie (c) de cette même figure.



Figure 4.18. Choix du couple (fen, fpas)

Il est à noter que pour faciliter nos calculs, nous travaillons uniquement avec des fenêtres dont la largeur (en nombre de points) est une puissance de 2. Si une valeur quelconque est entrée pour fen (cas (b) sur la figure 4.18), alors la fonction ajuste automatiquement la valeur de fen pour qu'elle soit égale à la puissance de 2 immédiatement supérieure et signale ce changement par une coloration du champ d'entrée de fen en rouge. De même, lorsque l'utilisateur fixe la valeur de fpas, la fonction retourne la valeur correspondante de fen sous la forme d'une puissance de 2.

### III.2.c. Détermination et application du seuil de détection des minima

Une fois les paramètres d'entrée correctement définis, l'appui sur le bouton "Calculer" lance l'algorithme qui, dans un premier temps, réalise le calcul du descripteur pour tous les fichiers de mesures sélectionnés. Les courbes associées sont ensuite tracées les unes à la suite des autres de sorte à obtenir le tracé du descripteur au cours du temps sur la durée totale de mesure (soit sur le nombre de fichiers de mesures traités multiplié par 1 ms). Sur ce tracé, chaque fichier est également délimité par des lignes verticales comme sur les graphes de la figure 4.19.

C'est à partir de la visualisation de cette courbe qu'est déterminé le seuil à appliquer pour compter les minima. Le seuil est défini comme un pourcentage de la valeur moyenne du descripteur afin de tenir compte des variations de l'amplitude du descripteur qui peuvent se produire d'une mesure à l'autre. La valeur du pourcentage est fixée visuellement, en traçant le seuil (sous la forme de droites horizontales) directement sur la courbe du descripteur pour différentes valeurs du pourcentage. Si le pourcentage, donc le seuil, est mal choisi (trop élevé ou trop faible), l'utilisateur le constate immédiatement sur la courbe du descripteur et peut ainsi le corriger. De plus, l'appui sur le bouton "Application du seuil" provoque le tracé automatique du diagramme de dispersion à partir des minima détectés sur la courbe du descripteur avec le seuil choisi. Il est donc également possible de vérifier que les paramètres, nombre de minima et somme des minima, ont été correctement extraits du descripteur avec le seuil défini. La figure 4.20 présente deux exemples (volontairement excessifs pour illustrer nos propos) de mauvais choix de seuil et illustre les conséquences de l'application d'un seuil trop élevé ou trop faible.

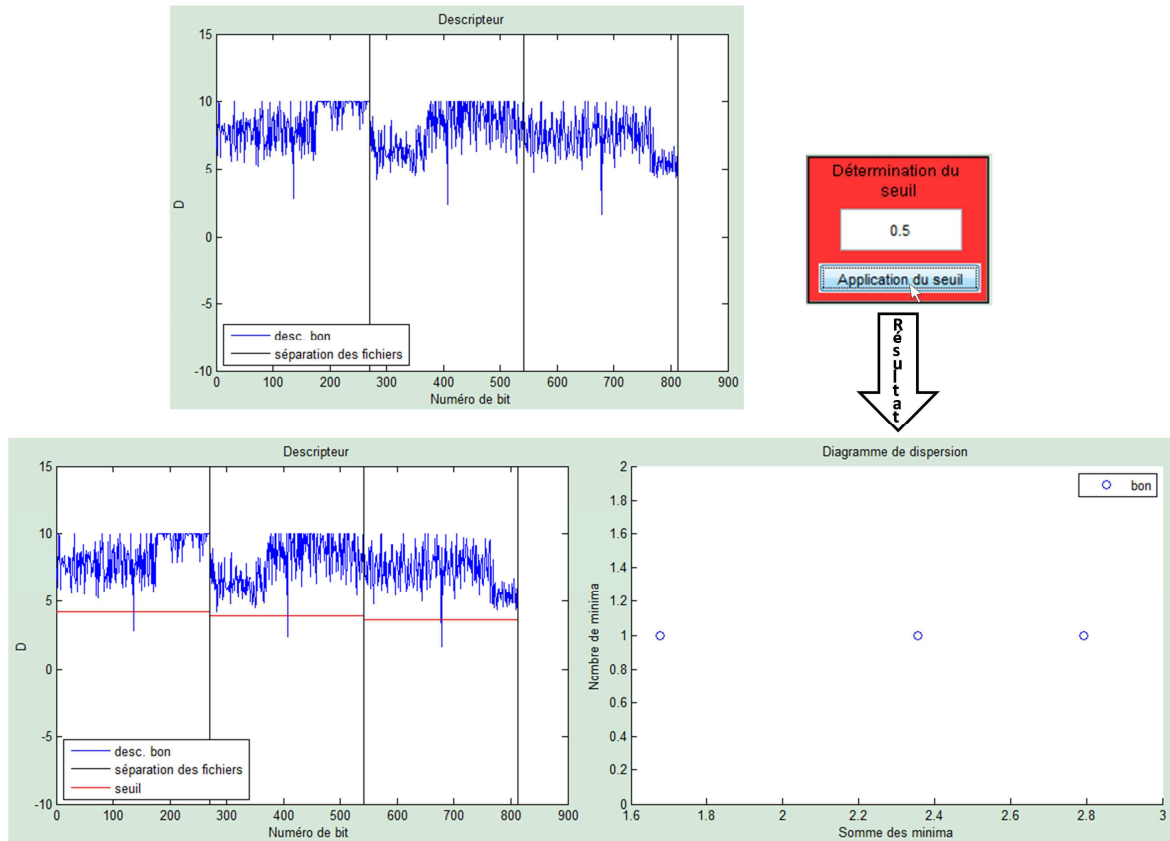


Figure 4.19. Détermination et application du seuil

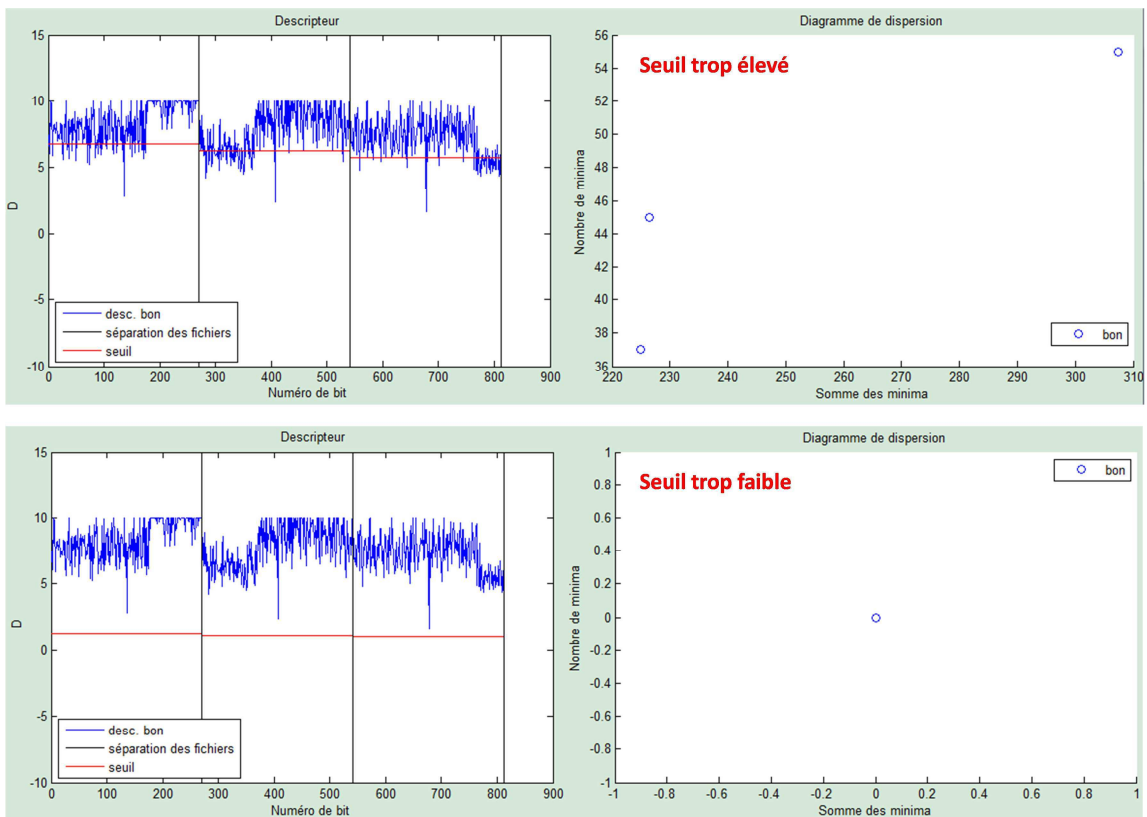


Figure 4.20. Exemples de mauvais choix de seuil



Sur ces exemples, si le seuil est trop élevé, comme sur la partie supérieure de la figure 4.20, de nombreux "faux minima" sont comptés sur la courbe du descripteur. Ainsi, sur le diagramme de dispersion le nombre de minima "explose" : 137 minima sont détectés sur l'ensemble des 3 fichiers de mesures traités quand les "vrais minima", résultant de l'apparition véritable d'un transitoire, ne sont en réalité qu'au nombre de 3. A l'inverse, si le seuil est trop faible (partie inférieure de la figure 4.20), le nombre de détections n'est pas suffisant et les deux paramètres (nombre et somme) sont sous-estimés. Aucun minimum n'est d'ailleurs détecté sur ce deuxième exemple.

Définir manuellement le seuil à appliquer en visualisant la courbe du descripteur et le diagramme de dispersion obtenu permet d'éviter les erreurs qui pourraient se produire dans le cas de l'application entièrement automatique d'un seuil prédéfini. C'est la raison pour laquelle nous avons retenu cette solution qui, à ce stade d'avancement de nos travaux, se révèle être la mieux adaptée. En effet, la définition d'un seuil automatique nécessite d'avoir de nombreux retours d'expérience et d'utilisation de notre méthode, ce dont nous ne disposons pas pour le moment.

III.2.d. Utilisation de l'intervalle de suivi dans le temps

Il peut être nécessaire de procéder à un suivi chronologique des éléments positionnés sur le diagramme de dispersion. Notamment si l'on cherche à mettre en évidence le fait que plusieurs points successifs sont venus se placer dans une zone du diagramme de dispersion, et donc que plusieurs éléments d'une même classe ont été successivement détectés. Ce point a été exposé dans la section II.4 lors de l'introduction de la classe "inter" et sa possible utilisation dans l'optique d'une application de notre méthode à bord des trains.

L'intervalle de suivi dans le temps permet de localiser et de tracer uniquement, à la fois sur la courbe du descripteur et sur le diagramme, les éléments associés à un certain nombre de fichiers de mesures successifs. Il suffit pour cela de choisir les numéros des fichiers de départ et de fin de l'intervalle et d'appuyer sur le bouton "Visualiser". La figure 4.21 reprend l'exemple des trois fichiers de mesures de la figure 4.19 avec application de l'intervalle [1,2]. Ainsi, seuls les fichiers n° 1 et 2 ont leur courbe du descripteur tracée et leur couple (nombre de minima, somme des minima) positionné et numéroté sur le diagramme de dispersion.

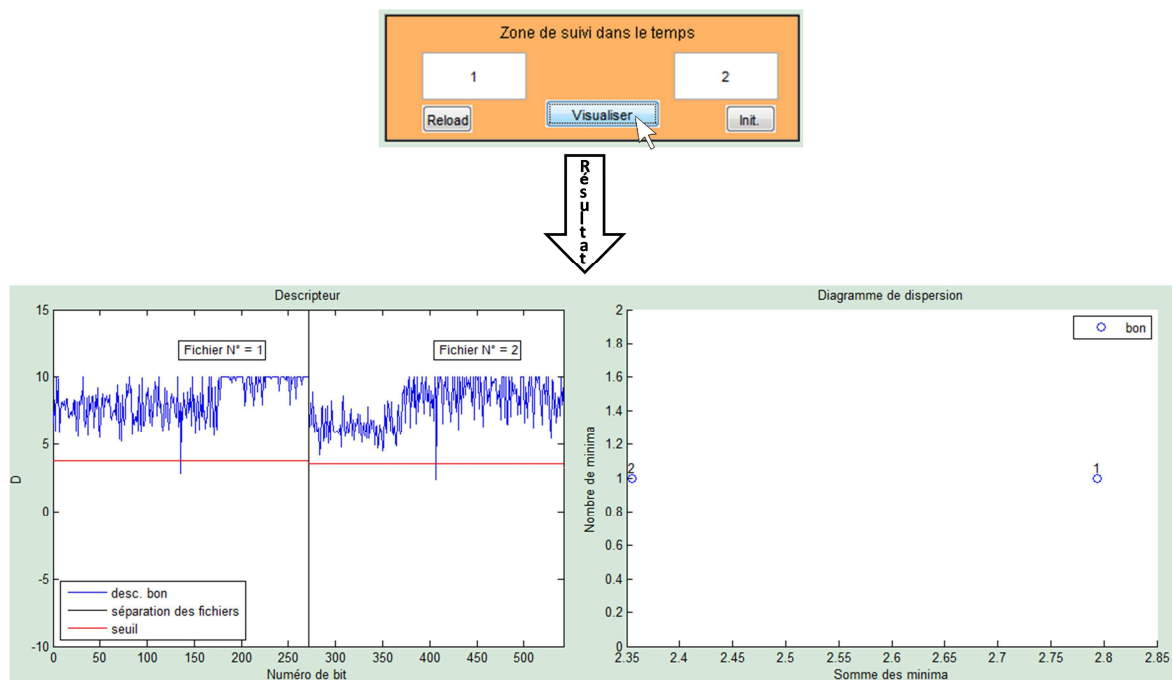


Figure 4.21. Intervalle de suivi dans le temps

III.2.e. Correspondance entre descripteur et diagramme de dispersion

Toujours dans le but d'identifier certains éléments, une fonction a été développée afin d'établir une correspondance entre courbe du descripteur et diagramme de dispersion. La première étape consiste à sélectionner, par un clic de la souris, le point associé au couple (nombre de minima, somme des minima) sur le diagramme (respectivement la courbe du descripteur) d'un fichier de mesures. S'affiche alors, à l'emplacement du clic, une petite fenêtre comportant le numéro du fichier, le nombre de minima détectés et la somme des minima. Par exemple, sur la figure 4.22 c'est le point associé au fichier n°2 qui est sélectionné sur le diagramme de dispersion et sur la figure 4.23, il s'agit du descripteur du fichier n°3. L'appui sur le bouton "C2" (respectivement "C1") conduit à l'affichage de la courbe du descripteur du fichier n°2 et du seuil qui lui a été appliqué (respectivement, la coloration en rouge du point n°3 du diagramme de dispersion).

L'établissement d'une correspondance entre courbe du descripteur et diagramme de dispersion peut avoir deux applications :

- identifier la classe d'un (ou plusieurs) fichier(s) de mesures qui nous intéresse(nt) tout particulièrement (sens descripteur → diagramme, illustré par la figure 4.23),
- pour le(s) point(s) dont la position sur le diagramme de dispersion est surprenante, vérifier que le seuil appliqué permet de correctement détecter les minima sur la courbe du descripteur (sens diagramme → descripteur, illustré par la figure 4.22).

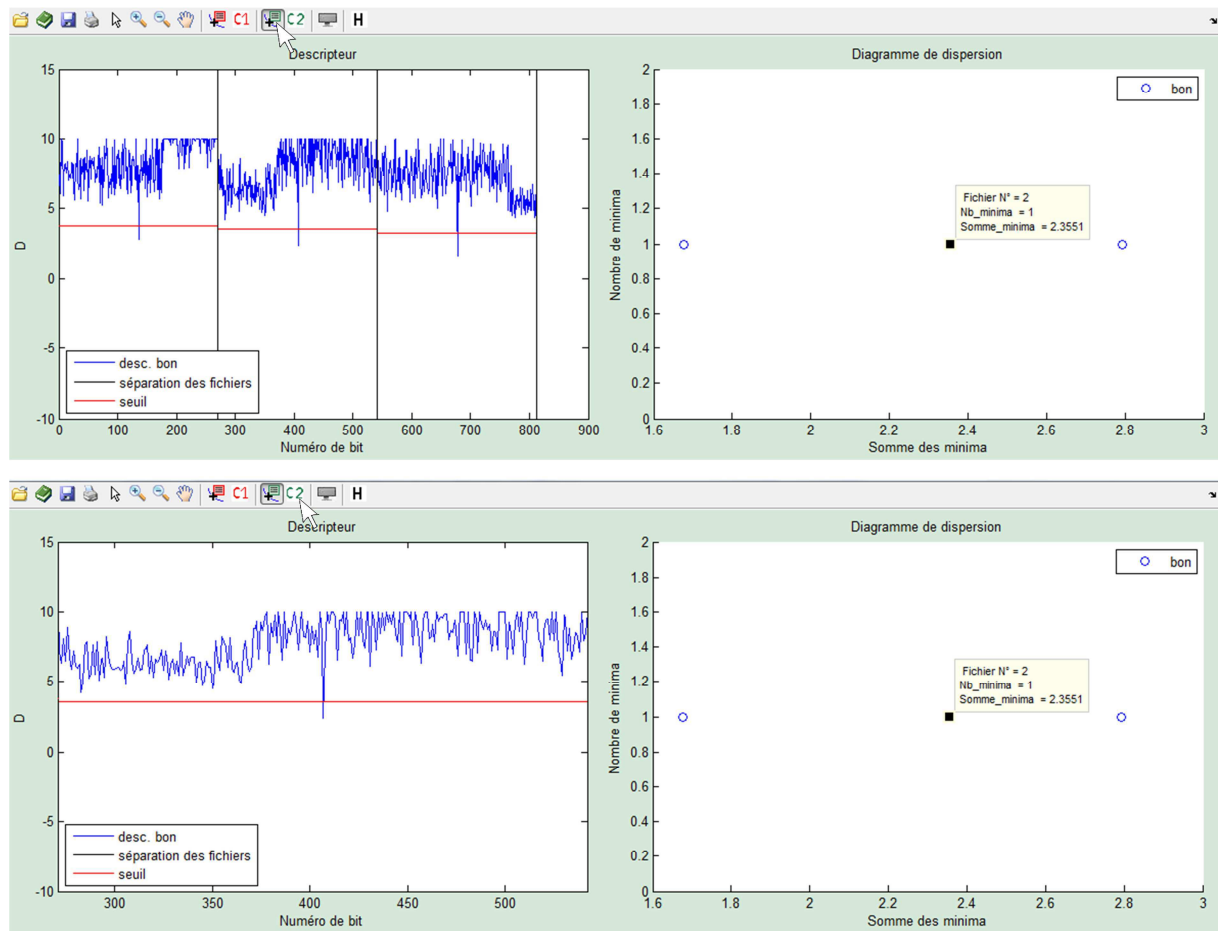


Figure 4.22. Correspondance diagramme de dispersion vers descripteur

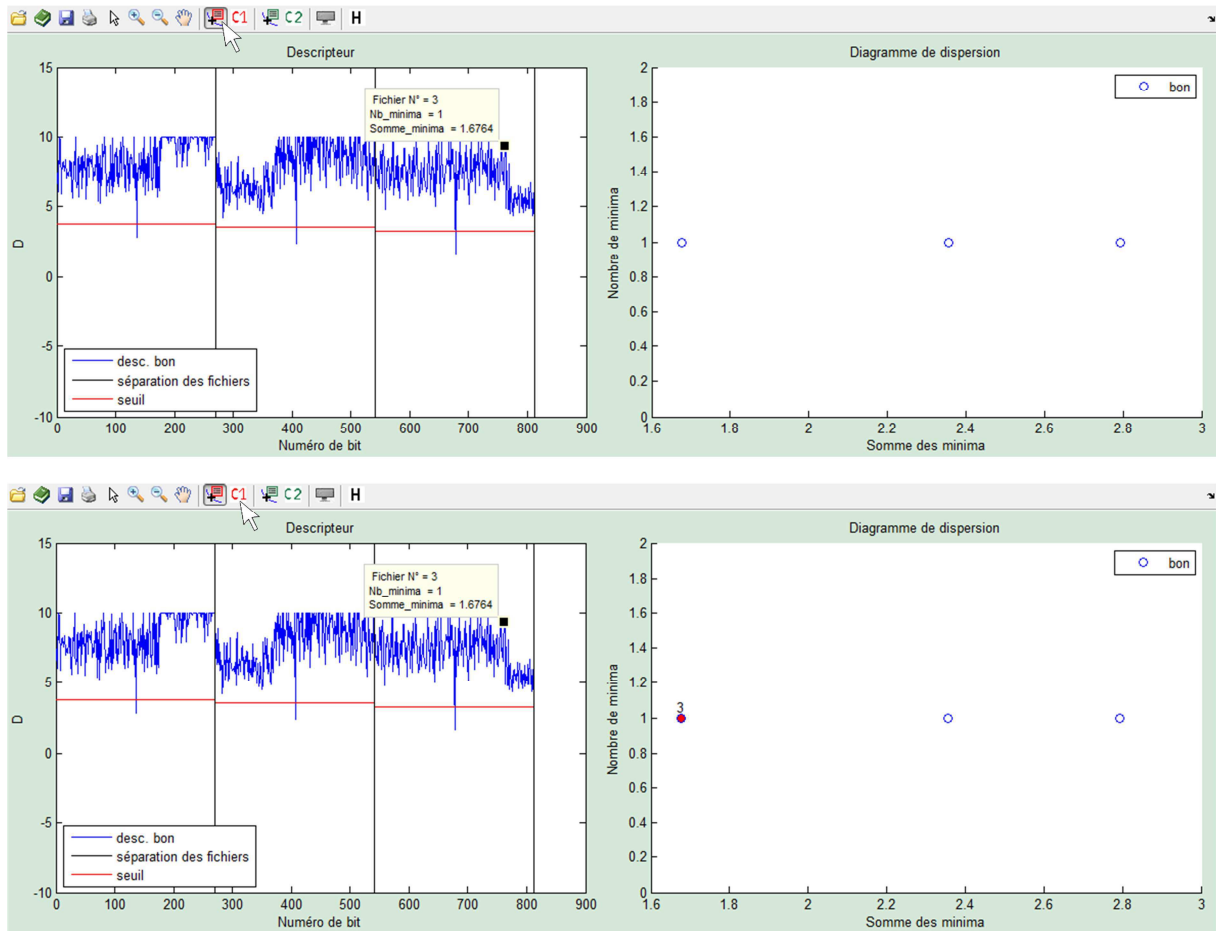


Figure 4.23. Correspondance descripteur vers diagramme de dispersion

*III.2.f. Barre d'outils*

Les autres fonctionnalités de notre interface sont regroupées sur la barre d'outils située dans le coin supérieur gauche de la fenêtre. Outre les fonctions de correspondance présentées dans le paragraphe précédent, il s'agit en fait de fonctions classiques : zoom, impression, enregistrement des graphes sous la forme d'images, sauvegarde et chargement des résultats (descripteur, seuil et diagramme de dispersion).



Figure 4.24. Barre d'outils

## Conclusion

Le principe de notre méthode de classification a été décrit étape par étape et appliqué à un scénario de bruit EM transitoire défini. Pour être applicable en environnement ferroviaire notre méthode nécessite l'établissement d'une base de référence représentative de toutes les conditions de bruit EM transitoire que peut y rencontrer le GSM-R. Il faut donc considérer d'autres scénarios de bruit (qui couvrent l'ensemble des conditions d'opération possibles à bord des trains) et leur appliquer le traitement qui vient d'être présenté de manière à définir de nouveaux couples (nombre de minima, somme des minima) à stocker dans la base. Une fois achevée la construction de la base de référence complète, le diagramme de dispersion qui lui est associé est tracé et devient le diagramme de référence. Au final, la classification des nouveaux éléments (dont la classe n'est pas connue) est réalisée "à l'aveugle", c'est-à-dire uniquement à partir du diagramme de référence. Ces derniers sont positionnés sur le diagramme de référence et leur classe est déterminée en fonction de la zone qu'ils occupent.

Avant de procéder à une classification "à l'aveugle", il convient de s'assurer de la robustesse de la méthode. Pour cela des tests sont effectués sur des scénarios construits à partir de signaux transitoires collectés à bord d'un train. Les fichiers de mesures enregistrés à l'oscilloscope sont labellisés en fonction de leur classe. Le traitement leur est appliqué afin d'extraire les couples (nombre de minima, somme des minima) qui sont ensuite placés sur le diagramme de référence. La méthode est validée si la grande majorité des éléments de test sont positionnés dans la zone correspondant effectivement à leur classe. Après validation, les éléments de test correctement identifiés sont ajoutés dans la base de référence afin de l'enrichir d'avantage et des essais à bord des trains peuvent être envisagés.

Le chapitre 5 présente les résultats obtenus par notre méthode de classification, de la construction de la base de référence complète jusqu'aux tests sur des signaux réellement acquis en environnement ferroviaire.

# Chapitre 5

## Résultats de la méthode de classification

---

### SOMMAIRE

---

<i>I - Construction de la base de référence avec les différents scénarios créés/simulés</i>	<b>104</b>
<i>II - Tests sur des signaux réellement acquis à bord d'un train</i>	<b>109</b>
<i>III - Tests sur des signaux acquis dans le métro de Barcelone</i>	<b>117</b>
<i>IV - Vers l'application automatique d'un seuil pour détecter et compter les minima</i>	<b>122</b>
<i>Conclusion</i>	<b>125</b>

---

Après avoir détaillé la méthode de classification des bruits EM transitoires en fonction de leur effet sur la qualité d'une communication GSM-R, ce chapitre présente les résultats obtenus lors de son évaluation. Comme nous l'avons expliqué à la fin du chapitre 4, nous avons besoin d'établir une base de référence la plus complète possible, cette base permettant par la suite de classer à l'aveugle tout nouvel élément inconnu. Dans le chapitre 2, nous avons donné un aperçu de quelques techniques de classification. Nous avons choisi la méthode des  $k$  plus proches voisins car elle nous semblait la plus simple à mettre en œuvre. C'est donc cette technique de classification qui est évaluée tout au long de ce chapitre. Une première évaluation est menée sur la base de mesures réalisées en laboratoire. Elle nécessite de considérer de multiples scénarios de bruit EM transitoire de manière à couvrir l'ensemble des situations de bruit auxquelles peut être confronté le GSM-R à bord des trains. Il est donc nécessaire d'effectuer un grand nombre de mesures. Une seconde évaluation est proposée en considérant cette fois des signaux transitoires acquis en environnement ferroviaire. L'objectif est de tester notre méthode sur des signaux réellement acquis en environnement ferroviaire afin de la valider complètement avant les essais en situation réelle de fonctionnement à bord d'un train sur une ligne équipée du système GSM-R.

Dans la première partie de ce chapitre, nous décrivons la méthode ayant permis de construire la base de référence. Nous présentons la technique de classification appliquée et procédons à son évaluation. Les deux autres sections sont consacrées à deux évaluations menées sur des signaux transitoires collectés à bord d'un train et à bord d'un métro. Dans ces deux dernières sections, nous commençons par présenter les scénarios de bruit EM transitoire construits à partir de ces signaux réels, puis nous fournissons les résultats de classification obtenus et nous procédons à leur analyse.

## I - Construction de la base de référence avec les différents scénarios créés/simulés

### 1.1 - Présentation des scénarios considérés

Plusieurs scénarios de bruit EM transitoire sont créés afin de diversifier les conditions de bruit EM auxquelles est confronté le GSM-R. Le but est de couvrir l'ensemble des scénarios susceptibles d'être rencontrés par le système de communication dans son environnement de fonctionnement. C'est en étudiant l'impact de chacun des scénarios envisagés que la base de référence est construite.

L'analyse de signaux transitoires collectés à bord de trains lors de plusieurs campagnes de mesures a permis d'établir les deux modèles caractéristiques de transitoire (sinus-expo et double exponentielle) précédemment présentés au chapitre 3.

Nous utilisons ces modèles pour créer nos scénarios de test. Tous sont construits sur la même base : un ou plusieurs modèle(s) de transitoire répété(s) avec un intervalle de temps (TI) fixe entre deux transitoires successifs. Cet intervalle prend 32 valeurs et varie entre de  $20 \mu\text{s}$  et  $1 \text{ ms}$ . Le scénario de bruit EM transitoire n°1 (déjà présenté au chapitre 4) se compose d'un sinus-expo. Le scénario n°2 est construit à partir du modèle double exponentielle filtré par un passe-haut. Les scénarios n°3 et 4 sont tous deux du même type : deux sinus-expo espacés respectivement de  $5 \mu\text{s}$  et  $100 \mu\text{s}$ . La figure 5.1 propose une illustration des quatre scénarios ainsi créés.

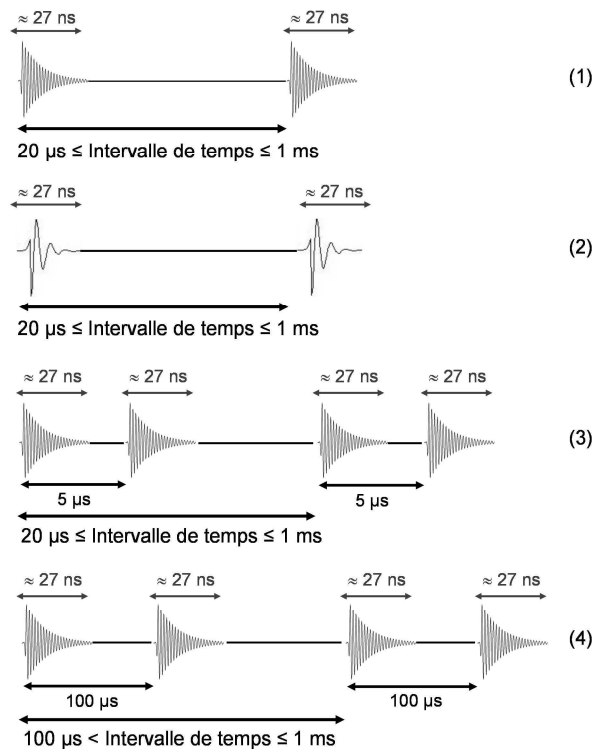


Figure 5.1. Scénarios de test n° 1, 2, 3 et 4

### 1.2 - Diagramme de dispersion des éléments de la base de référence

L'ensemble des configurations des quatre scénarios est produit en faisant varier le SNR et l'intervalle de temps entre deux transitoires selon les mêmes règles que celles décrites dans le chapitre 4. Au final, ce sont 530 éléments appartenant à la classe "bon" et 255 à la classe "mauv" qui viennent compléter la base de référence initialement construite à partir des éléments issus des tests avec le scénario n°1, portant ainsi à 776 le nombre de "bon" et 356 le nombre de "mauv". Le traitement décrit au chapitre précédent est ensuite appliqué à chaque fichier de mesures. Le diagramme de dispersion des éléments de cette base de référence est présenté à la figure 5.2. Nous observons que l'ajout de couples (nombre de minima, somme des minima) permet de mieux couvrir l'espace des paramètres et par conséquent de mieux identifier les parties de l'espace relatives à l'une et à l'autre des deux classes. En revanche, il apparait que la frontière entre les deux sous-espaces n'est pas linéaire. Cette non-linéarité se traduit par des chevauchements entre des éléments de deux classes distinctes. Nous avons montré dans le chapitre précédent que le nombre de minima et leur somme dépendaient de la valeur du seuil appliqué. Nous observons que ce seuil influe également sur la forme de la frontière. En effet, les chevauchements observés sont en fait des erreurs de "positionnement" provenant d'un mauvais comptage des minima lors de l'application du seuil de détection, ce qui engendre des erreurs à la fois sur le nombre et la somme. Selon la méthode de classification choisie, la non-linéarité de cette frontière sera à l'origine de la mauvaise classification des éléments inconnus dont la classe devra être déterminée uniquement à partir de la base de référence. Dans la section suivante, nous analysons le diagramme de dispersion en construisant la matrice de confusion des éléments de la base de référence.

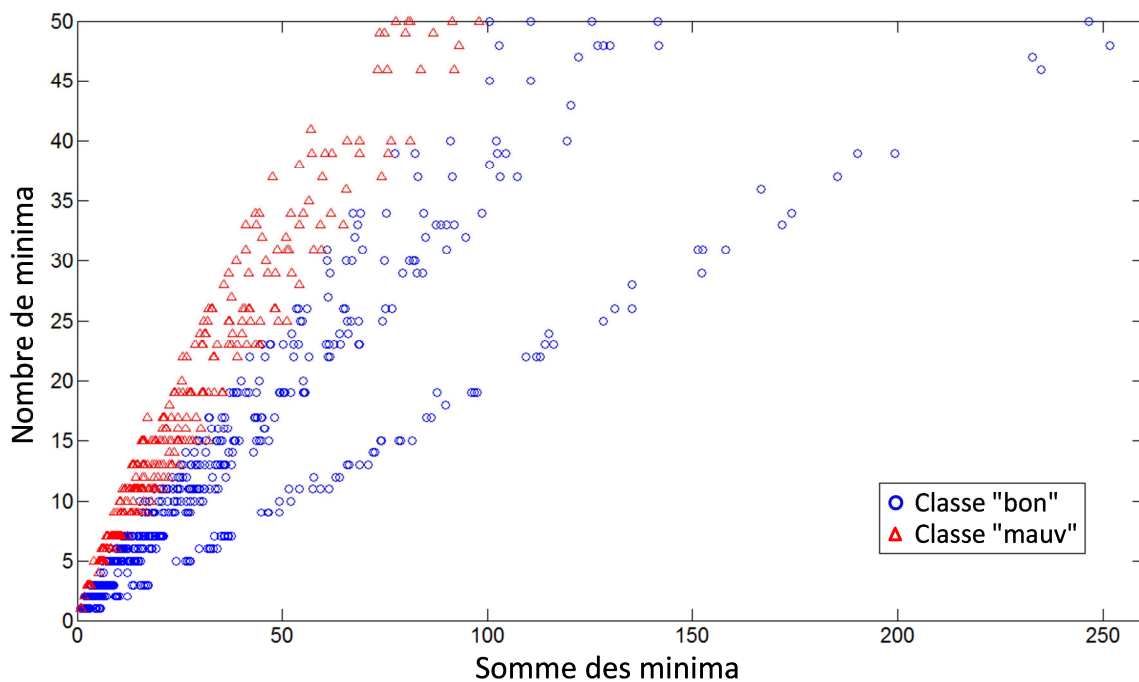


Figure 5.2. Diagramme de dispersion de référence

### 1.3 - Analyse et amélioration du diagramme de dispersion de référence

Comme nous l'indiquons dans l'introduction de ce chapitre, la technique des k-ppv a été retenue pour affecter chaque élément à classer à l'une ou l'autre des deux catégories car dans un espace de dimension 2 comme le nôtre, c'est la méthode la plus rapide et la plus facile à implémenter. Mais

d'autres méthodes existent comme nous l'avons vu au chapitre 2. Chaque nouvel élément est alors classé à partir de sa distance à ses plus proches voisins contenus dans une base d'apprentissage. Cette base d'apprentissage est constituée par notre base de référence.

Lorsque la base contient un nombre insuffisant d'éléments, la littérature préconise d'appliquer le principe du "leave-one-out" pour évaluer les performances du classifieur. Il s'agit de sortir un élément de la base et d'utiliser tous les autres pour le classer. La qualité du résultat de classification est mesurée à partir du contenu d'une matrice appelée "matrice de confusion". Cette matrice présente les taux de bonne et mauvaise classification de chaque classe.

Dans cette première évaluation, nous utilisons la règle des trois plus proches voisins. La règle des trois plus proches voisins reprend le même principe que la règle du plus proche voisin, décrite au chapitre 3, avec trois voisins au lieu d'un seul. Ainsi, on attribue à l'élément sorti, la classe des trois éléments dont il est le plus proche en termes de distance euclidienne. Ensuite, cet élément entre à nouveau dans la base et un autre en sort pour être classé à son tour selon le même principe.

La matrice de confusion est présentée à la figure 5.3. Elle montre un taux de bonne classification de 96.8% pour les éléments de la classe "bon" et de 91.3% pour ceux de la classe "mauv".

		(Estimé)	
		Bon	Mauv
(Réel)	Bon	751/776 0.968	25/776 0.032
	Mauv	31/356 0.087	325/356 0.913

Figure 5.3. Matrice de confusion de la base de référence

Nous remarquons que 25 couples "bon" sont classés comme "mauv" et, inversement 31 "mauv" sont classés comme "bon". Ces erreurs de classification s'expliquent, d'une part, par la nature non-linéaire de la frontière entre les deux classes, une non-linéarité qui s'illustre par des chevauchements comme évoqués dans la section I.2.

D'autre part, elles peuvent s'expliquer par une mauvaise labellisation lors de l'enregistrement des fichiers de mesures incriminés. Dans ce cas, la classe réelle affectée à l'élément n'est pas correcte. Comme vu au chapitre 3, la classe est attribuée en fonction de la valeur de BER mesurée. Or, il faut savoir que le BER n'est pas une mesure entièrement reproductible, sa valeur peut varier sur deux mesures consécutives. Pour illustrer ce phénomène, nous présentons en figure 5.4 (en page suivante) le résultat de 100 mesures consécutives du BER en présence d'évènements transitoires pour une même configuration de test (SNR = 0 dB et TI = 200 µs).

Il y apparaît que la valeur mesurée du BER varie de manière plus ou moins importante autour d'une valeur moyenne de 0.89%. Dans la plupart des cas, ces fluctuations sont sans conséquence pour la labellisation. Le problème se pose pour les valeurs de BER proches de 1.13% pour lesquelles on peut basculer d'une classe à l'autre du fait que le BER peut être inférieur à la valeur-limite entre les deux classes sur une mesure et supérieur à la suivante. C'est d'ailleurs le cas dans notre exemple : sur la totalité des mesures, la valeur du BER est à peine supérieure à la valeur-limite (1.13%) une seule fois (1.131% à la deuxième mesure). Etant donné le nombre important de configurations à considérer et par conséquent le nombre de mesures à réaliser, nous procédons à la labellisation d'un fichier de mesures à partir d'une seule mesure du BER. Dans le cas présenté à la figure 5.4, si la valeur du BER avait été 1.131% à la première mesure, le fichier aurait été classé comme "mauv" alors que les 99 autres mesures (suivantes) du BER assuraient qu'il appartenait à la classe "bon".



Il est difficile de déceler ce genre d'erreurs lors de la collecte des fichiers, c'est pourquoi il est nécessaire d'y remédier suite à l'analyse de la base.

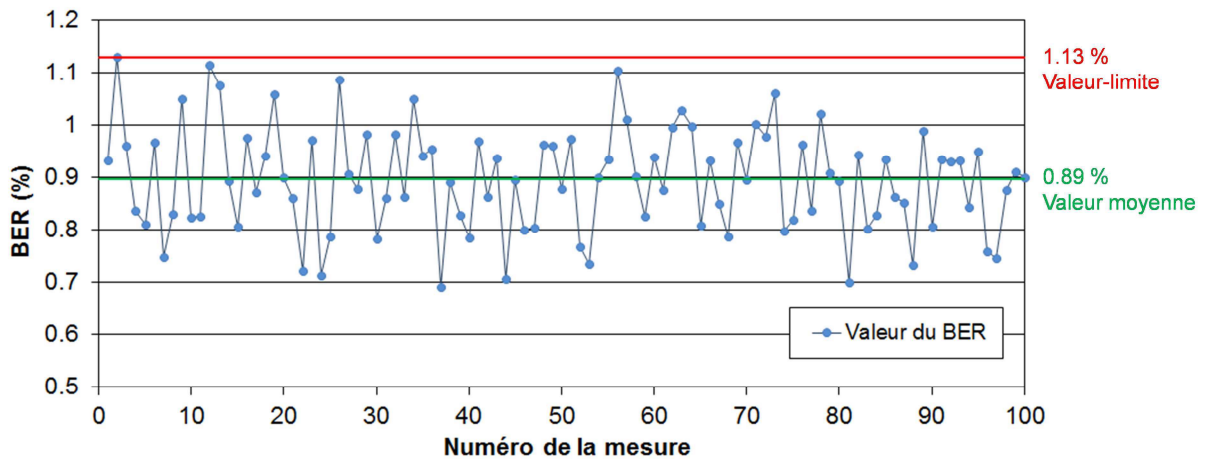


Figure 5.4. Valeurs du BER sur 100 mesures consécutives réalisées dans des conditions de test identiques

Rappelons que la base de référence constitue notre base d'apprentissage, par conséquent elle doit être fiable à 100% et ne pas comporter d'éléments qui peuvent être des sources d'erreur pour les classifications à venir. Par conséquent, les 56 couples "suspects" précédemment identifiés sont retirés de la base de référence. Nous ne cherchons pas à identifier individuellement la source de ces erreurs car ce n'est pas le but recherché. De plus, en raison du grand nombre d'éléments qui constituent notre base, le fait d'ôter ces quelques éléments "douteux" ne nuit pas à notre application.

Le nouveau diagramme de dispersion ainsi obtenu est donné par la figure 5.5.

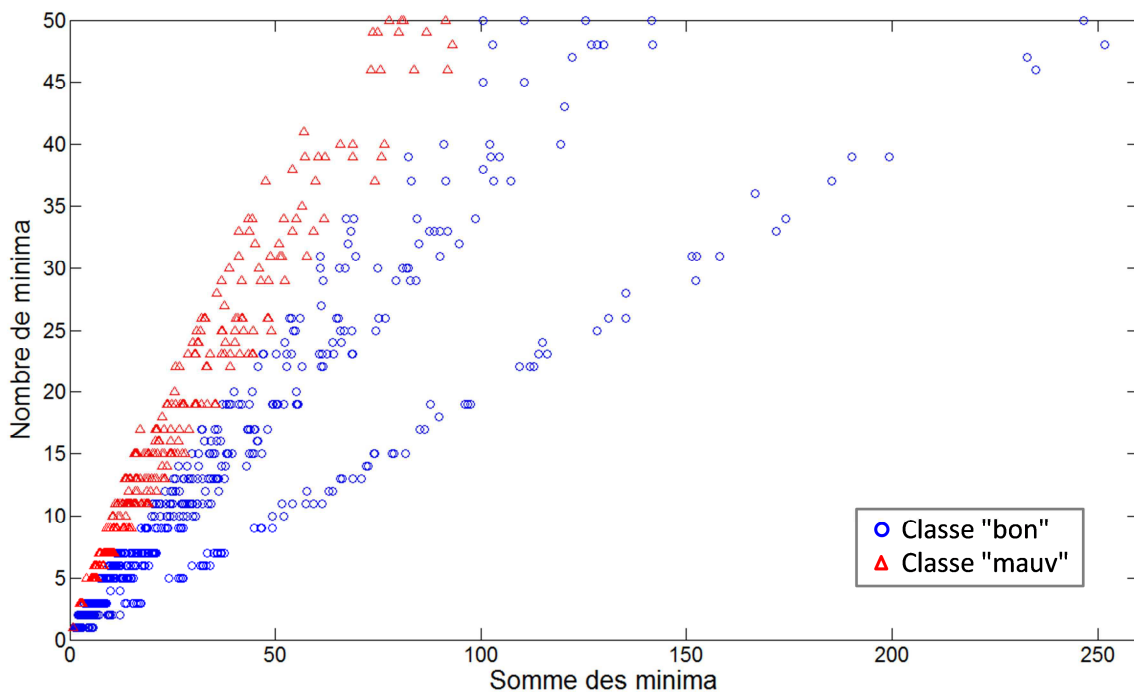


Figure 5.5. Diagramme de dispersion de référence après correction

Nous n'observons plus aucun chevauchement sur ce diagramme. De plus, nous parvenons à un taux de bonne classification de 100% pour les deux classes, ce qui constitue le but recherché.

A présent que nous disposons d'une base de référence entièrement fiable, nous pouvons procéder à des essais sur d'autres signaux que nous supposons de classe inconnue. Afin de considérer des signaux différents des modèles qui ont été employés jusqu'à présent, nous proposons de construire de nouveaux scénarios de test en utilisant des transitoires collectés lors de mesures réalisées à bord d'un train (section II) et d'un métro (section III).

## II - Tests sur des signaux réellement acquis à bord d'un train

### II.1 - Présentation des scénarios créés à partir de signaux transitoires réels

Les signaux transitoires utilisés pour créer nos nouveaux scénarios de test ont été enregistrés lors d'une précédente campagne de mesures à bord d'un TGV circulant sur une ligne non équipée du système de communication GSM-R. Ils ont été collectés à l'aide d'un oscilloscope numérique, sur une durée de 500 ns et avec une fréquence d'échantillonnage de 10 GHz. Parmi tous les transitoires collectés, nous en avons sélectionné deux qui possèdent des caractéristiques temporelles (durée et temps de montée) proches des valeurs typiquement observées [5.1] et dont l'impact sur la qualité d'une communication GSM-R a été étudié auparavant [5.2].

Le principe de construction reste le même : le signal transitoire enregistré est répété avec un intervalle de temps compris entre 20  $\mu$ s et 1 ms, comme illustré par la figure 5.6.

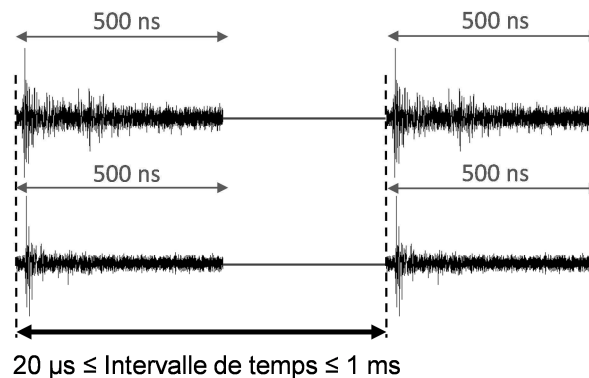


Figure 5.6. Scénarios de test utilisant des signaux transitoires enregistrés in situ

### II.2 - Diagrammes de dispersion et analyse des résultats

Les fichiers de mesures enregistrés lors de nos essais sont répartis équitablement (un sur deux) en deux bases :

- une base de développement,
- une base de test.

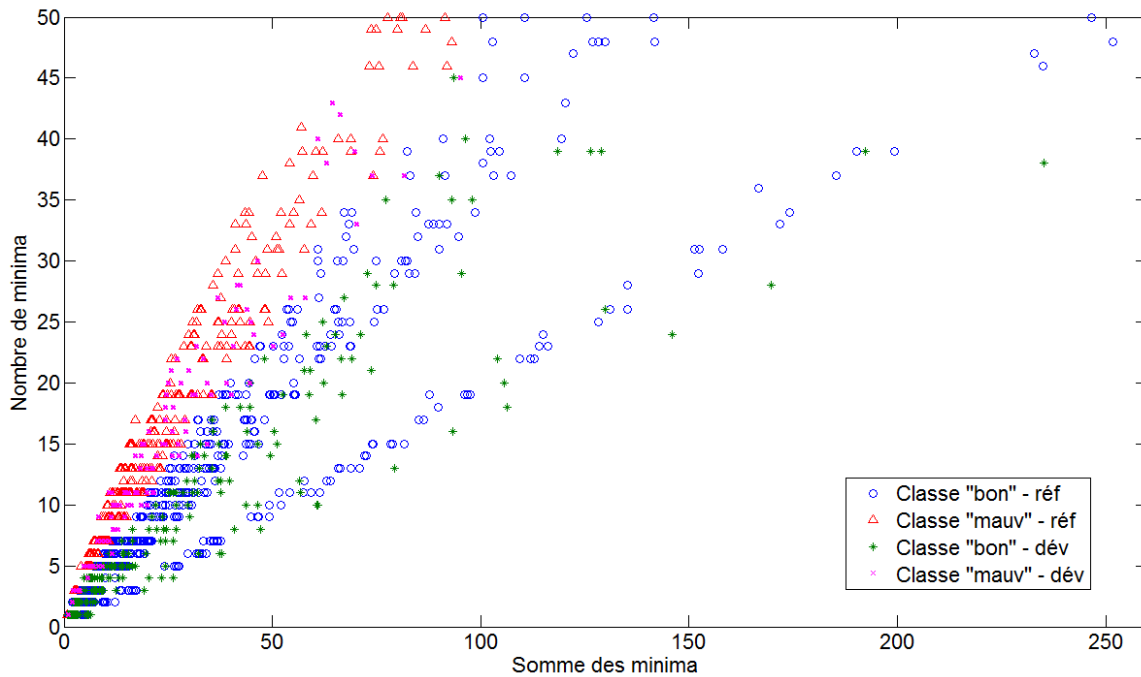
La base de développement sert à compléter la base de référence dont les éléments ont été produits à partir de transitoires générés sur la base des deux modèles précédemment définis. La classe de chaque élément de la base de développement est estimée en appliquant la technique des 3-ppv à partir de la base de référence. Tous les éléments de la base de développement qui sont correctement classés, sont intégrés à la base de référence initiale.

La base de test est quant à elle utilisée pour la dernière évaluation en aveugle avant l'application de notre méthode à bord d'un train. Chaque élément de la base de test est classé à partir de la base de référence enrichie.

Comme nous sommes toujours dans la phase d'apprentissage, nous avons besoin de connaître avec certitude la classe des couples (nombre de minima, somme des minima) extraits des fichiers de mesures qui ont servi à construire ces bases. Ainsi, nous pouvons valider la base de référence, l'enrichir et pour finir, tester le taux de bonne classification que nous obtenons alors avec la base de test. De ce fait, les éléments qui constituent les bases de développement et de test ont été préalablement labellisés en fonction de leur classe d'appartenance lors de la collecte des fichiers de mesures.

*II.2.a. Base de développement*

La base de développement compte 243 éléments de type "bon" et 115 "mauv". Ils sont placés sur le diagramme de référence afin d'observer leur positionnement par rapport aux couples de la base de référence qui y sont déjà positionnés. Sur la figure 5.7, les mêmes couleurs que celles des points de la base de référence sur la figure 5.5 ont été utilisées. Quant aux éléments de la base de développement, ils sont matérialisés en vert pour la classe "bon" et en mauve pour la classe "mauv".



**Figure 5.7. Diagramme de dispersion des éléments des bases de référence et de développement**

Globalement, les éléments de la base de développement viennent correctement se superposer aux éléments de référence en suivant la distribution établie des classes dans l'espace. Toutefois, des chevauchements (pour la plupart, des éléments "mauv" parmi les "bon") se produisent.

Le phénomène peut s'expliquer de la manière suivante. Il convient d'abord de rappeler que les fichiers contenant les signaux transitoires collectés à bord d'un train avaient été enregistrés avec une fréquence d'échantillonnage supérieure à celle utilisée pour collecter les fichiers de mesures avec l'oscilloscope lors de nos tests. Les scénarios créés avec ces signaux transitoires sont donc générés par l'AWG avec une fréquence d'échantillonnage supérieure à celle appliquée pour les collecter à l'oscilloscope en laboratoire. En termes de mesure, cela signifie que nous collectons moins de points de mesure à l'oscilloscope que nous en générons avec l'AWG par fenêtre de 1 ms. Ainsi, la valeur crête d'un signal transitoire peut être manquée lors de l'enregistrement à l'oscilloscope en laboratoire car le point peut ne pas être enregistré en raison du sous-échantillonnage appliqué. Il en résulte une sous-estimation du niveau de la puissance induite par ce transitoire sur la bande GSM-R et donc, une surestimation du descripteur D à l'instant d'apparition du transitoire. De ce fait, le minimum associé est plus élevé qu'il ne devrait l'être et peut se retrouver au-dessus du seuil fixé pour détecter les minima. Si c'est le cas, il n'est pas détecté lors du comptage des minima sur la courbe du descripteur. Par conséquent, le nombre de minima et la somme correspondante sont erronés puisque certains "vrais minima" ne sont pas comptabilisés. Sinon, il est détecté mais reste encore trop élevé, ce qui induit également une erreur (sous la forme d'une surestimation) sur la somme uniquement.

Dans les deux cas, lorsque le point est placé sur le diagramme de dispersion, il se décale et se retrouve dans le sous-espace des paramètres correspondant à la mauvaise classe. L'amplitude du décalage dépend du nombre de minima manquants ou de la surestimation de la somme.

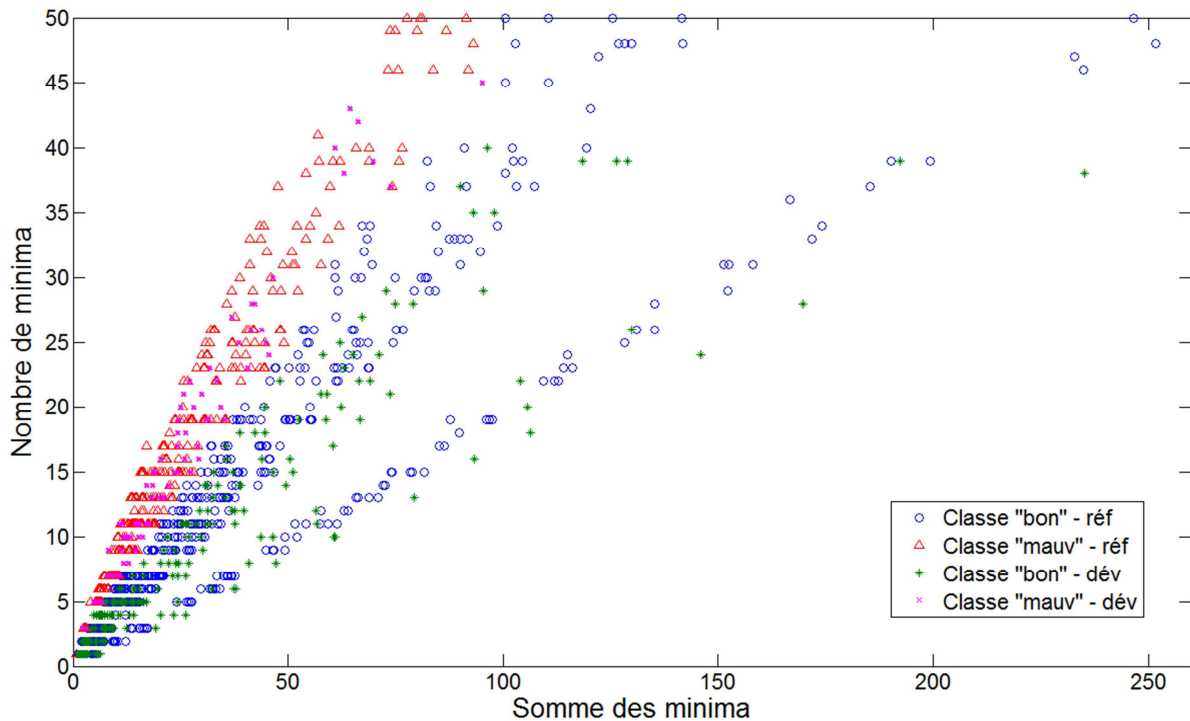
Si ces chevauchements s'expliquent, ils restent néanmoins la source d'erreurs de classification, comme le montre la matrice de confusion de la figure 5.8. Cette matrice est calculée en appliquant la règle des trois plus proches voisins aux éléments de la base de développement vis-à-vis des éléments de la base de référence.

		(Estimé)	
		Bon	Mauv
(Réel)	Bon	238/243 0.979	5/243 0.021
	Mauv	31/115 0.270	84/115 0.730

**Figure 5.8. Matrice de confusion de la base de développement**

Il s'avère en effet que de nombreuses erreurs (31) sont commises sur des éléments de la classe "mauv" qui sont, à tort, classés comme "bon". Or, comme nous pouvons le voir sur le diagramme de dispersion, les problèmes de chevauchement concernent, pour la quasi-totalité, des symboles de la classe "mauv". Il est donc logique que le taux de mauvaise classification des éléments "mauv" soit assez conséquent (27%).

Les couples (nombre de minima, somme des minima) correctement identifiés sont conservés pour enrichir notre base de référence et les autres sont éliminés. Nous obtenons alors le diagramme de dispersion, nommé « diagramme de référence complet » présenté en figure 5.9.



**Figure 5.9. Diagramme de dispersion des éléments des bases de référence et de développement après correction de cette dernière, nommé « diagramme de référence complet »**

*II.2.b. Tracé de la frontière entre les deux classes*

A partir de ces résultats, il nous est possible de dessiner manuellement la frontière entre les deux classes sur le diagramme de référence complet. Dans la mesure du possible, celle-ci est arbitrairement placée à égale distance des groupements de symboles "mauv" et "bon" qu'elle sépare. La figure 5.10 en montre le tracé. En observant la disposition des éléments, on s'aperçoit que la frontière est linéaire par partie, en forme de "marches d'escalier".

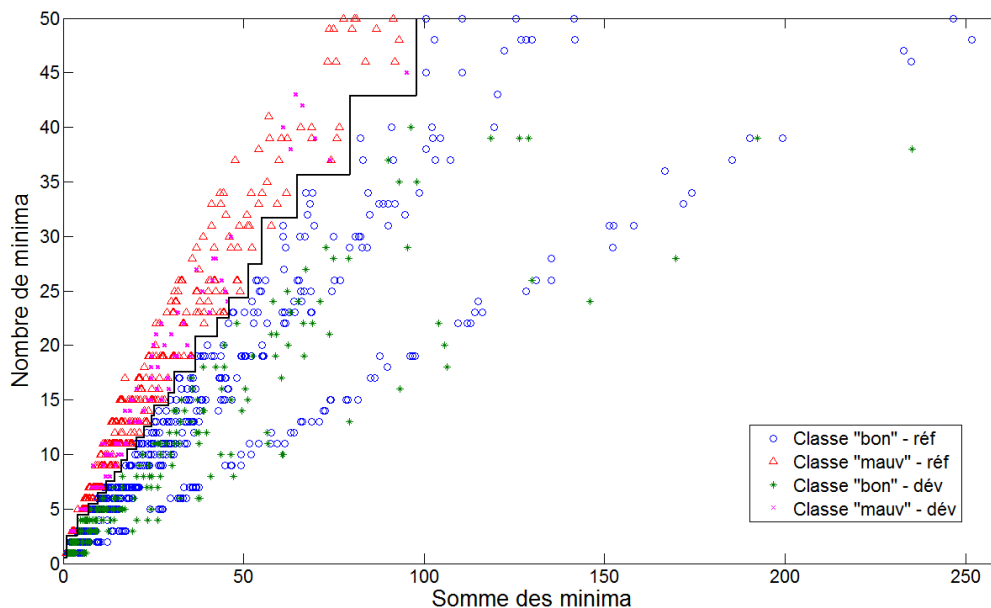
D'ailleurs, ce tracé concorde avec ce que nous avons observé au cours d'études préalables portant sur la susceptibilité du GSM-R face aux signaux transitoires [5.2], [5.3], [5.4]. Ces dernières avaient mis en évidence plusieurs éléments (cf. chapitre 1) :

- une même valeur de BER (donc une même qualité de communication) peut être atteinte pour deux couples (SNR, récurrence) différents,

- pour une valeur de SNR donnée, c'est la récurrence qui, elle seule, influe sur le BER et inversement.

Ici, le nombre de minima est représentatif de la récurrence et la somme, du SNR. Considérons dans un premier temps les segments verticaux de notre frontière pour lesquels le nombre de minima varie et la somme reste constante. Si cette dernière reste constante mais que le nombre de minima évolue, cela signifie que la valeur atteinte par chaque minima, donc le SNR, évolue également. Autrement dit, la récurrence des transitoires et le SNR varient alors que la qualité de la communication, elle, reste la même (pas de basculement entre les deux classes d'un bout à l'autre des segments verticaux) le long de ces segments. Ce qui montre bien qu'un même niveau de qualité peut être atteint pour deux couples (somme des minima ↔ SNR, nombre de minima ↔ récurrence) différents.

Intéressons-nous dans un second temps, aux segments horizontaux pour lesquels le nombre de minima reste constant mais la somme et la qualité de la communication évoluent. Si la somme diffère pour une même valeur de la récurrence, c'est que le SNR varie. Ce qui est en accord avec les courbes présentées au chapitre 1 qui montrent que pour une valeur de récurrence donnée, des mesures réalisées avec des SNR différents conduisent à des BER différents.



**Figure 5.10. Tracé de la frontière entre les deux classes à partir des éléments des bases de référence et de développement (corrigées)**

Le tracé de la frontière permet de découper le diagramme en deux zones, chacune associée à une classe afin de faciliter la visualisation des résultats obtenus avec les éléments de la base de test. Dans le cas présent nous obtenons une zone pour la classe "bon" (bonne qualité de communication) et une pour la classe "mauv" (qualité de communication insuffisante), comme sur la figure 5.11. Dans la section II.2.c, nous montrons comment ce graphe peut être employé. Précisons ici que cette méthode est uniquement proposée dans le but d'effectuer une évaluation subjective de la classification de façon

simple et rapide. Elle ne remplace en aucun cas l'analyse réalisée par le calcul de la matrice de confusion. Le taux de bonne classification doit toujours être calculé en appliquant la règle des trois plus proches voisins, comme cela a été fait pour les deux autres bases et les résultats présentés sous la forme d'une matrice.

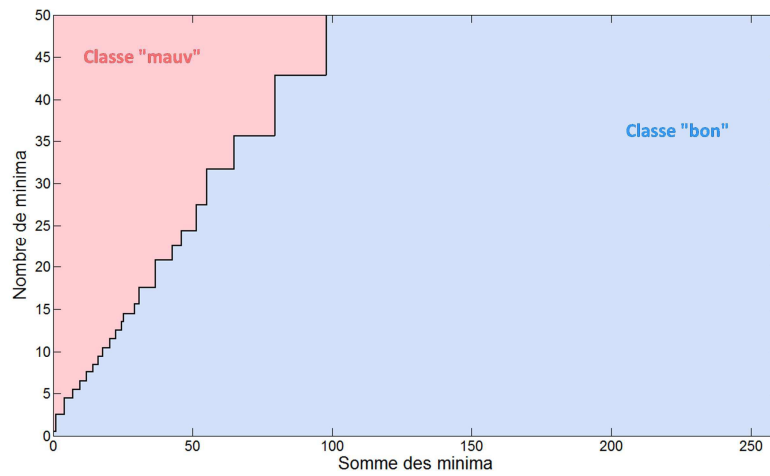


Figure 5.11. Séparation du diagramme de référence en deux zones de classification

*II.2.c. Base de test*

La base de test compte 242 éléments dans la classe "bon" et 114 dans la classe "mauv". Les couples (nombre de minima, somme des minima) sont positionnés sur un diagramme scindé en deux zones, comme le montre la figure 5.12. Les éléments des classes "bon" et "mauv" y sont matérialisés respectivement par des cercles verts et des cercles rouges.

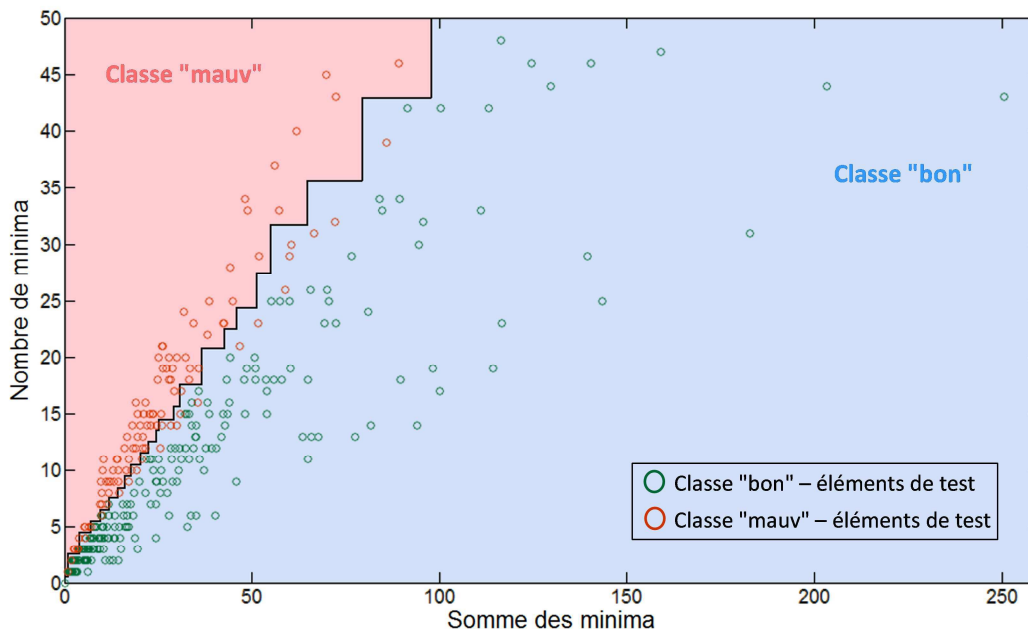


Figure 5.12. Diagramme de dispersion des éléments de test

Visuellement, nous constatons que tous les éléments de test de la classe "bon" sont bien localisés dans la zone bleue. Toutefois, certains éléments de la classe "mauv" le sont également. La matrice de confusion présentée à la figure 5.13 corrobore cette observation : 240 éléments "bon" sur 242 sont correctement classés (ce qui correspond à un taux de réussite de 99.2%) alors que 28 erreurs de classification sont commises sur les 114 couples que compte la classe "mauv" (soit un taux d'erreur de 24.6%).

		(Estimé)	
		Bon	Mauv
(Réel)	Bon	240/242 0.992	2/242 0.008
	Mauv	28/114 0.246	86/114 0.754

Figure 5.13. Matrice de confusion des éléments de test

### II.3 - Bilan et discussion des résultats

Une première base de référence a été créée à partir de différents scénarios construits à l'aide de modèles de transitoires caractéristiques des signaux issus du contact caténaire-pantographe. Après élimination des éléments constituant des sources d'erreur, celle-ci a été enrichie de nouveaux éléments provenant de scénarios similaires construits à partir de signaux acquis en situation réelle à bord d'un train. Ce qui a conduit à l'obtention de notre base de référence complète. Cette base de référence complète comporte l'ensemble des éléments qui ont servi à estimer la classe des éléments d'une base de test, construite à partir de transitoires réels. Le taux de bonne classification est supérieur à 99% pour la classe "bon". Ce taux atteint 75% pour la classe "mauv" et peut donc être amélioré.

#### II.3.a. Augmentation du nombre de voisins

Comme énoncé au paragraphe I.3, l'évaluation présentée précédemment a été obtenue avec la règle des trois plus proches voisins. C'est, en effet, la règle qui nous semblait la plus appropriée dans le cas d'un problème à deux dimensions qui plus est très linéaire.

A titre comparatif, nous avons également appliqué les règles des cinq et sept plus proches voisins pour évaluer si cela permettait d'améliorer ou non les taux de bonne classification. Il s'est avéré que l'augmentation du nombre de voisins conduit à un taux de bonne classification de 100% pour les éléments de la classe "bon". En revanche, quel que soit le nombre de voisins considérés, le taux de bonne classification sur la classe "mauv" diminue légèrement (73.68% au lieu de 75.4%). Cela est dû au fait que les éléments de la base de référence complète appartenant à la classe "bon" sont plus nombreux que ceux de la classe "mauv". Ainsi, les éléments pour lesquels il est plus difficile de trouver plus de trois voisins de la même classe sont plus nombreux parmi les "mauv". De ce fait un élément "mauv" peut être faussement classé comme "bon" car ayant moins de cinq voisins appartenant à la classe "mauv" et cinq (ou plus) parmi les autres plus proches (même si plus éloignés que les quelques voisins "mauv") de la classe "bon".



II.3.b. Compensation de la perte de dynamique liée au sous-échantillonnage lors de la collecte des fichiers à l'oscilloscope

Comme précédemment expliqué dans la section II.2.a, les fichiers de mesures utilisés pour construire les bases de développement et de test ont été sous-échantillonnés lors de leur enregistrement à l'oscilloscope. Ce sous-échantillonnage est la raison d'une perte de dynamique des signaux correspondants. Ainsi les valeurs des minima et leur somme sont surestimées, ce qui provoque un "décalage" des éléments des bases de développement et de test sur le diagramme de dispersion de référence.

Pour compenser cette perte de dynamique et ainsi corriger les valeurs des sommes des minima, nous proposons d'appliquer un gain  $\epsilon$  à la somme des minima pour chaque couple (nombre de minima, somme des minima). Le tableau de la figure 5.14 montre les résultats de classification obtenus en appliquant la règle des trois plus proches voisins aux éléments de la base de développement pour différentes valeur de  $\epsilon$ .

	BASE DE DEVELOPPEMENT			
	"Bon" correctement identifiés	"Mauv" correctement identifiés	Erreurs "bon" en "mauv"	Erreurs "mauv" en "bon"
Résultats initiaux	238/243 0.979	84/115 0.73	5/243 0.021	31/115 0.27
$\epsilon = 0.9$	234/243 0.963	102/115 0.887	9/243 0.037	13/115 0.113
$\epsilon = 0.85$	228/243 0.938	106/115 0.922	15/243 0.062	9/115 0.078
$\epsilon = 0.8$	216/243 0.889	110/115 0.957	27/243 0.111	5/115 0.043
$\epsilon = 0.7$	188/243 0.774	113/115 0.983	55/243 0.226	2/115 0.017

Figure 5.14. Compensation de la perte de dynamique sur la base de développement

Nous constatons que c'est en appliquant un coefficient multiplicateur  $\epsilon=0.85$  que nous obtenons les meilleures performances globales. En effet, avec cette valeur, nous diminuons nettement le taux d'erreur de la classe "mauv", tout en conservant un bon taux de bonne classification pour la classe "bon". Nous conservons donc cette valeur de  $\epsilon=0.85$  et l'appliquons aux éléments de la base de test. Les résultats ainsi obtenus sont présentés sur la figure 5.15.

	BASE DE TEST			
	"Bon" correctement identifiés	"Mauv" correctement identifiés	Erreurs "bon" en "mauv"	Erreurs "mauv" en "bon"
Résultats initiaux	240/242 0.992	86/114 0.754	2/242 0.008	28/114 0.246
$\epsilon = 0.85$	217/242 0.897	106/114 0.93	25/242 0.103	8/114 0.07

Figure 5.15. Compensation de la perte de dynamique sur la base de test

A nouveau, nous observons une augmentation du taux de bonne classification de la classe "mauv" qui passe de 75.4% à 93%. Dans le même temps celui de la classe "bon" diminue, un peu plus fortement que dans le cas de la base de développement puisqu'il passe de 98.8% à 89.7%.

Ce résultat est acceptable pour deux raisons :

- un taux de bonne classification de 90% est très satisfaisant,
- le taux de bonne classification des éléments de la classe "mauv" est supérieur à 90%. Pour l'application à bord des trains, il nous semble plus utile d'être capable d'identifier au mieux les situations critiques pour le GSM-R. D'où la nécessité d'avoir un taux de bonne classification élevé sur les éléments de la classe "mauv". Ainsi, en favorisant légèrement la reconnaissance des éléments "mauv", on anticipe sur la façon dont notre méthode de classification pourrait être appliquée à bord des trains et sur ce qu'elle pourrait apporter en termes de détection de situations critiques pour la qualité des transmissions GSM-R.

Utiliser un gain est une solution à l'utilisation de différentes fréquences d'échantillonnage. A l'avenir, cette différence devra être proscrite car la compensation ne sera pas réalisable sur les fichiers "inconnus". Les fichiers devront tous être collectés à l'oscilloscope avec une fréquence de 2.5 GHz de manière à :

- se voir appliquer le même post-traitement que celui appliqué pour construire la base de référence (largeur de la fenêtre de Gabor, facteur de recouvrement...),
- fournir des couples (nombre de minima, somme des minima) qui peuvent être placés (sans qu'il y ait de "décalage") sur le diagramme de référence.

### III - Tests sur des signaux acquis dans le métro de Barcelone

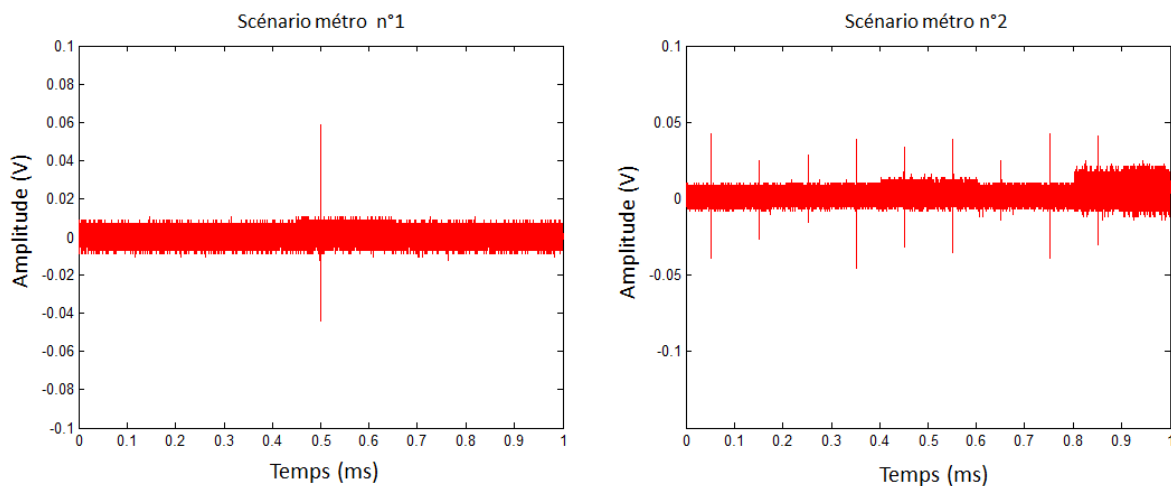
Jusqu'à maintenant, nous avons effectué des tests sur des signaux contenant des perturbations EM transitoires couvrant les bandes de fonctionnement du GSM-R. Une question se pose alors : comment réagit la méthode face à des perturbations EM qui couvrent d'autres bandes que celles du GSM-R ? Pour y répondre, nous utilisons des enregistrements de transitoires couvrant des fréquences plus basses que celles du GSM-R pour construire de nouveaux scénarios de test. Le but est de vérifier que le détecteur défini précédemment n'engendre aucune fausse détection.

#### III.1 - Présentation des signaux de test utilisés

Cette nouvelle base de test regroupe des mesures effectuées dans le métro de Barcelone dont l'alimentation est assurée par une caténaire. Le but de ces mesures était d'étudier les signaux transitoires issus du contact caténaire-pantographe en environnement tunnel. Les antennes étaient fixées à l'avant de la rame de métro sous le pantographe et connectées à un oscilloscope qui réalisait les acquisitions sur des fenêtres de 1 ms avec une fréquence d'échantillonnage de 10 GHz.

Parmi tous les enregistrements réalisés lors de cette campagne, nous en sélectionnons deux qui présentent des caractéristiques temporelles différentes afin de diversifier les scénarios de test. Le premier contient un seul transitoire dont l'amplitude pic est nettement supérieure au niveau de bruit. Le second est quant à lui constitué d'une série de transitoires régulièrement espacés et dont l'amplitude varie et reste supérieure au niveau de bruit.

Ces enregistrements sont rejoués à l'aide de l'AWG et mélangés à un signal GSM-R via notre banc de test. Le mélange « transitoire + signal de communication » ainsi créé, qui est reçu par le mobile GSM-R, est enregistré à l'oscilloscope sur une durée de 1 ms et avec une fréquence d'échantillonnage de 2.5 GHz. La figure 5.16 montre les deux enregistrements réalisés à l'entrée du mobile.



**Figure 5.16. Scénarios construits à partir de signaux transitoires collectés à bord du métro (amplitudes à l'entrée du mobile GSM-R)**

#### III.2 - Analyse de la couverture spectrale des signaux de test utilisés

La couverture spectrale de nos deux enregistrements est étudiée à l'aide de l'application TIFANY, développée au cours de son stage par Benjamin Ruytoor. Cette application permet le calcul de transformations temps-fréquence sur des données volumineuses en un temps réduit [5.5]. Nous

employons une TFCT avec une fenêtre (rectangulaire) de 2048 points et un facteur de recouvrement de 50%, soit des paramètres d'entrée identiques à ceux utilisés pour notre transformation de Gabor optimisée. Les représentations temps-fréquence obtenues avec TIFANY sont fournies par les figures 5.17 et 5.18.

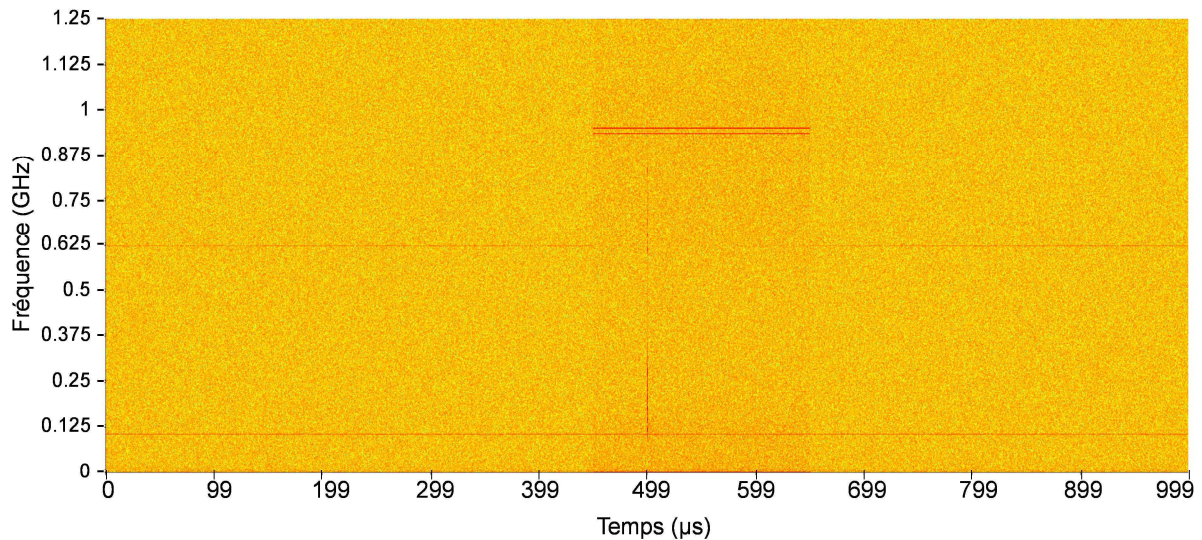


Figure 5.17. Couverture spectrale du scénario n°1

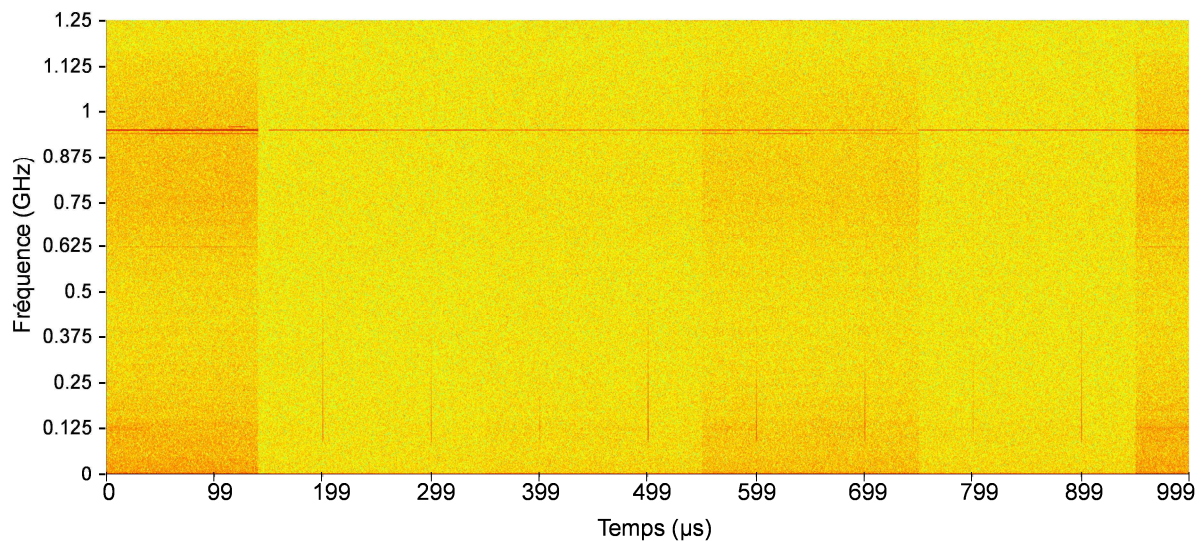
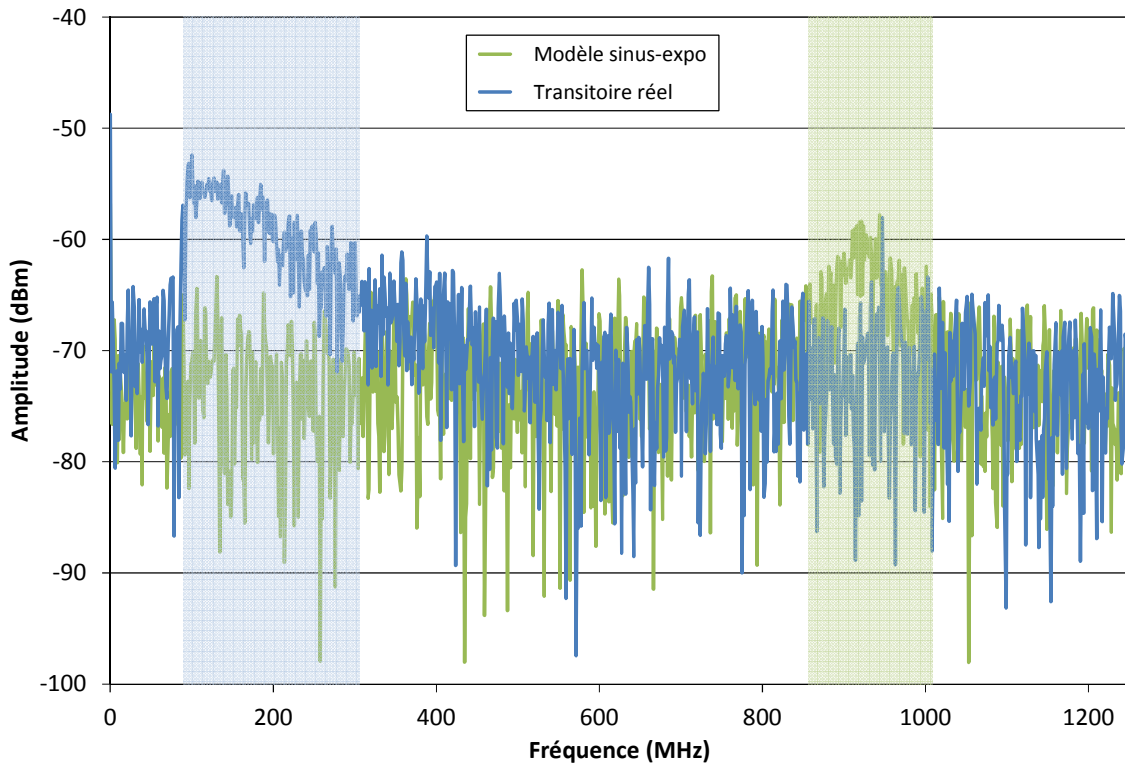


Figure 5.18. Couverture spectrale du scénario n°2

Les deux graphes montrent que les transitoires contenus dans ces scénarios ne couvrent pas (n°2) ou très peu (n°1) les bandes de fonctionnement du GSM-R mais des fréquences plus basses qui ne font pas partie de notre bande d'étude.

Cette observation est confirmée par l'étude du spectre de l'un des transitoires du scénario n°2. Le spectre du transitoire qui se produit aux environs de  $t=499 \mu s$  est représenté sur la figure 5.19. A titre de comparaison, la représentation spectrale du modèle sinus-expo y est superposée. Sur cette figure, le spectre du signal transitoire acquis dans le métro est en bleu et celui du modèle en vert. Nous employons le même code-couleur pour faire apparaître les bandes de couverture en fréquences respectives des deux signaux. Ainsi que précédemment observé, la puissance du transitoire du scénario

n°2 est concentrée en basses fréquences (100-300 MHz) alors que pour le modèle sinus-expo elle est bien centrée sur la bande du GSM-R.



**Figure 5.19. Comparaison des spectres du modèle sinus-expo et du transitoire issu du scénario n°2**

Dès lors, les signaux transitoires contenus dans ces nouveaux scénarios de test ne peuvent pas impacter la qualité des communications GSM-R. De ce fait, le classifieur ne devrait pas leur attribuer de classe.

Comme expliqué au chapitre 4, pour calculer le descripteur il nous faut déterminer la puissance générée par les perturbations transitoires sur le canal GSM-R. Celle-ci est estimée par la valeur maximale sur la bande 921-925 MHz à l'exception du canal GSM-R employé par la communication. Or ici, le niveau de puissance ramené sur la bande 921-925 MHz n'est pas représentatif des transitoires contenus dans ces fichiers de mesures car ceux-ci couvrent des fréquences plus basses. Ainsi, les valeurs maximales extraites sur la bande du GSM-R ne devraient pas permettre de faire apparaître des minima, sur la courbe du descripteur, résultant effectivement de l'apparition des transitoires. Pour le vérifier, dans la section III.3, nous calculons le descripteur de ces deux fichiers de mesures.

### *III.3 - Courbes obtenues pour le descripteur*

La figure 5.20 présente la courbe du descripteur obtenue en présence d'un unique signal transitoire couvrant les bandes du GSM-R. Dans le cas de notre scénario n°1, nous obtenons la courbe de la figure 5.21. Nous constatons une différence importante dans la typologie de ces courbes. En effet, sur la figure 5.20 nous observons une succession de fluctuations autour d'une valeur moyenne et un minimum qui se dégage nettement de ces fluctuations, marquant ainsi l'apparition d'un signal transitoire. Ce n'est pas le cas de la courbe du descripteur présentée en figure 5.21. Ceci nous empêche



de définir un seuil visuel qui permet de calculer le nombre de minima et la somme correspondante. Par conséquent, aucun couple (nombre de minima, somme des minima) ne peut être extrait de cette courbe de descripteur. Nous ne pouvons donc pas appliquer notre méthode de classification au scénario de test n°1.

Une solution réside dans la création d'une classe de rejet correspondant aux scénarios dans lesquels aucun phénomène transitoire n'apparaît dans les bandes de fréquences du GSM-R. Ce qui correspond aux cas où le couple (nombre de minima, somme des minima) vaut (0,0). Le scénario n°1 sera ainsi attribué à cette classe de rejet.

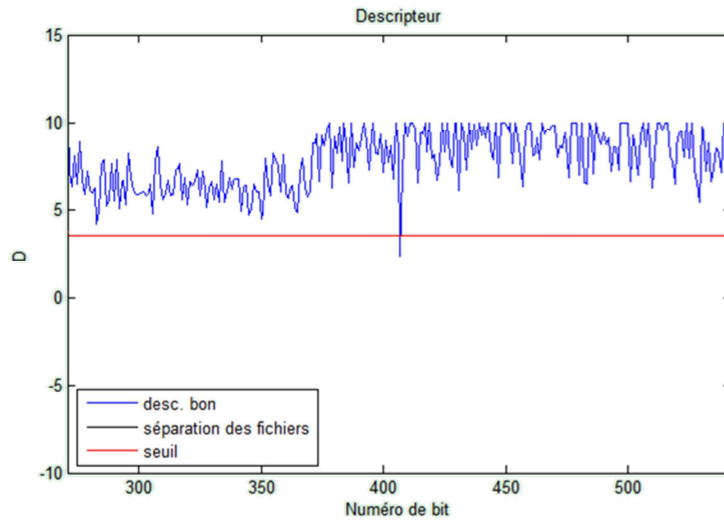


Figure 5.20. Exemple de courbe du descripteur obtenue en présence d'un unique événement transitoire couvrant les bandes du GSM-R

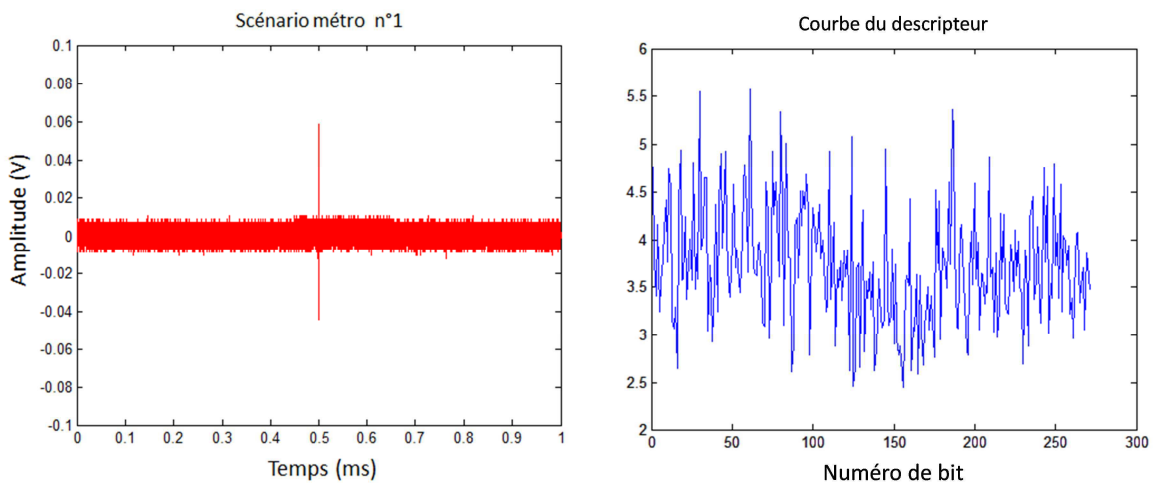
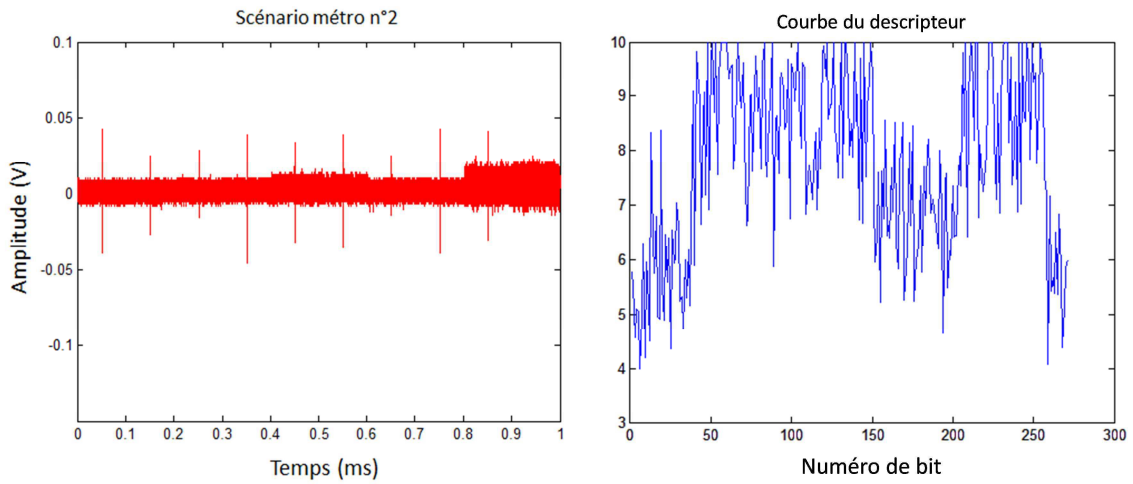


Figure 5.21. Courbe du descripteur obtenue pour le scénario n°1

La figure 5.22, qui fournit la courbe du descripteur obtenue pour le scénario n°2, est une autre illustration du phénomène décrit précédemment. Le scénario n°2 sera lui aussi attribué à la classe de rejet.



**Figure 5.22. Courbe du descripteur obtenue pour le scénario n°2**

Dans les sections précédentes, nous avons montré que notre méthode permet de classer, en une classe "bon" ou une classe "mauv", les scénarios de bruit EM contenant des transitoires qui couvrent les bandes du GSM-R.

Avec ce nouvel exemple nous montrons que le descripteur et les deux paramètres qui en sont extraits permettent d'écarter les scénarios dans lesquels aucun transitoire n'apparaît dans les bandes de fréquences du GSM-R. Une classe de rejet peut ainsi être créée pour éviter la production de fausses détections.

## IV - Vers l'application automatique d'un seuil pour détecter et compter les minima

Comme nous l'avons expliqué au chapitre 4, pour le moment nous choisissons de définir manuellement le seuil à appliquer, pour compter les minima, en visualisant la courbe du descripteur. Sur la base des quelques retours d'expérience et d'utilisation de notre méthode dont nous disposons, nous définissons, dans cette section, une méthode de seuillage entièrement automatique. Ce seuil est défini à partir d'informations extraites de la courbe du descripteur.

Il s'agit de calculer "l'erreur" commise sur l'estimation de la puissance du signal GSM-R en l'absence de transitoire. Cette erreur, notée  $\Delta$ , est égale à

$$\Delta = \text{maximum} - \text{moyenne} \quad 5.1$$

où *maximum* correspond à la valeur maximale du descripteur (nécessairement atteinte en l'absence de transitoire) et *moyenne* représente la valeur moyenne du descripteur (qui, compte tenu de la durée très brève d'un transitoire, équivaut à la valeur moyenne des fluctuations de la puissance du signal GSM-R en l'absence de transitoire).

Cette erreur fournit la marge à appliquer de part et d'autre de la valeur moyenne du descripteur pour identifier les minima représentatifs de l'apparition effective de signaux transitoires. Le nouveau seuil est alors défini par la relation suivante :

$$\text{nouveau seuil} = \text{moyenne} - \Delta \quad 5.2$$

Sur la figure 5.23, nous représentons les différentes informations extraites de la courbe du descripteur qui conduisent à la définition de ce nouveau seuil (en rouge). A titre de comparaison, le seuil initialement défini (comme un pourcentage de la valeur moyenne du descripteur) et appliqué pour cette courbe de descripteur y est également tracé (en vert).

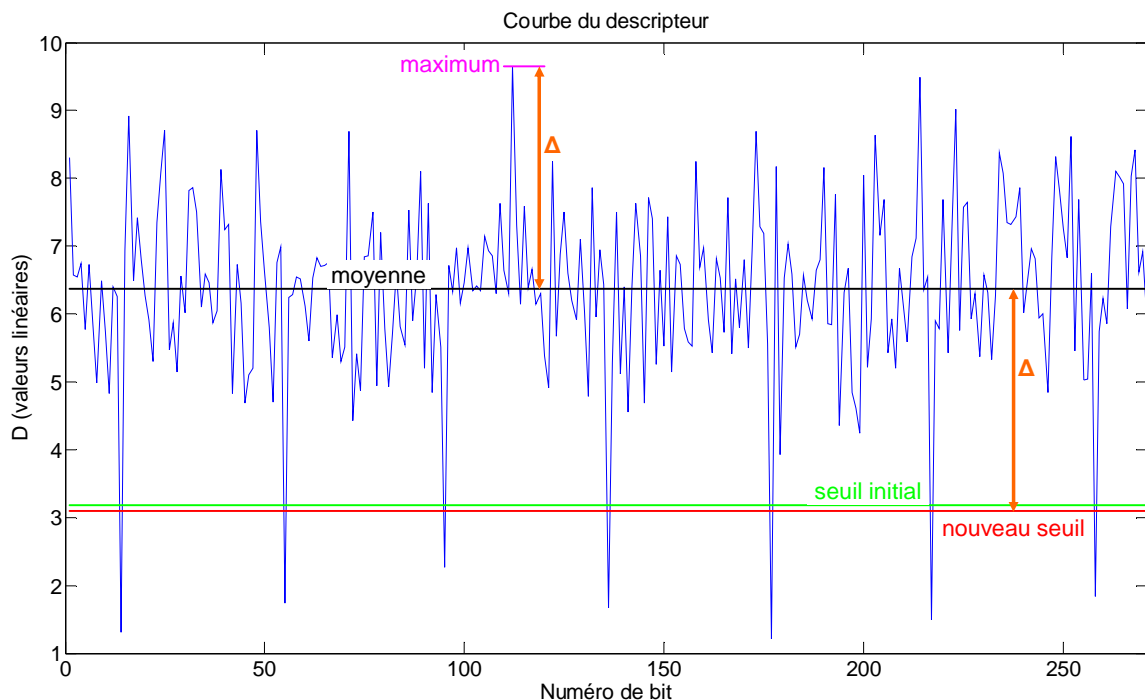


Figure 5.23. Définition du nouveau seuil



Nous constatons que le nouveau seuil est très proche de celui qui avait été défini initialement pour détecter et compter les minima lors du précédent traitement. De ce fait, avec ce nouveau seuil, nous devrions parvenir à des résultats équivalents aux précédents au niveau des couples (nombre de minima, somme des minima) extraits des courbes de descripteur. Pour le vérifier, nous sélectionnons arbitrairement quelques courbes de descripteur sur lesquelles nous appliquons le nouveau seuil pour détecter les minima. Nous comparons, dans le tableau de la figure 5.24, les résultats ainsi obtenus avec les résultats précédemment obtenus avec le seuil manuel.

Fichier n°	Seuil initial		Nouveau seuil	
	Nb	Somme	Nb	Somme
1	5	7.78	5	7.78
2	5	8.26	5	8.26
3	7	11.53	7	11.53
4	3	4.55	3	4.55
5	3	3.65	3	3.65
6	15	24.32	15	24.32
7	25	42.22	25	42.22
8	11	14.04	11	14.04
9	13	16.06	12	14.35
10	1	0.76	1	0.76

Figure 5.24. Comparaison des paramètres extraits (nombre de minima, somme des minima) avec application du seuil initial et du nouveau seuil

Sur les 10 fichiers testés, il n'existe qu'un cas où le nouveau seuil fournit des résultats différents. En effet, un minimum semble ne pas avoir été compté sur le fichier n°9 avec le nouveau seuil. La figure 5.25 illustre ce phénomène.

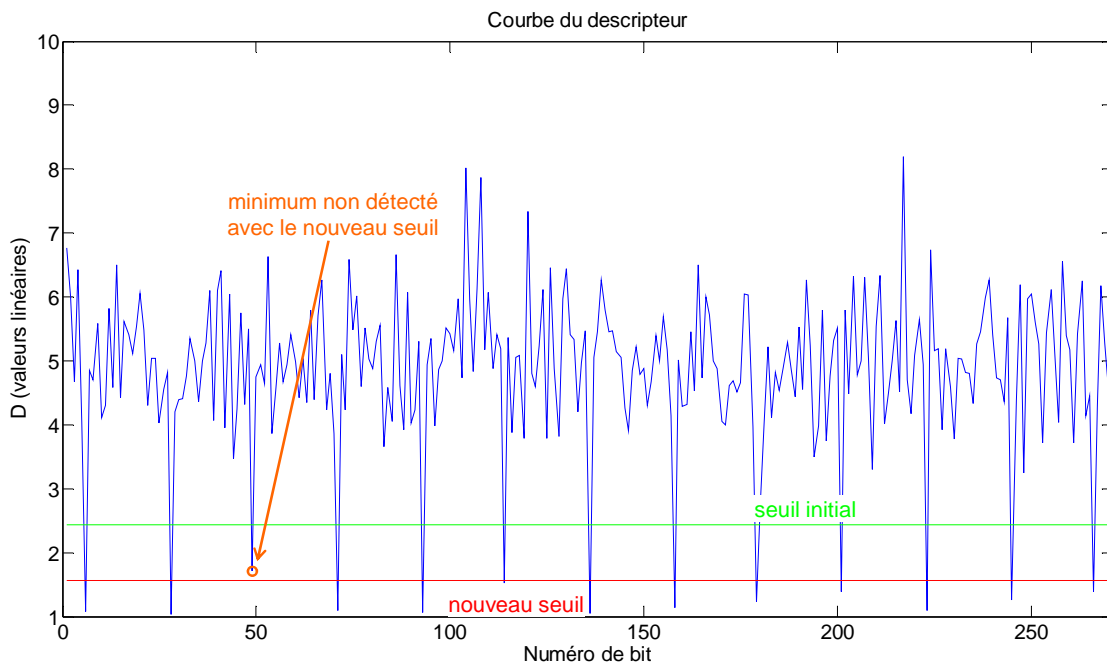


Figure 5.25. Cas de non-détection d'un minimum avec le nouveau seuil dans le cas du fichier n°9

Il apparaît qu'un minimum se trouve au-dessus du nouveau seuil et, par conséquent n'est pas détecté. Le seuil initial, qui lui avait été fixé manuellement en visualisant la courbe du descripteur, permet de détecter tous les minima présents sur la courbe.

Dès lors, pour valider complètement le nouveau seuil, il faudrait procéder à un grand nombre d'essais afin de s'assurer qu'il permet de compter correctement les minima et que les cas de non-détection sont très rares. Dans l'immédiat nous ne pouvons pas le faire car les courbes de descripteur dont nous disposons ont été construites en imposant une limite à 10 sur l'axe des ordonnées puisqu'au départ nous ne nous intéressions pas aux trop fortes valeurs de  $D$ . La grande majorité des courbes est donc écrêtée, ce qui ne posait pas de problème pour le traitement initial mais qui rend impossible la détermination du véritable maximum, valeur indispensable à la détermination du nouveau seuil.

Il s'agit donc, pour le moment, uniquement d'une piste d'amélioration possible de notre méthode qui devra être testée à plus grande échelle pour en démontrer la pertinence.

## Conclusion

A l'aide des quatre scénarios construits à partir de modèles de transitoires, une base de référence a pu être établie. Cette base a été analysée, corrigée et enrichie de nouveaux éléments provenant d'essais avec des signaux transitoires collectés à bord d'un train. A l'issue de ces manipulations, nous avons obtenu une base de référence complète pour laquelle le taux de bonne classification des deux classes est de 100%. Afin de valider notre méthode, cette base a été utilisée pour procéder à la classification de fichiers de test. Ces fichiers provenaient de deux séries de tests réalisées avec des signaux transitoires collectés à bord d'un train, pour l'une et d'un métro en environnement tunnel pour l'autre. Les éléments de la première série ont pu être classés avec notre méthode, avec un taux de bonne classification de plus de 99% pour la classe "bon" et de plus de 75% pour la classe "mauv". Rappelons que de nombreuses erreurs de classification proviennent du sous-échantillonnage lors de la mesure à l'oscilloscope et qu'après compensation de la perte de dynamique qui en résulte, nous parvenons à un taux de bonne classification de 89.7% pour la classe "bon" et de 93% pour la classe "mauv". Afin d'éviter les problèmes liés au sous-échantillonnage et d'optimiser les performances de notre méthode, nous devons toujours être précautionneux quant à la qualité des mesures.

Les fichiers de la deuxième série ont confirmé la robustesse et la pertinence de la méthode de classification puisque les transitoires qu'ils contenaient n'ont, à juste titre, pas été classés. En effet, étant donné qu'ils ne couvraient pas la bande GSM-R, ils ne devaient pas être pris en compte car n'ayant pas d'impact sur la qualité des communications. La méthode de classification a ainsi prouvé sa fiabilité.

Pour finir, nous avons également présenté une nouvelle définition du seuil à appliquer pour détecter et compter les minima sur la courbe du descripteur. Sa détermination est entièrement automatique et repose sur l'extraction, de la courbe du descripteur, de deux informations : valeur maximale et valeur moyenne. Toutefois, des expérimentations supplémentaires sont nécessaires pour valider complètement cette approche qui a, pour le moment, été testée sur 10 fichiers seulement.



# CONCLUSION GENERALE ET PERSPECTIVES

Les travaux présentés dans cette thèse visent à contribuer à l'évolution nécessaire des équipements et des méthodologies définies dans les normes CEM afin de s'adapter aux nouveaux systèmes de communication sans fil employés dans le monde des transports. Nous avons choisi le domaine ferroviaire comme cadre applicatif et le système de transmission numérique (GSM-R), progressivement installé à bord des TGV sur l'ensemble du territoire européen, comme objet d'étude. Nous avons limité notre étude à un seul type de perturbation EM, à savoir les signaux transitoires issus du contact glissant entre le fil caténaire et le pantographe.

La méthodologie proposée permet la classification des bruits EM transitoires en fonction de leur impact sur la qualité des communications GSM-R. Cependant, elle s'avère tout à fait transposable à d'autres systèmes de transmission numérique vis-à-vis de la présence de signaux EM transitoires.

Elle nécessite pour cela, en amont, une étude de la susceptibilité du système de communication à protéger vis-à-vis de la présence de signaux EM transitoires. En effet, cette étude permet d'identifier les facteurs les plus influents sur le système et dont le descripteur doit rendre compte. Dans le cadre de cette thèse, celle-ci avait été réalisée quelques mois auparavant au laboratoire.

Le développement de notre méthode s'est déroulé en plusieurs phases que l'on peut résumer comme suit.

La première étape a consisté en la mise en place d'un banc de tests permettant de reproduire, en laboratoire, les conditions EM dans lesquelles la radio sol-trains GSM-R est amenée à fonctionner. Ce banc permet également d'effectuer l'ensemble des mesures temporelles et fréquentielles nécessaires au développement de notre méthodologie et de procéder à la classification des éléments utilisés pour la construction de notre base de référence.

Dans une deuxième étape, nous avons développé le post-traitement à appliquer aux mesures temporelles afin de réaliser la classification des bruits EM transitoires en fonction de leur impact sur les transmissions GSM-R. Cela a nécessité la définition d'un descripteur, dans notre cas un paramètre (estimé à partir des mesures temporelles) représentatif de l'impact des signaux EM transitoires sur la qualité des communications GSM-R. Ce descripteur est calculé à partir des coefficients issus d'une transformée temps-fréquence que nous avons optimisée de manière à en limiter les calculs à la bande de fréquences du lien descendant GSM-R (921-925 MHz). Une fois la courbe du descripteur tracée, un seuil linéaire est appliqué pour déterminer et compter le nombre de minima présents sous le seuil ainsi que leur somme. Ce couple (nombre de minima, somme des minima) est ensuite placé sur un diagramme de dispersion afin d'observer dans quelle zone de l'espace ainsi défini il se situe. Les premiers essais ont montré que nous parvenions à une séparation satisfaisante des deux classes. Ceci valide le fait que notre descripteur et les deux paramètres qui en sont extraits sont pertinents pour la caractérisation de la qualité des signaux GSM-R.

Dans un troisième temps, une base de référence composée de couples (nombre de minima, somme des minima) de classe connue a été construite à partir de fichiers de mesures collectés en présence de divers scénarios de bruit EM transitoire. Cette base sera le seul élément sur lequel reposera la classification des futurs éléments de classe inconnue. Elle est représentative des différents types de signaux transitoires qui peuvent être rencontrés en environnement ferroviaire et pourra être complétée au fur et à mesure des expérimentations (et à terme, des utilisations) à bord des trains.

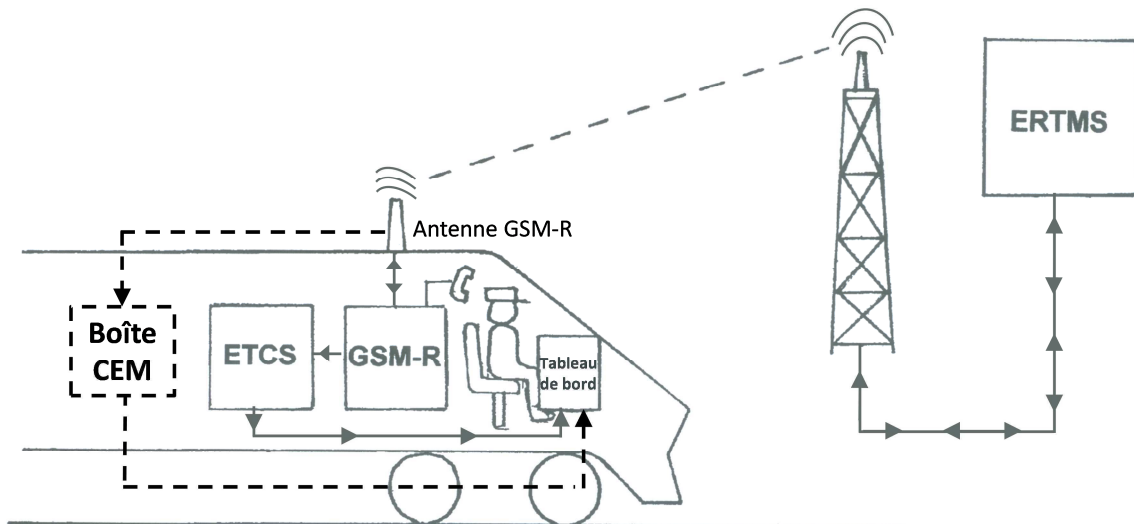
La dernière phase a été consacrée à la réalisation d'essais de classification sur des éléments provenant de mesures en présence de signaux transitoires collectés in situ afin de valider notre base de référence

et par la même notre méthodologie. Au cours de ces essais, notre méthode de classification a démontré sa pertinence et sa robustesse.

Au final, notre méthodologie pourrait trouver son application à bord des trains à grande vitesse européens équipés du système GSM-R où elle pourrait servir à identifier les zones critiques pour la qualité des communications et alerter des risques de dégradation de la qualité. A bord d'un train, elle pourrait prendre la forme d'une "boîte CEM", telle qu'illustrée sur la figure ci-dessous. Cette boîte contiendrait, par exemple, une carte FPGA (Field-Programmable Gate Array) sur laquelle seraient implémentés l'ensemble des algorithmes de traitement. Elle serait connectée à l'antenne GSM-R, fixée en toiture, afin de collecter les signaux reçus par cette dernière durant le trajet et les traiter en temps-réel. Ce système devrait opérer en parallèle et indépendamment du système GSM-R, sans en perturber le fonctionnement ni créer de perturbations supplémentaires pour celui-ci.

Les résultats, quant à eux, pourraient être enregistrés pour être transmis par la suite à l'opérateur ferroviaire ou s'afficher sur le tableau de bord du conducteur sous la forme d'alertes pour l'avertir en cas de risque de perte de disponibilité du système GSM-R.

Durant la phase de déploiement et de test ou d'opérations de maintenance du GSM-R, cet outil pourrait également servir à localiser les zones critiques dans lesquelles la qualité serait insuffisante pour assurer une bonne qualité de service.



Possible implémentation de notre travail à bord d'un train

De plus, cette méthode s'avérerait d'autant plus efficace qu'elle emploie l'antenne réellement utilisée par le système à protéger. Par conséquent, la mesure temporelle est de meilleure qualité puisque l'on rejette les basses fréquences qui sont responsables des problèmes de dégradation de la résolution verticale. Cela permet également de caractériser précisément les signaux qui parviennent à l'entrée du système.

Enfin, on pourrait imaginer implémenter l'algorithme de calcul du descripteur directement sur un équipement de mesure existant, comme par exemple, un analyseur temps-réel. Il rejoindrait alors la liste des traitements déjà proposés par l'équipement qui réalise d'ores et déjà des calculs de FFT sur des données en temporel pour obtenir des spectres en temps-réel. Pour les équipements de mesure existants, cet ajout constituerait une évolution qui, à l'heure actuelle et au vu de la multiplication des systèmes numériques installés à bord des véhicules, semble de plus en plus nécessaire.

Les pistes d'amélioration de cette méthode de classification sont multiples. Considérons, tout d'abord, le seuil de détection des minima. Au chapitre 4, une interface de calcul et de visualisation a été développée afin de procurer une application aisée de la méthode. Elle permet également de définir, manuellement, le seuil à appliquer pour détecter et compter les minima sur les courbes du descripteur. A la fin du chapitre 5, nous avons proposé la définition d'un nouveau seuil qui pourrait être utilisé pour extraire les paramètres "nombre de minima" et "somme des minima" des courbes du descripteur, sans nécessiter de visualisation préalable de ces dernières. Les premiers tests effectués avec ce nouveau seuil se sont révélés encourageants. Toutefois, de nombreux essais devront être réalisés pour valider complètement cette nouvelle approche entièrement automatique. En attendant que cette piste d'amélioration ait été pleinement explorée et validée, nous préconisons de continuer de procéder à la détermination du seuil en visualisant le descripteur.

Le second point d'amélioration concerne les paramètres extraits de la courbe du descripteur qui sont, pour le moment, au nombre de deux. En effet, si on augmentait le nombre de paramètres, on se placerait dans un espace de classification de dimension supérieure, qui pourrait permettre une meilleure séparation des classes. Cela pourrait contribuer à résoudre les problèmes de séparation rencontrés aux abords de la frontière entre les classes. Dans un premier temps, trouver un troisième paramètre pourrait, par exemple, consister à calculer la valeur moyenne des minima ou encore à utiliser la valeur minimale.

En fonction du nombre de paramètres choisis pour construire l'espace dans lequel doit être résolu le problème de classification, il serait possible (et peut-être optimal) d'avoir recours à l'une des autres méthodes de classification présentées dans le chapitre 2. En effet, pour des espaces de dimension supérieure, la règle des  $k$  plus proches voisins ne présente pas autant d'intérêt qu'en dimension 2 où elle est très simple à mettre en œuvre en comparaison avec les autres méthodes de classification.





---

# GLOSSAIRE

AWG	Arbitrary Waveform Generator (générateur de signaux arbitraires)
BER	Bit Error Rate (taux d'erreur bit)
BTS	Base Transceiver Station (station de base)
CEM	Compatibilité ElectroMagnétique
CWD	Choï-Williams Distribution (distribution de Choï-Williams)
CWT	Continuous Wavelet Transform (transformée en ondelettes continue)
DGT	Discrete Gabor Transform (transformée de Gabor discrète)
DWT	Discrete Wavelet Transform (transformée en ondelettes discrète)
EM	ElectroMagnétique
ERTMS	European Rail Traffic Management System
ETCS	European Train Control System
FFT	Fast Fourier Transform
FPGA	Field-Programmable Gate Array
GMM	Gaussian Mixture Model (modèle de mélange gaussien)
GMSK	Gaussian Minimum Shift Keying (modulation à déplacement minimal avec filtrage gaussien)
GSM-R	Global System for Mobile communications - Railways
k-ppv	k plus proches voisins
RBW	Resolution Band Width (bande de résolution)
RTF	Représentation Temps-Fréquence
SNR	Signal-to-Noise Ratio (rapport signal à bruit)
SPEC	Spectrogramme
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine (séparateur à vaste marge)
TDMA	Time Division Multiple Access (AMRT : Accès Multiple à Répartition dans le Temps)
TF	Temps-Fréquence
TFCT	Transformée de Fourier à Court Terme
TI	Time Interval (intervalle de temps)
VBS	Voice Broadcast Services (appels ou annonces diffusés)
VGCS	Voice Group Call Services (appels de groupe)
WD	Wigner-Ville Distribution (distribution de Wigner-Ville)
WPT	Wavelet Packet Transform (décomposition en paquets d'ondelettes)
WT	Wavelet Transform (transformée en ondelettes)



# LISTE DES FIGURES

Figure 1.1. Illustration d'un problème de réception.....	14
Figure 1.2. Gabarit d'émission .....	17
Figure 1.3. Caractéristiques temporelles des signaux transitoires définis dans la norme G1000-4-4 et des transitoires collectés à bord d'un train .....	18
Figure 1.4. Illustration de la variation de l'intervalle de temps entre deux transitoires successifs à bord d'un train.....	19
Figure 1.5. Bandes de fréquences allouées au GSM-R .....	20
Figure 1.6. Description du système TDMA employé par le GSM-R.....	21
Figure 1.7. Evolution de la puissance du signal GSM-R reçu à bord du train au cours de son déplacement.....	21
Figure 1.8. Origine des perturbations EM transitoires reçues par l'antenne GSM-R fixée en toiture .....	22
Figure 1.9. Etude de la puissance générée par les perturbations EM transitoires sur la bande du lien descendant du GSM-R.....	23
Figure 1.10. Principe de la mesure .....	24
Figure 1.11. Evolution du BER en fonction de l'intervalle de temps entre deux transitoires successifs pour trois valeurs du SNR .....	25
Figure 1.12. Les différentes étapes de l'approche proposée dans cette thèse.....	27
Figure 2.1. Illustration du principe de la classification dans le cas où deux paramètres sont utilisés pour définir l'espace dans lequel le problème de classification doit être résolu.....	30
Figure 2.2. Règle des k plus proches voisins dans une espace de dimension 2 .....	31
Figure 2.3. Relation entre les couches successives de neurones .....	32
Figure 2.4. Principe de la phase d'apprentissage sur un réseau de neurones.....	32
Figure 2.5. Principe des séparateurs à vaste marge.....	33
Figure 2.6. Classification d'un élément inconnu avec les séparateurs à vaste marge .....	33
Figure 2.7. Principales méthodes de classification non-supervisées .....	34
Figure 2.8. Dendrogramme (à gauche) et méthode de détermination du nombre de classes (à droite).....	35
Figure 2.9. Principe de la représentation temps-fréquence .....	37
Figure 2.10. Familles de représentations temps-fréquence .....	38
Figure 2.11. Résolutions en temps et en fréquence de la transformée de Fourier à court terme .....	39
Figure 2.12. Représentations temporelles des trois fenêtres considérées .....	40
Figure 2.13. Transformées de Fourier des trois fenêtres considérées .....	41
Figure 2.14. Comparaison des résultats obtenus lors de l'utilisation de ces trois fenêtres pour le calcul du spectre d'un même signal temporel.....	41
Figure 2.15. Résolutions en temps et en fréquence de la transformée en ondelettes .....	42
Figure 2.16. Matrice des coefficients de la transformée en ondelettes .....	43
Figure 2.17. Schéma de principe de la transformée en ondelettes discrète.....	45
Figure 2.18. Arbre de décomposition en paquets d'ondelettes à l'ordre 3.....	46
Figure 2.19. Distribution de Wigner-Ville d'un signal de saut de fréquence .....	48
Figure 2.20. Distribution de Choi-Williams du même signal de saut de fréquence que pour la distribution de Wigner-Ville.....	49
Figure 3.1. Mélange « signal + bruit » généré.....	52
Figure 3.2. Signal transitoire employé.....	52
Figure 3.3. Matrice des coefficients de la transformée.....	54
Figure 3.4. Matrice construite après seuillage des coefficients de la transformée .....	55
Figure 3.5. Identification des colonnes susceptibles de contenir un transitoire .....	55
Figure 3.6. Principe de regroupement des colonnes de détection et détermination des instants d'apparition associés .....	56
Figure 3.7. Principe d'élimination des détections redondantes par « fenêtre de trois ».....	57
Figure 3.8. Affichage des résultats.....	58
Figure 3.9. Fonction d'ondelette Db3.....	61
Figure 3.10. Représentation de la décomposition sous la forme d'un arbre.....	62

Figure 3.11. Représentation de la valeur absolue des coefficients de la décomposition sous la forme d'un arbre.....	63
Figure 3.12. Cartographie de l'énergie.....	64
Figure 3.13. Matrice d'énergie E.....	64
Figure 3.14. Construction de la matrice d'énergie normalisée.....	65
Figure 3.15. Représentation de la première colonne de U pour les enregistrements de bruit des séries de mesures 1 et 2.....	67
Figure 3.16. Représentation de la première colonne de U pour les transitoires des séries de mesures 1 et 2.....	67
Figure 3.17. Repérage et sélection des nœuds significatifs sur l'arbre de décomposition.....	68
Figure 3.18. Sélection des nœuds 2 et 4 sur l'ensemble des enregistrements.....	68
Figure 3.19. Diagramme de dispersion des 4 groupes.....	69
Figure 3.20. Calcul de toutes les distances entre l'élément à identifier et ses voisins.....	69
Figure 3.21. Essai de la méthode de classification sur les transitoires de la série de mesures 2.....	70
Figure 3.23. Objectif de notre approche.....	73
Figure 3.24. Banc de mesures/tests développé.....	74
Figure 3.25. Exemple de séquence de transitoires générée par l'AWG.....	75
Figure 4.1. Principe utilisé pour optimiser la transformée de Gabor.....	82
Figure 4.2. Représentation de la matrice des coefficients de notre transformée de Gabor optimisée.....	83
Figure 4.3. Exemple de représentation temps-fréquence obtenue avec notre transformée.....	83
Figure 4.4. Détection de la fréquence GSM-R employée par la communication.....	84
Figure 4.5. Puissances sur le canal GSM-R effectivement employé et celui détecté.....	85
Figure 4.6. Principe de regroupement des fenêtres lors du calcul du descripteur.....	86
Figure 4.7. Exemple de courbe du descripteur.....	87
Figure 4.8. Exemple d'extraction des paramètres à partir de la courbe du descripteur pour deux fichiers de mesures.....	88
Figure 4.9. Scénario de bruit EM transitoire n°1.....	89
Figure 4.10. Représentation de base réf. 1.....	89
Figure 4.11. Diagramme de dispersion des éléments de base réf. 1.....	90
Figure 4.12. Tableau de correspondance entre le Rxqual et le BER.....	91
Figure 4.13. Utilisation de 3 classes pour le diagramme de dispersion.....	92
Figure 4.14. Détermination du meilleur seuil possible.....	93
Figure 4.15. Influence du seuil sur le diagramme de dispersion.....	94
Figure 4.16. Vue d'ensemble de l'interface de calcul et visualisation.....	95
Figure 4.17. Définition des paramètres de calcul.....	96
Figure 4.18. Choix du couple (fen, fpas).....	97
Figure 4.19. Détermination et application du seuil.....	98
Figure 4.20. Exemples de mauvais choix de seuil.....	98
Figure 4.21. Intervalle de suivi dans le temps.....	99
Figure 4.22. Correspondance diagramme de dispersion vers descripteur.....	100
Figure 4.23. Correspondance descripteur vers diagramme de dispersion.....	101
Figure 4.24. Barre d'outils.....	101
Figure 5.1. Scénarios de test n° 1, 2, 3 et 4.....	104
Figure 5.2. Diagramme de dispersion de référence.....	105
Figure 5.3. Matrice de confusion de la base de référence.....	106
Figure 5.4. Valeurs du BER sur 100 mesures consécutives réalisées dans des conditions de test identiques.....	107
Figure 5.5. Diagramme de dispersion de référence après correction.....	107
Figure 5.6. Scénarios de test utilisant des signaux transitoires enregistrés in situ.....	109
Figure 5.7. Diagramme de dispersion des éléments des bases de référence et de développement.....	110
Figure 5.8. Matrice de confusion de la base de développement.....	111
Figure 5.9. Diagramme de dispersion des éléments des bases de référence et de développement après correction de cette dernière, nommé « diagramme de référence complet ».....	111
Figure 5.10. Tracé de la frontière entre les deux classes à partir des éléments des bases de référence et de développement (corrigées).....	112
Figure 5.11. Séparation du diagramme de référence en deux zones de classification.....	113
Figure 5.12. Diagramme de dispersion des éléments de test.....	113
Figure 5.13. Matrice de confusion des éléments de test.....	114

<i>Figure 5.14. Compensation de la perte de dynamique sur la base de développement.....</i>	<i>115</i>
<i>Figure 5.15. Compensation de la perte de dynamique sur la base de test.....</i>	<i>115</i>
<i>Figure 5.16. Scénarios construits à partir de signaux transitoires collectés à bord du métro (amplitudes à l'entrée du mobile GSM-R).....</i>	<i>117</i>
<i>Figure 5.17. Couverture spectrale du scénario n°1.....</i>	<i>118</i>
<i>Figure 5.18. Couverture spectrale du scénario n°2.....</i>	<i>118</i>
<i>Figure 5.19. Comparaison des spectres du modèle sinus-expo et du transitoire issu du scénario n°2.....</i>	<i>119</i>
<i>Figure 5.20. Exemple de courbe du descripteur obtenue en présence d'un unique évènement transitoire couvrant les bandes du GSM-R .....</i>	<i>120</i>
<i>Figure 5.21. Courbe du descripteur obtenue pour le scénario n°1 .....</i>	<i>120</i>
<i>Figure 5.22. Courbe du descripteur obtenue pour le scénario n°2 .....</i>	<i>121</i>
<i>Figure 5.23. Définition du nouveau seuil.....</i>	<i>122</i>
<i>Figure 5.24. Comparaison des paramètres extraits (nombre de minima, somme des minima) avec application du seuil initial et du nouveau seuil .....</i>	<i>123</i>
<i>Figure 5.25. Cas de non-détection d'un minimum avec le nouveau seuil dans le cas du fichier n°9.....</i>	<i>123</i>
<i>Possible implémentation de notre travail à bord d'un train .....</i>	<i>128</i>

# LISTE DES TABLEAUX

<i>Tableau 1. Bilan comparatif des deux familles de représentations temps-fréquence</i>	38
<i>Tableau 2. Ensemble de variation des paramètres des signaux de test</i>	53
<i>Tableau 3. Bilan comparatif des résultats obtenus avec les deux transformées temps-fréquence</i>	59
<i>Tableau 4. Résultats des essais de classification</i>	71

---

# REFERENCES

## CHAPITRE 1

**[1.1] Fatma Rouissi, Olivier Delangre, Virginie Degardin, Martine Lienard, Marc Heddebaut, Virginie Deniau, Pierre Degauque**

Intra-vehicle digital communications: characterization and simulation of the electromagnetic environment

*Comptes Rendus Physique Tome 10 / fascicule 1: "Nouvelles approches en compatibilité électromagnétique", pp 22-30, 2009.*

**[1.2] David Middleton**

Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance

*IEEE Transactions on electromagnetic compatibility, vol. 21, no. 3, August 1979.*

**[1.3] David Middleton**

Procedures for Determining the Parameters of the First-Order Canonical Models of Class A and Class B Electromagnetic Interference

*IEEE Transactions on electromagnetic compatibility, vol. 21, no. 3, August 1979.*

**[1.4] David Middleton**

Statistical-Physical Models of Electromagnetic Interference

*IEEE Transactions on electromagnetic compatibility, vol. 19, no. 3, August 1977.*

**[1.5] David Middleton**

An introduction to statistical communication theory

*Editions McGraw-Hill, New York, 1960.*

**[1.6] EN 50121**

Railway applications – Electromagnetic compatibility – Part 1 to 5

*European Standard, 2006.*

**[1.7] EN 61000-4-2**

Electromagnetic Compatibility – Part 4: Testing and Measurement techniques – Section 2: Electrostatic discharge immunity test

*European Standard, April 2001.*

**[1.8] EN 61000-4-3**

Electromagnetic Compatibility – Part 4: Testing and Measurement techniques – Section 3: Radiated, radio-frequency, electromagnetic field immunity test

*European Standard, August 2002.*

**[1.9] EN 61000-4-4**

Electromagnetic Compatibility – Part 4: Testing and Measurement techniques – Section 4: Electrical fast transient/burst immunity test

*European Standard, June 1995.*

**[1.10] EN 61000-4-5**

Electromagnetic Compatibility – Part 4: Testing and Measurement techniques – Section 5: Surge immunity test

*European Standard, June 1995.*

**[1.11] ETSI TS 145 001**

Digital cellular telecommunications system (Phase 2+); Physical layer on the radio path; General description

*Technical specification produced by ETSI 3rd Generation Partnership Project (3GPP), V11.0.0, Oct. 2012.*

**[1.12] ETSI TS 145 002**

Digital cellular telecommunications system (Phase 2+); Multiplexing and multiple access on the radio path

*Technical specification produced by ETSI 3rd Generation Partnership Project (3GPP), V11.1.0, Feb. 2013.*

**[1.13] T. Hammi, N. Ben Slimen, V. Deniau, J. Rioult and S. Dudoyer**

Comparison between GSM-R coverage level and EM noise level in railway environment

*9th International Conference on Intelligent Transport System-Telecommunications (ITST 2009), pp 123-128, Lille, France, October 2009.*

**[1.14] UIC Project EIRENE**

System Requirements Specification

*Spécifications que doit respecter le système de communication numérique ferroviaire européen, v14.0, 21 Oct. 2003.*

**[1.15] S. Dudoyer, N. Ben Slimen, V. Deniau and M. Berbineau**

Reliability of the GSM-R Communication System against Railway Electromagnetic Interferences.

*9th World Congress on Railway Research (WCRR 2011), Lille, France, 22-26 May 2011.*

**[1.16] ETSI TS 145 005**

Digital cellular telecommunications system (Phase 2+); Radio transmission and reception

*Technical specification produced by ETSI 3rd Generation Partnership Project (3GPP), V10.0, April 2011.*

**[1.17] Mohamed Nedim Ben Slimen**

Recherche de Procédures de Caractérisation de l'Environnement Electromagnétique Ferroviaire Adaptées au Contexte des Systèmes de Communications Embarqués

*Thèse de doctorat de l'Université des Sciences et Technologies de Lille (Lille 1), soutenue le 18 décembre 2009.*

**[1.18] S. Dudoyer, V. Deniau, R. Adriano, M. N. Ben Slimen, J. Rioult, B. Meyniel and M. Berbineau**

Study of the Susceptibility of the GSM-R Communications to the Electromagnetic Interferences of the Rail Environment

*IEEE Transactions on electromagnetic compatibility, vol. 54, no. 3, June 2012.*

**[1.19] S. Dudoyer, R. Adriano, V. Deniau, J. Rioult, N. Ben Slimen, B. Meyniel**

Testing of the GSM-R System against Electromagnetic Disturbances present in the Railway Environment

*9th International Conference on Intelligent Transport System-Telecommunications (ITST 2009), pp 117-122, Lille, France, Octobre 2009.*

**[1.20] Stephen Dudoyer, Virginie Deniau, Nedim Ben Slimen and Ricardo Adriano**

Susceptibility of the GSM-R Transmissions to the Railway Electromagnetic Environment

*Ouvrage InTech 2012, ISBN: 978-953-51-0448-3.*

*Disponible en ligne : <http://www.intechopen.com/books/infrastructure-design-signalling-and-security-in-railway/susceptibility-of-the-gsm-r-transmissions-to-the-railway-electromagnetic-environment>*



- [1.21] **N. Ben Slimen, V. Deniau, J. Rioult, S. Dudoyer, and S. Baranowski**  
 Statistical characterisation of the EM interferences acting on GSM-R antennas fixed above moving trains  
*Eur. Phys. J. Appl. Phys.*, vol. 48, 21202, 2009.
- [1.22] **CISPR 16-1**  
 Spécifications des méthodes et des appareils de mesure des perturbations radioélectriques et de l'immunité aux perturbations radioélectriques – Partie 1 : Appareils de mesure des perturbations radioélectriques et de l'immunité aux perturbations radioélectriques  
*Comité International Spécial Des Perturbations Radioélectriques*.

## CHAPITRE 2

- [2.1] **Richard O. Duda, Peter E. Hart, and David G. Stork**  
 Pattern Classification (2nd Edition)  
*Edited by Wiley-Interscience, 2000.*
- [2.2] **A. Ganapathiraju, J. E. Hamaker, J. Picone**  
 Applications of support vector machines to speech recognition  
*IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2348-2355, Aug. 2004.
- [2.3] **Wang Jin-Jia, Jiang Ke-Mei, Ma Chong-Xiao**  
 GMM-based Detection Methods in EEG-based Brain-Computer Interfaces  
*First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)*, pp. 779-782, 17-19 Sept. 2010.
- [2.4] **T. S. Tabatabaei, S. Krishnan, A. Anpalagan**  
 SVM-based classification of digital modulation signals  
*IEEE International Conference on Systems Man and Cybernetics (SMC)*, pp. 277-280, 10-13 Oct. 2010.
- [2.5] **Guillaume Bouchard**  
 Les modèles génératifs en classification supervisée et applications à la catégorisation d'images et à la fiabilité industrielle  
*Thèse de doctorat de l'université Joseph Fourier – Grenoble 1, soutenue en 2005.*
- [2.6] Classification non supervisée  
 Eléments de cours disponibles à : <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-explo-classif.pdf>
- [2.7] **Sandun Kodituwakku**  
 Radial Kernel based Time-Frequency Distributions with Applications to Atrial Fibrillation Analysis  
*PhD Student presentation, The Australian National University, Canberra, Australia.*
- [2.8] **Mickael Duquenoy**  
 Analyse Temps - Fréquence appliquée à l'imagerie SAR polarimétrique  
*Thèse de doctorat de l'Institut d'Electronique et de Télécommunications de Rennes, soutenue à Rennes le 20 octobre 2009.*
- [2.9] **François Auger, Franz Hlawatsch et al.**  
 Temps-fréquence – concepts et outils  
*Traitement du signal et de l'image, Hermès, Paris, 2005.*

**[2.10] Ulrich Günther**

Advanced NMR Processing

*EuroLabCourse "Advanced Computing in NMR Spectroscopy", Florence, Sept. 2001.*

**[2.11] F. Hlawatsch and G. F. Boudreaux-Bartels**

Linear and Quadratic Time-Frequency Signal Representations

*IEEE Signal Processing Mag., vol. 9, no. 2, April 1992, pp. 21-67.*

**[2.12] Kartik Kariya, B RajaSekhar, C.H.Ravichandrababu**

The Gabor Transform, SFTF and CWT Invertibility and generalized Parseval's like theorem.

**[2.13] C. Demars**

Chapitre 1 – Représentation de type Gabor

LIMSI-CNRS, 2000.

**[2.14] B. Friedlander, B. Porat**

Detection of transient signals by the Gabor representation

*IEEE Transactions on Acoustics, Speech and Signal Processing, Feb. 1989, Volume 37 – issue 2, pp 169-180.*

**[2.15] André Quinquis, Cornel Ioana**

Représentations Temps-Fréquence et Temps-Echelle

*Rapport interne du Centre de recherche "extraction et exploitation de l'information en environnements incertains", ENSIETA, Brest, France, Janvier 2002.*

**[2.16] Michel Vacher et Dan Istrate**

Notes de lecture sur la Transformée en Ondelettes

*Rapport interne CLIPS-IMAG, Septembre 2004.*

**[2.17] Concettina Buccella, Mauro Feliziani, Giuliano Manzi**

Accurate Detection of Low Entropy Cable Faults by Wavelet Transform

*Présentation.*

**[2.18] Robi Polikar**

"The Wavelet Tutorial"

<http://users.rowan.edu/~polikar/WAVELETS/WTpart3.html>

**[2.19] Robi Polikar**

The Wavelet Tutorial

<http://users.rowan.edu/~polikar/WAVELETS/WTpart4.html>

**[2.20] Marwa Chendeb**

Détection et classification des signaux non stationnaires par utilisation des ondelettes - Application aux signaux électromyographiques utérins

*Thèse de doctorat de l'université de technologie de Troyes, soutenue le 14 mars 2006.*

**[2.21] F. Auger, C. Doncarli**

Etude des propriétés théoriques de quelques représentations temps-fréquence proposées récemment

*Treizième colloque GRETSI, Juan-les-pins, 16-20 Septembre 1991.*

**[2.22] Satish Rajagopalan, José A. Restrepo, José M. Aller, Thomas G. Habetler, Ronald G. Harley**

Selecting Time-Frequency Representations for Detecting Rotor Faults in BLDC Motors Operating Under Rapidly Varying Operating Conditions

*Publication IEEE, 2005.*

**[2.23] Stéphane Mallat**

Une exploration des signaux en ondelettes

*Les Editions de l'Ecole Polytechnique, Novembre 2000, ISBN : 2-7302-0733-3.*

**[2.24] Matteo Gandetto, Marco Guainazzo, Carlo S. Regazzoni**

Use of Time-Frequency Analysis and Neural Networks for Mode Identification in a Wireless Software-Defined Radio Approach

*EURASIP Journal on Applied Signal Processing, December 2004, pp 1778–1790.*

**[2.25] Anna H. Pryor, Marianne Mosher, David G. Lewicki**

The Application of Time-Frequency Methods to HUMS

*Presented at the American Helicopter Society's 57th Annual Forum, Washington D.C., 9-11 May 2001.*

**[2.26] Guillermo G. Gaunard and Hans C. Strifors**

Applications of (Wigner-Type) Time-Frequency Distributions to Sonar and Radar Signal Analysis.

**[2.27] A.C.A. Figueiredo, M.F.F. Nave and JET-EFDA Contributors**

Time-Frequency Analysis of Nonstationary Fusion Plasma Signals: A Comparison between the Choi-Williams Distribution and Wavelets

*15th HTPD Conference, San Diego California, 18-22 April 2004.*

**CHAPITRE 3****[3.1] N. Ben Slimen, V. Deniau, J. Rioult, S. Dudoyer, S. Baranowski**

Statistical characterisation of the EM Interferences acting on GSM-R antennas fixed above moving trains

*The European Physical Journal - Applied Physics, vol. 48, 2009.*

**[3.2] Alain Dufaux**

Detection and Recognition of Impulsive Sound Signals

*Thèse de doctorat de l'Université de Neuchâtel, soutenue en janvier 2001.*

**[3.3] David L. Donoho and Iain M. Johnstone**

Ideal spatial adaptation by wavelet shrinkage

*Biometrika (1994), 81, 3, pp 425-455.*

**[3.4] Philippe Ravier**

Détection de transitoires par ondelettes adaptées - Critères d'adaptation fondés sur les statistiques d'ordre supérieur

*Thèse de doctorat de l'Institut National Polytechnique de Grenoble, soutenue le 6 juillet 1998.*

**[3.5] S. Dudoyer, V. Deniau, M. Ayad, M. Heddebaut**

Analyse d'environnements électromagnétiques transport à l'aide de transformations temps-fréquence

*JFMMA 2011, Tanger, Maroc, 16-18 Mars 2011.*

**[3.6] Rachel E. Learned**

Wavelet Packet Based Transient Signal Classification

*Master of Science Thesis, September 1992.*

**[3.7] David C. Lay**

Linear Algebra and Its Applications, 4<sup>th</sup> edition

*Editions Pearson, chapitre 7 (ISBN: 0321385179).*

**[3.8] S. Dudoyer, V. Deniau, R. Adriano, M. N. Ben Slimen, J. Rioult, B. Meyniel and M. Berbineau**

Study of the Susceptibility of the GSM-R Communications to the Electromagnetic Interferences of the Rail Environment

*IEEE Transactions on electromagnetic compatibility*, vol. 54, no. 3, June 2012.

**[3.9] Rohde & Schwarz**

CMU operating manual, v4-50.

**[3.10] Recommandation UIT-T K.48**

Prescriptions de CEM pour les équipements de télécommunication - Recommandation relative aux familles de produits, Sept. 2006.

## CHAPITRE 4

**[4.1] S. Dudoyer, V. Deniau, R. Adriano, M. N. Ben Slimen, J. Rioult, B. Meyniel and M. Berbineau**

Study of the susceptibility of the GSM-R communications to the electromagnetic interferences of the rail environment

*IEEE Trans. on EMC*, vol. 54, no. 3, June 2012.

**[4.2] Xavier Lagrange, Philippe Godlewski, Sami Tabbane**

Réseaux GSM-DCS

*Editions Hermès, Paris, 1995,1996.*

**[4.3] Recommandation UIT-T K.48**

Prescriptions de CEM pour les équipements de télécommunication - Recommandation relative aux familles de produits, Sept. 2006.

**[4.4] Nabil Ech charraq**

Développement d'interfaces graphiques pour la classification d'environnements électromagnétiques

*Rapport de stage License informatique, 2012.*

## CHAPITRE 5

**[5.1] N. Ben Slimen, V. Deniau, J. Rioult, S. Dudoyer, and S. Baranowski**

Statistical characterisation of the EM interferences acting on GSM-R antennas fixed above moving trains

*Eur. Phys. J. Appl. Phys.*, vol. 48, 21202, 2009.

**[5.2] S. Dudoyer, V. Deniau, R. Adriano, M. N. Ben Slimen, J. Rioult, B. Meyniel and M. Berbineau**

Study of the Susceptibility of the GSM-R Communications to the Electromagnetic Interferences of the Rail Environment.

*IEEE Transactions on electromagnetic compatibility*, vol. 54, no. 3, June 2012.

**[5.3] S. Dudoyer, R. Adriano, V. Deniau, N. Ben Slimen and B. Meyniel**

Testing of the GSM-R System against Electromagnetic Disturbances present in the Railway Environment

*9th International Conference on Intelligent Transport System-Telecommunications (ITST2009), Lille, France, Oct. 2009, pp. 117–122.*

**[5.4] Stephen Dudoyer, Virginie Deniau, Sébastien Ambellouis, Marc Heddebaut, Andrea Mariscotti**

Recherche de descripteurs adaptés à l'analyse du bruit EM agissant sur la qualité des transmissions GSM-R.

*CEM 2012, Rouen, France, 25-27 Avril 2012.*

**[5.5] Benjamin Ruytoor**

TIFANY (TIme Frequency ANALYser)

*Rapport de stage License informatique, 2012.*



# BIBLIOGRAPHIE DE L'AUTEUR

## Revue

**S. Dudoyer, V. Deniau, R. Adriano, M. N. Ben Slimen, J. Rioult, B. Meyniel and M. Berbineau**  
Study of the Susceptibility of the GSM-R Communications to the Electromagnetic Interferences of the Rail Environment.

*IEEE Transactions on electromagnetic compatibility, vol. 54, no. 3, June 2012.*

**S. Dudoyer, V. Deniau, S. Ambellouis, M. Heddebaut and A. Mariscotti**

Classification of Transient EM Noises Depending on their Effect on the Quality of GSM-R Reception.  
*IEEE Transactions on electromagnetic compatibility, accepted on 31<sup>th</sup> December 2012, published online: 13<sup>th</sup> February 2013.*

## Chapitre d'ouvrage

**Stephen Dudoyer, Virginie Deniau, Nedim Ben Slimen and Ricardo Adriano**

Susceptibility of the GSM-R Transmissions to the Railway Electromagnetic Environment

*Ouvrage Intech "Infrastructure Design, Signalling and Security in Railway", 2012*

Disponible à : <http://www.intechopen.com/books/infrastructure-design-signalling-and-security-in-railway/susceptibility-of-the-gsm-r-transmissions-to-the-railway-electromagnetic-environment>

## Conférences

**S. Dudoyer, V. Deniau, N. Ben Slimen, J. Rioult, B. Meyniel**

Comparaison entre différentes méthodologies de test d'immunité des transmissions GSM-R.

*CEM 2010, Limoges, France, 7-9 Avril 2010.*

**S. Dudoyer, V. Deniau, M. Ayad, M. Heddebaut**

Analyse d'environnements électromagnétiques transport à l'aide de transformations temps-fréquence.

*JFFMA 2011, Tanger, Maroc, 16-18 Mars 2011.*

**S. Dudoyer, N. Ben Slimen, V. Deniau and M. Berbineau**

Reliability of the GSM-R Communication System against Railway Electromagnetic Interferences.

*9th World Congress on Railway Research (WCRR 2011), Lille, France, 22-26 May 2011.*

**S. Dudoyer**

Contribution à la détection et à la reconnaissance de bruits électromagnétiques ciblées en fonction de leurs effets sur les transmissions numériques.

*Journée des doctorants 2011, Villeneuve d'Ascq, France, 7 Juin 2011.*

**S. Dudoyer, V. Deniau, M. Heddebaut, S. Ambellouis, A. Mariscotti**

Méthode de classification de bruits électromagnétiques en fonction de leurs effets sur la qualité des communications GSM-R.

*Assemblée générale du GDR Ondes, Nice, France, 24-26 Octobre 2011.*

**Stephen Dudoyer, Virginie Deniau, Sébastien Ambellouis, Marc Heddebaut, Andrea Mariscotti**  
Recherche de descripteurs adaptés à l'analyse du bruit EM agissant sur la qualité des transmissions GSM-R.

*CEM 2012, Rouen, France, 25-27 Avril 2012.*

**Prix de la meilleure présentation de la conférence**

**Stephen Dudoyer, Virginie Deniau, Sébastien Ambellouis, Marc Heddebaut et Andrea Mariscotti**

Classification de bruits EM transitoires en fonction de leur effet sur la qualité des transmissions GSM-R.

*Journées des doctorants 2012, Villeneuve d'Ascq, France, 13-14 Juin 2012*

**Virginie Deniau, Stephen Dudoyer, Sébastien Ambellouis, Marc Heddebaut, Andrea Mariscotti**  
Researching of observables adapted to the analysis of EM noise acting on the quality of the GSM-R transmissions.

*EMC Europe 2012, Rome, Italy, 17-21 September 2012.*

**V. Deniau, S. Dudoyer, M. Heddebaut, A. Mariscotti, A. Marrese, N. Pasquino**

Test bench for the evaluation of GSM-R operation in the presence of electric arc interference

*Electrical Systems for Aircraft, Railway and Ship Propulsion (ESARS), Bologna, Italy, 16-18 October 2012.*





## **Résumé :**

Avec la multiplication des systèmes électroniques analogiques et numériques embarqués dans les moyens de transports actuels, l'environnement EM (ElectroMagnétique) devient de plus en plus riche en signaux de toutes sortes et, par conséquent, il devient plus complexe à caractériser.

Dans ces travaux de thèse, nous nous intéressons à un système numérique en particulier : le GSM-R (Global System for Mobile communications - Railways). Il s'agit du nouveau système de communication radio numérique portant la signalisation actuellement en cours de déploiement sur le réseau ferroviaire européen afin d'assurer l'interopérabilité des trains à grande vitesse en Europe. Ainsi, tous les pays d'Europe utiliseront le même système, ce qui facilitera la circulation des trains transfrontaliers d'un pays à l'autre.

En environnement ferroviaire, le GSM-R doit faire face à différentes sources de bruit EM, notamment aux perturbations EM transitoires provenant du contact glissant entre le fil caténaire et le pantographe. Ces perturbations couvrent de larges bandes de fréquences dont, potentiellement, celles du GSM-R.

Nous proposons une méthode de classification qui permet de prédire l'effet des perturbations EM transitoires sur la qualité des transmissions GSM-R.

Cette méthode de classification pourrait être mise en place à bord des trains où elle permettrait d'identifier et localiser les zones critiques pour la qualité des transmissions GSM-R tout au long du trajet.

Du point de vue de la standardisation, ces travaux de recherche pourraient contribuer à l'évolution nécessaire des équipements et des méthodologies définies dans les normes CEM afin de s'adapter aux nouvelles problématiques émanant de la multiplicité des systèmes et protocoles de communication sans fils employés dans le monde des transports.

## **Summary:**

With the proliferation of analog and digital electronic systems in the current means of transport, the EM (Electromagnetic) environment becomes richer and richer in all kinds of signals and, therefore, it becomes more difficult to characterize.

In this thesis, we focus on a particular digital system: the GSM-R (Global System for Mobile communications - Railways). It is the new digital radio communication system under deployment on the European rail network in order to ensure the interoperability of high-speed trains in Europe. Then, all countries in Europe will use the same system, which will facilitate the movement of cross-border trains from a country to another one.

In the railway environment, the GSM-R is subject to different EM noise sources, including transient EM interferences coming from the sliding contact between the pantograph and the catenary. These disturbances cover wide frequency bands including, potentially, those of the GSM-R system. We propose a classification method for predicting the effect of transient EM disturbances on the quality of GSM-R transmissions.

This classification method could be implemented on trains where it could identify and locate critical areas for the quality of GSM-R transmissions along the journey.

From the point of view of standardization, these research works could contribute to the necessary evolution of equipment and methods defined in EMC standards in order to adapt to the new problems arising from the multiplicity of wireless communication systems and protocols employed nowadays in the world of transports.