UNIVERSITÉ DE
SHERBROOKE

Faculté de génie

Département de génie électrique et de génie informatique

APPROCHES PARAMÉTRIQUES POUR LE CODAGE AUDIO MULTICANAL

Mémoire de maîtrise ès sciences appliquées
Spécialité : génie électrique et génie informatique

Jimmy LAPIERRE

Sherbrooke (Québec), Canada Avril 2007

One of the most striking facts about our ears is that we have two of them and yet we hear one acoustic world; only one voice per speaker [1]

[1] E. C. Cherry and W. K. Taylor. Some further experiments on the recognition of speech, with one and two ears. *Journal of the Acoustic Society of America*, 26:554-559, 1954.

SOMMAIRE

Afin de répondre aux besoins de communication et de divertissement, il ne fait aucun doute que la parole et l'audio doivent être encodés sous forme numérique. En qualité CD, cela nécessite un débit numérique de 1411.2 kb/s pour un signal stéréophonique. Une telle quantité de données devient rapidement prohibitive pour le stockage de longues durées d'audio ou pour la transmission sur certains réseaux, particulièrement en temps réel (d'où l'adhésion universelle au format MP3). De plus, ces dernières années, la quantité de productions musicales et cinématographiques disponibles en cinq canaux et plus ne cesse d'augmenter. Afin de maintenir le débit numérique à un niveau acceptable pour une application donnée, il est donc naturel pour un codeur audio à bas débit d'exploiter la redondance entre les canaux et la psychoacoustique binaurale. Le codage perceptuel et plus particulièrement le codage paramétrique permet d'atteindre des débits manifestement inférieurs en exploitant les limites de l'audition humaine (étudiées en psychoacoustique). Cette recherche se concentre donc sur le codage paramétrique à bas débit de plus d'un canal audio.

ABSTRACT

In order to fulfill our communications and entertainment needs, there is no doubt that speech and audio must be encoded in digital format. In “CD” quality, this requires a bit-rate of 1411.2 kb/s for a stereo signal. Such a large amount of data quickly becomes prohibitive for long-term storage of audio or for transmitting on some networks, especially in real-time (leading to a universal adhesion to the MP3 format). Moreover, throughout the course of these last years, the number of musical and cinematographic productions available in five channels or more continually increased. In order to maintain an acceptable bit-rate for any given application, it is obvious that a low bit-rate audio coder must exploit the redundancies between audio channels and binaural psychoacoustics. Perceptual audio coding, and more specifically parametric audio coding, offers the possibility of achieving much lower bit-rates by taking into account the limits of human hearing (psychoacoustics). Therefore, this research concentrates on parametric audio coding of more than one audio channel.

REMERCIEMENTS

En premier lieu, je désire remercier vivement et très sincèrement mon directeur, le professeur Roch Lefebvre. D'une part, c'est grâce à lui que j'ai entrepris des études aux cycles supérieurs. D'autre part, ses conseils et son appui ont été appréciés tout au long de ces dernières années.

Ensuite, je désire exprimer une reconnaissance tout aussi incontournable à ma conjointe, Geneviève Veilleux, dont les encouragements et le soutien moral furent déterminants dans mes accomplissements.

De même, un merci tout spécial à mon bon ami Martin Bouchard, qui m'a rendu le parcours universitaire bien plus agréable en raison de son support et de son écoute.

Ma gratitude va également à mes collègues du GRPA, dont plusieurs ont participé à des évaluations subjectives réalisées pour ce mémoire. J'aimerais remercier plus particulièrement Jonathan Fillion-Deneault, Marie Oger, Catherine Lemyre ainsi que Philippe Gournay, qui m'ont aidé à cheminer autant au niveau académique que personnel. Merci aussi à Danielle Poirier, dont le travail, souvent dans l'ombre, nous facilite la vie à tous !

Je ne saurais passer sous silence le soutien financier de l'entreprise VoiceAge, du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), de tous les gens qui ont contribué à la Fondation de l'Université de Sherbrooke et de M^{e} Louis Lagassé qui a créé le Fonds Suzanne et Jacques Lagassé pour venir en aide aux étudiants de deuxième cycle en génie électrique. Le soutien financier offert par ces différentes instances m'aura certes permis de mener à terme mon projet de maîtrise.

Enfin, mes derniers remerciements vont à mes supporteurs inconditionnels, mes parents, qui m'ont toujours appuyé dans tous mes choix et réalisations. Par le fait même, je dédie cet ouvrage à ma mère qui aurait tant voulu poursuivre des études, mais qui n'a pas eu cette chance.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

1. INTRODUCTION

Le codage de la parole et de l'audio sont employés dans une grande variété d'applications, telles que les communications personnelles et le divertissement multimédia. Que ce soit directement par Internet ou indirectement par l'entremise d'un fournisseur de services de téléphonie, de radio ou de télévision, les représentations numériques permettent un stockage et une transmission fiable et efficace des signaux sonores. Toutefois, le volume de ces données peut devenir rapidement prohibitif, d'où les nombreux algorithmes de compression. Ceux-ci ont pour but de réduire la redondance et parfois la non-pertinence des informations contenues à l'intérieur d'une représentation audionumérique donnée. Alors qu'un algorithme dit « sans perte » comme FLAC a pour seul but de réduire la redondance dans la représentation binaire de l'information, une approche « avec pertes » comme celle employée dans MP3 exploite les limites de la perception auditive humaine en sacrifiant certaines informations jugées peu importantes pour ainsi atteindre des taux de compression considérablement plus élevés.

1.1 Psychoacoustique

La psychoacoustique est l'étude scientifique de la perception subjective des sons par des humains. Ce domaine de recherche s'intéresse entre autres aux limites de notre appareil auditif, que ce soit les limites absolues ou la résolution de celui-ci. Par exemple, il est généralement reconnu que les humains peuvent normalement entendre les sons entre 20Hz et 20kHz. On sait aussi que la résolution en fréquence et en intensité de notre système auditif est relativement logarithmique et non linéaire.

Parmi les sujets d'intérêt de la psychoacoustique particulièrement pertinents pour le codage audio, on y retrouve l'étude du masquage. En effet, les phénomènes de masquage temporels et fréquentiels sont employés dans les codecs « avec pertes » pour ainsi éviter de représenter de l'information qui est peu ou pas du tout audible. Par exemple, un algorithme qui détecte une note à une fréquence de 440Hz avec une intensité de 60dB pourrait choisir d'allouer très peu de bits à un autre son d'une intensité de 30dB dans la zone de 420 à 500Hz.

1.2 Codage de la parole et de l'audio

En plus d'être basés sur la psychoacoustique, les codecs perceptuels conçus pour la parole et ceux pour l'audio composent deux grandes familles. Alors qu'un codec pour l'audio s'appuie d'abord et avant tout sur les limites de l'audition humaine, un codec de parole se base davantage sur un modèle de production de la parole. L'avantage de ce dernier n'est pas facilement transférable à un codec audio générique, car il serait inefficace d'avoir un modèle de production de source sonore pour tous les sons qui existent dans la musique. Par contre, il est tout de même possible de concevoir des modèles paramétriques qui exploitent certaines limites de l'audition humaine.

1.3 Codage perceptuel de plus d'un canal audio

Le codage à bas débit de plus d'un canal audio se base principalement sur des modèles paramétriques qui exploitent la psychoacoustique binaurale, c'est-à-dire la perception des sons dans l'espace. L'objectif est généralement d'encoder un seul canal de façon traditionnelle, accompagné d'une représentation paramétrique, très compacte en termes de bits, du lieu et de l'ambiance générale des sons. La résolution fréquentielle et temporelle limitée du système auditif est exploitée, ainsi que les mécanismes de localisation comme l'amplitude relative des sons à chaque oreille. Ces modèles ont progressé de la simple modification des ratios d'énergie entre deux canaux en stéréo [1] aux modèles plus modernes couverts dans ce document.

1.4 Défis pour le codage audio multicanal

En examinant l'état de l'art existant au début de ces travaux de recherche, il était possible de constater que le codage perceptuel de l'audio à bas débit semblait s'appuyer principalement sur la psychoacoustique monaurale. En effet, mis à part quelques travaux récents en stéréo [2] et en multicanal [3], peu d'efforts semblaient avoir été consacrés à l'intégration de la psychoacoustique binaurale au codage paramétrique de plus d'un canal audio depuis l'avènement du « intensity stereo » en 1994 [1]. Une opportunité se présentait donc pour l'avancement de l'état de l'art dans ce domaine. D'ailleurs, c'est ce qui s'est produit pendant la durée de ces travaux. En

mars 2004, l'organisation MPEG a demandé des propositions sur le codage audio « spatial 5.1 » à très bas débit [4], maintenant connu sous l'appellation plus générale « MPEG Surround ». Durant la même période, le codage paramétrique de la stéréo a été mis au point [5] et adapté à un standard international à l'intérieur du 3GPP [6]. Pour sa part, la standardisation de la technologie « MPEG Surround » est attendue pour le début de 2007. L'exploitation des connaissances grandissantes en psycho-acoustique binaurale pour le codage de plus d'un canal audio était et demeure aujourd'hui un axe de recherche active et encore remplie de défis.

1.5 Contributions au domaine de recherche

Ces travaux de maitrise ont permis d'apporter des contributions significatives au codage audio stéréo, et par le fait même, au codage audio multicanal. Ces innovations sont des améliorations à l'état de l'art en codage stéréo paramétrique. Celles-ci exploitent une redondance dans les paramètres qui ont été proposés pour le standard [7] afin de réduire le débit requis pour transmettre l'information de phase. Cela est rendu possible sans diminuer la qualité sonore et sans augmenter ni la complexité, ni le délai de l'algorithme. Les solutions proposées dans ce mémoire permettent aussi d'éviter l'accroissement du délai et de la complexité lorsque le module de stéréo paramétrique et le codec mono subjacent ne partagent pas la même structure d'analyse temps-fréquence. Un bon exemple d'application pour ces optimisations est certainement le codec AMR-WB+ [8], qui possède une technologie hybride alternant entre les domaines temporel et fréquentiel selon les caractéristiques du signal d'entrée.

Plus précisément, une première contribution permet de réduire le débit nécessaire pour transmettre l'information de phase sans aucun impact sur la qualité sonore et sans aucune augmentation de la complexité et du délai algorithmiques.

Une deuxième contribution permet de compenser l'énergie d'un signal stéréo sommé dans un seul canal au décodeur. Sachant que l'énergie de la demi-somme de deux signaux n'est pas égale à la demi-somme des énergies de chaque canal (à cause de la phase et de la corrélation entre les canaux qui varient), le standard prévoit un

mélange (down-mix) à l'encodeur où l'énergie est compensée dans chaque sous-bande. Dans le cas d'un codec mono qui n'opère pas dans le même domaine temps-fréquence que la partie stéréo, cela ajoute un délai algorithmique important, sans compter la complexité engendrée par une transformée inverse supplémentaire. Alternativement, l'approche proposée permet d'effectuer cette compensation au décodeur sans perte de qualité. L'encodeur mono peut donc traiter sans délai le signal correspondant à la demi-somme des canaux stéréo.

Ces innovations ont été présentées à la 120^{e} convention de l'AES à Paris, France, en mai 2006 [9].

1.6 Sommaire du mémoire

Ce mémoire se concentre sur le codage perceptuel de plusieurs canaux audio à très bas débit, c'est-à-dire un codage paramétrique basé sur la psychoacoustique binaurale. Ces technologies permettent aux codeurs audio perceptuels modernes, comme les codecs AMR-WB+ [8] et Enhanced aacPlus [10], d'être étendus au codage efficace de deux ou plusieurs canaux audio. Dans un premier temps, le chapitre 2 présente les connaissances pertinentes à ce problème en psycho-acoustique, avec un accent sur la psychoacoustique binaurale. Puis, les technologies de codage audio stéréo et multicanal sont couvertes au chapitre 3, avec un accent cette fois-ci sur les techniques à bas débit. Ensuite, le chapitre 4 détaille les contributions apportées à ce domaine [9]. Finalement, le chapitre 5 conclut en présentant une analyse des performances des algorithmes proposés.

2. FONDEMENTS PSYCHOACOUSTIQUES

De nombreux paramètres du système auditif humain ont été étudiés dans le passé, d'où le domaine de la psychoacoustique, qui étudie les phénomènes se rapportant à la physiologie et au mécanisme psychosensoriel de l'audition. Ce chapitre présente un résumé des connaissances dans ce domaine, en accordant plus d'importance aux phénomènes exploités dans les codeurs perceptuels actuels. Un accent particulier est mis sur la psychoacoustique binaurale, dans le but de mieux comprendre les techniques de codage paramétrique conçus pour de multiples canaux.

2.1 Introduction à la psychoacoustique

En premier lieu, une introduction à la psychoacoustique s'impose. Il s'agit ici d'un résumé des éléments pertinents au codage perceptuel de l'audio à partir des œuvres de référence classiques [11-13] et non d'une revue exhaustive de la matière.

2.1.1 LA PHYSIQUE DES SONS

Notre système auditif est un système complexe qui nous permet de capter des variations de pression dans le temps, $P(t)$. Comparées aux variations de pression atmosphérique, les variations de pression rapides dans le temps causées par des sources sonores sont extrêmement faibles. L'unité de pression acoustique est le Pascal (Pa). En psychoacoustique, les valeurs de variation de pression acoustique pertinentes sont celles entre 10^{-5} Pa (le seuil absolu de l'audition) et 10^{2} Pa (le seuil de la douleur). Une autre mesure physique pertinente est l'intensité acoustique, définie comme étant la quantité d'énergie traversant par unité de temps l'unité de surface orientée parallèlement au front d'onde, en watt/m^2. La mesure du niveau de pression acoustique $\beta(t)$ en décibels (dB), la pression acoustique $P(t)$ et l'intensité acoustique $I(t)$ sont reliées par l'équation suivante :

$$\beta = 20\log\left(\frac{P}{P_0}\right) = 10\log\left(\frac{I}{I_0}\right) \quad \text{(dB)}. \tag{2.1}$$

Les valeurs de référence standards sont $P_0 = 20\mu\text{Pa}$ et $I_0 = 10^{-12}\ \text{W/m}^2$.

2.1.2 LE SEUIL DE L'AUDITION

Le seuil de l'audition (threshold of hearing) est défini comme étant la quantité d'énergie minimale requise pour qu'une tonalité pure (sinusoïde) soit détectée par le système auditif dans un environnement autrement silencieux. Deux mesures sont fréquemment utilisées : le « MAP » est le seuil d'audibilité d'une pression acoustique présentée très près du tympan d'une seule oreille et le « MAF » est le seuil d'audibilité d'un champ acoustique présenté aux deux oreilles (avec des haut-parleurs en chambre anéchoïque). Le seuil de l'audition binaural (MAF) correspondant à la ligne inférieure de la figure 1 est souvent approximé par l'équation non linéaire suivante :

$$SA(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4 \quad \text{(dB)}. \qquad (2.2)$$

De toute évidence, des variantes de ces données existent, car ce seuil de détection est déterminé de façon expérimentale. De surcroit, des variations de ce seuil de plus ou moins 20dB selon les individus sont considérées « normales ». Par ailleurs, ce seuil a tendance à augmenter avec l'âge, surtout en ce qui a trait aux hautes fréquences. On notera que des tonalités de durée suffisante (300ms) sont généralement utilisées pour effectuer ces mesures. Dans le cas de durées plus courtes, les chances de détection varient presque linéairement avec la durée. Il s'agit alors du phénomène d'intégration temporelle :

$$(I - I_L) \cdot t = I_L \cdot \tau = CST\ , \qquad (2.3)$$

où I est le seuil de détection d'une tonalité de durée t, I_L est l'intensité minimale représentant un stimulus (le seuil de détection d'un pulse de longue durée) et τ est la constante de temps de l'intégration temporelle dans le système auditif. Dans l'ensemble, toutes ces données peuvent s'avérer utiles au codage de l'audio, car elles définissent à la fois le niveau minimal requis pour un signal devant être codé et le niveau de bruit de quantification qui peut être introduit sans être perçu. Toutefois, en pratique, notre habileté à déceler des sons de faible amplitude est davantage une fonction du bruit ambiant. Les phénomènes de masquage s'avèrent donc d'une plus grande importance pour les codeurs audio.

2.1.3 LA PERCEPTION DE L'INTENSITÉ DES SONS

La sonie (loudness en anglais) est « l'intensité subjective » des sons, ou le caractère de la sensation auditive liée essentiellement à la pression acoustique. Notre perception de l'intensité du son varie de façon approximativement logarithmique avec l'énergie acoustique, ce qui nous permet de percevoir des sons ayant des intensités de 10^{-12} à 1 watt/m^2. Cependant, pour une même énergie acoustique, cette perception (subjective) de l'intensité varie selon la fréquence. Par conséquent, il existe au moins deux échelles de mesure pour la sonie. D'abord, le phone est la pression acoustique en décibels d'une onde plane d'incidence frontale de 1kHz qui est perçue comme ayant la même intensité que le son mesuré. La figure 1 illustre les contours de sonie équivalente en phones. Par ailleurs, cette figure illustre que nous sommes plus particulièrement sensibles à la gamme de fréquences de la parole, c'est-à-dire entre 400 et 5000 Hz.

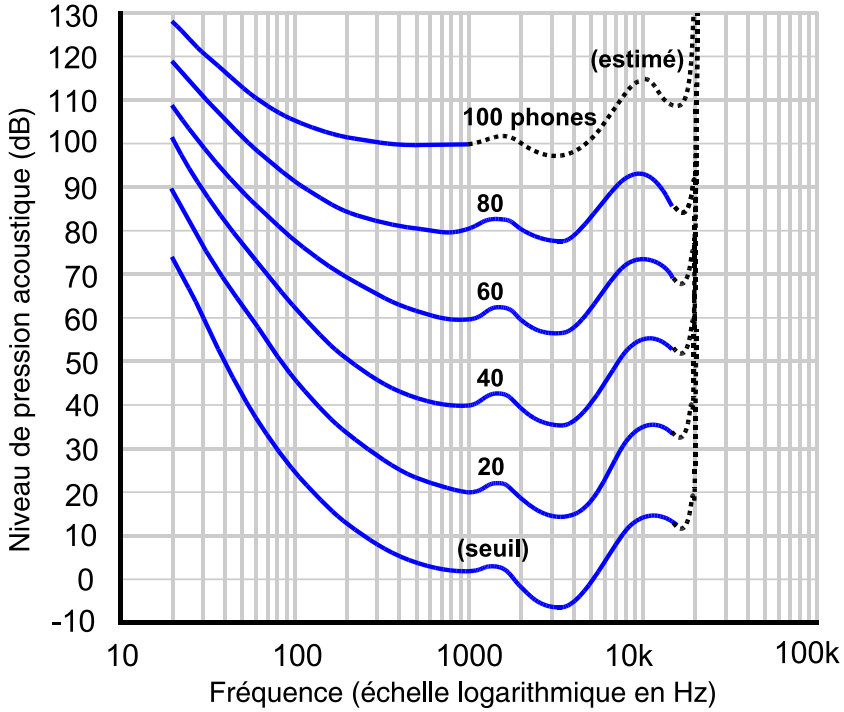

Figure 1 – Contours de sonie équivalente en phones (ISO 226: 2003)

Parallèlement, la sone est une mesure alternative de la sonie. En plus de considérer notre perception de l'intensité en fréquence, elle est graduée d'une façon à prendre en compte le fait qu'une augmentation de l'intensité acoustique de 10dB est perçue

comme une multiplication par deux de l'intensité. Par convention, 40 phones sont l'équivalent d'une sone. Donc, 50 phones sont l'équivalent de deux sones, 60 phones sont l'équivalent de quatre sones, et ainsi de suite.

Finalement, notre perception de l'intensité d'un son varie aussi en fonction de la largeur de bande de celui-ci, de sa durée, ainsi que de certains phénomènes de masquage. Les modèles pour la mesure de la sonie sont un domaine de recherche en soi, donc ils ne seront pas traités en plus grand détail dans ce document.

2.1.4 La perception de la fréquence des sons

La tonie (pitch en anglais) est le caractère subjectif d'un son par lequel on lui donne une place dans l'échelle musicale. La mesure physique (objective) qui lui est associée est la fréquence, exprimée en Hertz (Hz). Comme pour l'intensité, la relation entre notre perception subjective des fréquences et leur mesure objective (en Hz) est approximativement logarithmique.

De plus, tout comme la sonie varie en fonction de la fréquence, la relation entre la tonie et la fréquence n'est pas constante lorsque l'intensité change. En général, la tonie d'une fréquence sous 2kHz diminue lorsque l'intensité acoustique augmente et la tonie d'une fréquence supérieure à 4kHz augmente lorsque l'intensité augmente. Ces changements sont de l'ordre de 1% pour des fréquences entre 1 et 2kHz, mais peuvent atteindre jusqu'à 5% à d'autres fréquences.

De surcroit, notre perception des fréquences est différente au-delà de 5kHz. Des études démontrent que seules les tonalités en dessous de cette fréquence possèdent une place dans notre échelle musicale. Même si des différences fréquentielles sont perçues, une séquence de tonalités situées uniquement au dessus de 5kHz n'est généralement pas perçue comme étant une mélodie. Il est aussi très difficile de transposer une séquence de tonalités (une mélodie) à de telles fréquences. Aussi, les gens qui sont en mesure de nommer des notes de façon absolue (sans référence) ont peu de succès à des fréquences supérieures à 4-5kHz. Ces informations semblent expliquer en partie le succès des techniques d'extension de bande paramétrique dans les codeurs perceptuels modernes à bas débit [8, 10].

2.1.5 LES VARIATIONS SONORES MINIMALES PERCEPTIBLES

Notre système auditif possède une certaine résolution en intensité, en temps (durée), en fréquence et en phase. Des variations sonores en dessous de ces seuils ne sont pas perceptibles. Cette section résume ces différentes résolutions, connues sous le nom de seuils différentiels. Plus précisément, un seuil différentiel est défini comme étant la variation minimale requise par un signal pour être perçue. Donc, on s'intéresse aux changements minimaux perceptibles au niveau de la fréquence, de la durée, ou plus particulièrement de l'intensité des signaux sonores.

En ce qui a trait à l'amplitude, on dénote souvent un décibel comme étant la résolution du système auditif. Toutefois, notre habileté à détecter une variation d'intensité dépend de la nature du son (tonalité ou bruit), de sa fréquence et de son intensité nominale. Dans des conditions idéales, le seuil de détection de la variation d'amplitude d'un bruit blanc (20Hz à 20kHz) d'au moins 30dB est pratiquement constant à 0.7dB. Lorsqu'il s'agit d'une tonalité pure, ce seuil a plutôt tendance à diminuer lorsque l'intensité augmente. Cette dépendance sur la nature du son est en partie liée à la technique de mesure, car pour éviter des « clics » lors des tests, on emploie des signaux modulés en amplitude. Or, un sinus modulé en amplitude fait apparaitre deux lobes secondaires. Parallèlement, si la fréquence nominale et l'amplitude (>30dB) sont constantes, le seuil de détection diminue lorsque la largeur de bande d'un bruit augmente. Il va de même lorsque la durée augmente.

Le système auditif est aussi limité par une certaine résolution fréquentielle. Le seuil de détection minimale audible pour une variation en fréquence dépend de la fréquence nominale. De plus, les valeurs mesurées dépendent de la technique utilisée. D'une part, si le test consiste à présenter d'abord une tonalité de référence de fréquence f (en Hz) et ensuite une deuxième de fréquence $f+\Delta f$, la résolution du système auditif peut être approximée par la relation suivante [11, 14] :

$$\log\left(\Delta f\right) \approx 0.02637\sqrt{f} - 0.5139\,. \tag{2.4}$$

Cette technique a l'avantage d'être relativement simple, mais elle a une forte dépendance sur la mémoire auditive. D'autre part, si le test consiste à déterminer si

une tonalité modulée en fréquence est identique à la même tonalité sans modulation, la discrimination en fréquence Δf peut être approximée comme étant constante à 3.6Hz jusqu'à f = 514.3Hz pour augmenter linéairement à raison de 0.007f Hz par la suite. Cela signifie que la résolution fréquentielle du système auditif est excellente; autour de 0.7% pour les fréquences au dessus de 500Hz. Toutefois, ce pouvoir de résolution fréquentielle diminue progressivement lorsque la vitesse de modulation augmente (passé environ 8Hz). Cette technique de mesure semble mieux isoler le facteur de mémoire auditive par rapport à la première technique, où deux tonalités de fréquences différentes sont tout simplement comparées.

Le pouvoir de résolution du système auditif est aussi influencé par le masquage. Cette résolution diminue rapidement lorsqu'un bruit de fond est ajouté. Par exemple, on peut penser à un exemple simple où le masquage fréquentiel « cache » les lobes secondaires introduits par la modulation en amplitude d'une tonalité pure.

2.1.6 LE MASQUAGE ET LA SÉLECTIVITÉ EN FRÉQUENCE

Cette partie discute de la sélectivité en fréquence du système auditif. Il s'agit de notre habileté à détecter un son en présence d'un autre son. La sélectivité en fréquence du système auditif est normalement démontrée et mesurée par l'entremise du phénomène de masquage fréquentiel. Ainsi, le masquage (fréquentiel et temporel) est défini comme suit :

1) Le processus selon lequel le seuil d'audibilité pour un son (masqué) est augmenté par la présence d'un autre son (masquant).

2) La quantité d'augmentation du seuil d'un son (masqué) par l'autre son (masquant). Cette quantité est normalement exprimée en décibels (dB).

En général, on remarque qu'un son est plus facilement masqué lorsque le son masquant possède des composantes fréquentielles proches ou identiques à celui-ci. Cela se produit car un signal excite, à l'intérieur du système auditif, une gamme de fréquences plus large que celle contenue dans celui-ci. En guise d'illustration, la figure 2 présente le patron d'excitation d'une onde de 1kHz à diverses intensités.

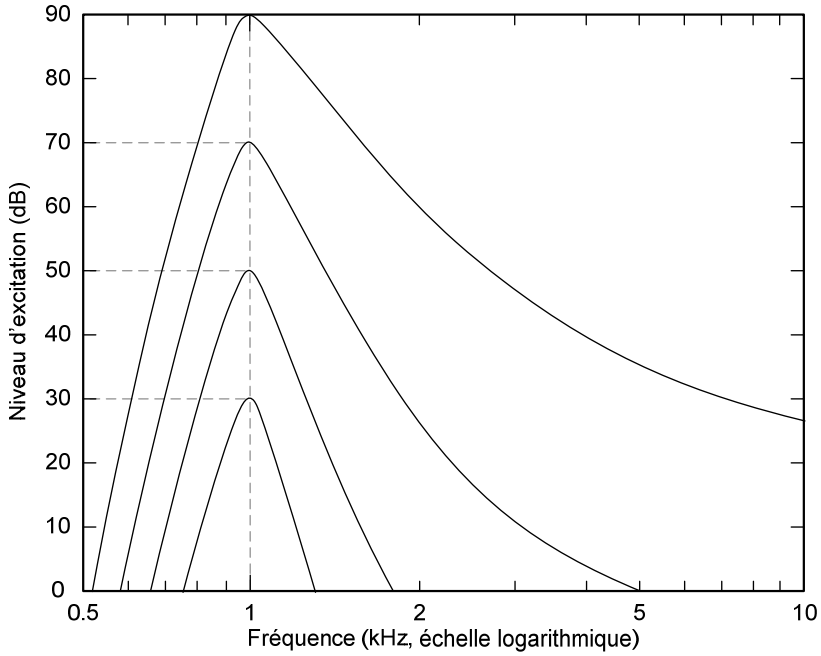

Figure 2 – Patron d'excitation d'une tonalité de 1kHz à 30, 50, 70 et 90 dB

La figure 2 permet de constater que le seuil de détection de l'audition risque fort bien d'augmenter par rapport au seuil minimal lorsqu'un autre signal est présent, car l'excitation produite par cette onde affecte plusieurs autres fréquences. Il est aussi possible de constater que ce phénomène est non linéaire, car il dépend de l'intensité.

En effet, c'est ce qui se produit, et de nombreux codeurs perceptuels calculent ces courbes de masquage fréquentiel pour déterminer la quantité de bruit de quantification permis. Dans bien des cas, cela correspond au nombre de « bits » à employer dans chaque sous-bande pour la quantification de l'information spectrale.

Les patrons d'excitation permettent aussi de déterminer l'allure des « filtres auditifs » qui peuvent servir à la modélisation du système auditif, en assumant qu'ils sont suivis de simples détecteurs d'énergie. Ces filtres auditifs peuvent être vus comme un banc de filtres passe-bandes qui se superposent largement, sont de largeurs croissantes en fréquence et ont une réponse en fréquence non linéaire en fonction de l'intensité. Un tel réseau de filtres suivi de détecteurs d'énergie peut effectivement servir à modéliser les effets de masquage fréquentiel étudiés dans cette section. Une simplification de ces filtres auditifs est présentée à la section suivante qui couvre l'étude de ce qui est connu sous le nom de « bandes critiques ».

2.1.7 LES BANDES CRITIQUES

Les bandes critiques réfèrent normalement à une simplification du phénomène de masquage fréquentiel, où l'on suppose que le système auditif est composé de filtres rectangulaires. Cette version « approximée » est tout de même reliée à la sélectivité en fréquence du système auditif et elle est d'une grande importance pour les codeurs audio perceptuels. Contrairement au filtre auditif, les bandes critiques sont considérées comme étant un banc de filtres passe-bandes rectangulaires (idéaux) qui ne se recouvrent pas. Une des techniques utilisées pour évaluer la largeur de ces bandes utilise la définition suivante : un bruit à bande étroite d'un niveau donné est toujours perçu avec la même intensité, tant que la bande passante de celui-ci ne dépasse pas la largeur d'une bande critique. Dans ce cas, l'unité de mesure de cette échelle est le Bark. Elle est graduée de 1 à 24 et elle a été publiée sous forme de tableau (Zwicker) avec comme extrémités (en Hz) : [0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500] et comme centres respectifs : [50, 150, 250, 350, 450, 570, 700, 840, 1000, 1170, 1370, 1600, 1850, 2150, 2500, 2900, 3400, 4000, 4800, 5800, 7000, 8500, 10500, 13500]. Toutefois, il peut être utile de transformer des fréquences de l'échelle Hertz à l'échelle Bark. Dans ce cas, plusieurs approximations empiriques existent, telles que :

$$z = 13\arctan\left(\frac{19f}{25000}\right) + 3.5\arctan\left(\left(\frac{f}{7500}\right)^2\right) \quad \text{(Bark)}, \tag{2.5}$$

et

$$z = \frac{26.81f}{1960+f} - 0.53 \quad \text{(Bark)}, \tag{2.6}$$

qui peut d'ailleurs être « améliorée » aux extrémités, ce qui donne la relation :

$$z = \begin{cases} 0.01f & \text{pour } f < 421.613 \\ 1.22\zeta - 5.0686 & \text{pour } f > 6542.85 \\ \zeta - 0.53 & \text{ailleurs} \end{cases} \quad \Bigg| \quad \zeta = \frac{26.81f}{1960+f} \quad \text{(Bark)}, \tag{2.7}$$

où les valeurs 421.613 et 6542.85 sont obtenues en résolvant pour f les équations $0.01f = \zeta - 0.53$ et $1.22\zeta - 5.0686 = \zeta - 0.53$ respectivement. De plus, l'équation (2.7) est facilement réversible, ce qui permet aussi de passer de l'échelle Bark à des Hertz.

Par ailleurs, les largeurs des bandes critiques, en Hertz, peuvent aussi être approximées par l'équation suivante :

$$BW_c(f) = 25 + 75\left[1 + 1.4\left(f/1000\right)^2\right]^{0.69} \quad \text{(Hz)}. \tag{2.8}$$

De façon moins précise, la « règle du pouce » dicte une largeur de 100Hz pour les fréquences sous 500Hz et une largeur de f/5 pour les fréquences plus élevées.

En dépit du nombre de publications en accord avec ces données, de nombreuses autres études ont confirmé des valeurs bien différentes de celles présentées jusqu'ici. Notamment, parce que la technique de mesure employée est très différente. Donc, pour bien faire la différence, le terme ERB (equivalent rectangular bandwidth) sera employé au lieu de « bandes critiques ». L'ERB est basé sur un test conçu par Patterson (1976) qui consiste à mesurer le seuil de masquage d'une tonalité noyée dans un bruit situé de chaque côté de celle-ci, en variant la largeur de la zone sans bruit, tel que présenté à la figure 3 [11].

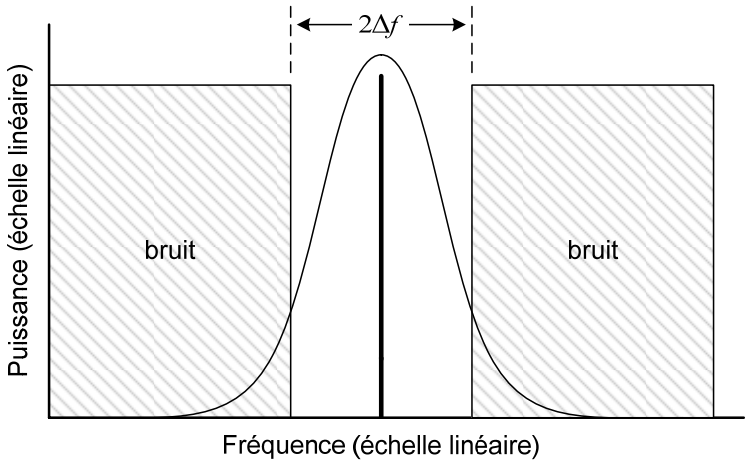

Figure 3 – Mesure du ERB par la technique de Patterson (1976)

Mesurée de cette façon, la largeur d'un ERB est approximée par l'équation suivante :

$$ERB(f) = 0.10794f + 24.7 \quad \text{(Hertz)}. \tag{2.9}$$

D'ailleurs, une échelle fréquentielle peut être substituée par une échelle « ERB », notée ici par « ERB_N », en employant l'approximation suivante :

$$ERB_N(f) = 21.4\log_{10}(0.00437f + 1). \tag{2.10}$$

Il est à noter que la largeur d'un ERB est toujours moins large que les bandes critiques classiques. Cela dit, ce type d'échelle perceptuelle permet d'améliorer les performances subjectives des codeurs audio, qui ont avantage à analyser les signaux audio regroupés de cette façon au lieu d'employer une échelle fréquentielle linéaire ou purement logarithmique.

2.1.8 LA RÉSOLUTION TEMPORELLE

Jusqu'ici, seulement l'amplitude et le contenu fréquentiel des signaux audio ont été considérés. Cependant, la résolution temporelle du système auditif est d'une égale importance. Une grande part de l'information contenue dans la parole et la musique est véhiculée par des variations auditives dans le temps, et non par le contenu des périodes stationnaires. De nombreuses techniques existent pour mesurer la résolution temporelle du système auditif, dont la détection d'intervalles de silence dans un bruit à large bande et la détection de doubles-clics à séquence inversée. Dans les deux cas, les sons doivent conserver le même spectre d'amplitude (par exemple, inverser un son dans le temps ne change pas son spectre). D'une façon générale, ces tests permettent de déterminer que la résolution temporelle du système auditif est de l'ordre de deux à trois millisecondes (2-3ms). Cette résolution peut diminuer jusqu'à cinq millisecondes (5ms) lors des expériences effectuées avec des tonalités pures.

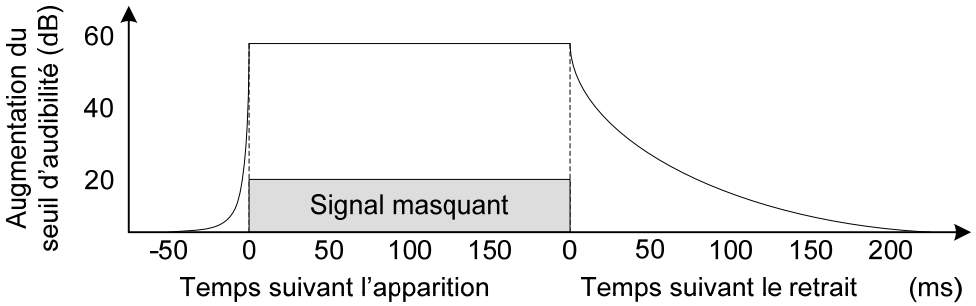

Figure 4 – Phénomène de masquage temporel

Parallèlement, seul le masquage fréquentiel a été considéré jusqu'ici. Cependant, les phénomènes de masquage peuvent aussi se produire dans le domaine temporel. Même après avoir retiré un signal masquant, il peut continuer à influer sur le seuil de l'audition durant une période pouvant atteindre jusqu'à 200ms. La durée et le niveau de ce masquage dépendent de l'intensité et de la durée du signal masquant. De surcroit, bien que cet effet soit encore mal compris, un signal masquant peut aussi avoir une influence sur le seuil de l'audition quelques millisecondes avant l'apparition de celui-ci. La figure 4 illustre le phénomène de masquage temporel autour d'un signal masquant. Toutefois, cet effet s'atténue pour devenir pratiquement inexistant lorsque l'entrainement des auditeurs augmente, ce qui laisse croire qu'il est davantage lié à une confusion avec le signal masquant qu'à un véritable phénomène de masquage. Ce phénomène controversé est quand même fréquemment mis de l'avant pour justifier l'efficacité de certaines techniques appliquées dans le codage audio, comme la réduction dynamique des tailles de blocs et la mise en forme du bruit de quantification dans le temps (TNS, ou temporal noise shaping).

2.2 Psychoacoustique binaurale

Une grande part des travaux sur le codage audio perceptuel ne prend pas en compte les études plus ou moins récentes en psychoacoustique binaurale. Toutefois, certaines de ces lacunes ont été rectifiées durant la période de cette recherche. Cela dit, cette section présente une revue de la littérature en psychoacoustique binaurale pertinente au codage perceptuel. De nombreux détails additionnels se retrouvent dans certains ouvrages classiques tels que [13, 15-18].

2.2.1 LA LOCALISATION D'UNE SOURCE

D'abord, il ne faut pas présumer qu'il est suffisant de préserver l'emplacement des sources d'un signal pour y préserver l'ambiance stéréophonique. L'écho et la réverbération, qui apportent une qualité « spatiale » à un évènement sonore, ne sont pas des facteurs déterminés uniquement par la position des sources. Toutefois, la préservation du lieu des sources sonores, en plus de l'ambiance, est un aspect primordial pour tout codeur audio perceptuel (stéréo ou multicanal).

Le système auditif, à partir de deux oreilles seulement, peut localiser des sons dans l'espace tridimensionnel, c'est-à-dire l'angle d'arrivée horizontale (gauche versus droit), l'angle d'arrivée verticale (haut versus bas) et la distance. De surcroit, même si la localisation est plus précise à l'aide de deux oreilles, elle peut aussi se baser sur l'information présente à l'entrée d'une seule.

Toutefois, la plupart des expériences sur ce sujet sont réalisées avec des écouteurs, où l'image est normalement perçue comme étant à l'intérieur de la tête. Ces expériences permettent de contrôler indépendamment certaines variables comme le retard et l'intensité entre les canaux. En général, la source est alors perçue à l'intérieur de la tête, et la tâche de l'auditeur consiste à projeter la position de cette source sur une ligne unidimensionnelle entre ses deux oreilles. Dans ce cas, le terme latéralisation est employé au lieu de localisation. Par exemple, une source sonore présentée avec une différence d'intensité de 30dB entre les deux oreilles est perçue comme étant positionnée complètement à l'extrémité de la tête dont l'intensité est la plus élevée, et cette position varie presque linéairement lorsque la différence d'intensité (en dB) est réduite. Parallèlement, une source sonore présentée avec un retard d'une milliseconde à une oreille par rapport à l'autre est perçue comme étant du côté de la tête dont le signal a été présenté en premier, et encore une fois, cette position varie presque linéairement lorsque ce délai est réduit.

Suite à ces précisions, voici les principaux indicateurs utilisés pour la localisation (dans les trois dimensions) d'une source sonore :

1) Le temps d'arrivée inter-canal, surtout efficace sous 750 Hz.
2) La différence d'amplitude inter-canal, surtout au dessus de 1000 Hz.
3) Le pavillon de l'oreille, qui agit à titre de filtre en fonction de la direction.
4) Si permis, de légers mouvements de tête qui font varier tous les indicateurs.
5) L'intensité du son direct par rapport à celle des réverbérations (distance).
6) L'atténuation des hautes fréquences (distance).
7) L'effet Doppler (sources sonores mobiles).

Les effets numérotés de 1 à 4 sont les indicateurs les plus prédominants pour déterminer l'angle d'incidence d'une source sonore. Les effets 5 et 6 sont davantage

des indices pour juger de la distance d'une source sonore. Finalement, l'effet Doppler concerne les sources sonores mobiles, soit un aspect négligeable lors de l'écoute d'haut-parleurs. Les quatre premiers indicateurs seront précisés dans ce qui suit.

Le temps d'arrivée inter-canal peut être estimé à partir d'un modèle de tête sphérique, comme illustré à la figure 5 et précisé par l'équation (2.11),

$$ITD(\theta) = \frac{r}{c}\left[\sin(\theta) + \theta\right]. \tag{2.11}$$

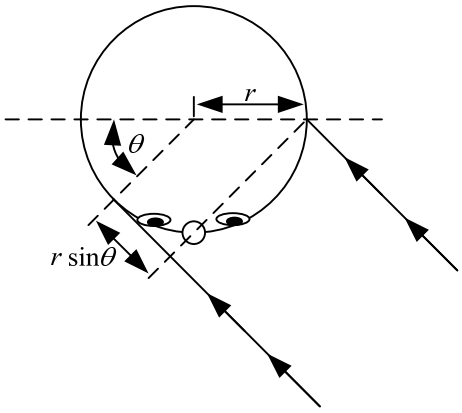

Figure 5 – Estimation du délai inter-canal avec un modèle sphérique

Dans la figure 5, r représente le rayon de la tête, c la vitesse du son, et θ l'angle d'arrivée (horizontale) de l'onde sonore. Il existe aussi des formules plus complètes qui tiennent compte de l'élévation verticale (azimut), de la forme ellipsoïdale de la tête et de la position des oreilles. Un tel modèle réduit l'effet du cône de confusion, car le ITD n'y est pas constant, contrairement au modèle sphérique [19]. L'effet des légers mouvements de tête est alors très important; il permet d'éviter le problème du cône de confusion et facilite même la localisation d'une source à l'aide d'une seule oreille au point de rivaliser avec la localisation à l'aide des deux oreilles [11].

Le ITD demeure un indice important pour la localisation. Toutefois, autour de 1.5 kHz, la longueur d'une onde acoustique sinusoïdale avoisine la distance entre les deux oreilles (par exemple : $345\ \text{m/s} \div 1.5\ \text{kHz} = 23\ \text{cm}$). Donc, pour les sinusoïdes de fréquence supérieure à 750 Hz, la relation de phase entre les deux oreilles devient ambigüe et le système auditif ne peut assumer une valeur unique de ITD. Cependant,

il est erroné d'affirmer que les signaux au dessus de cette fréquence sont inutiles à la localisation, car dans le cas d'un son composé, l'enveloppe temporelle peut varier suffisamment lentement pour être détectée aisément [11]. De plus, un signal de fréquence trop élevée pour être utile à la localisation ne signifie pas pour autant qu'il est impossible de percevoir la présence d'une différence de phase. L'allure temporelle des déclenchements d'un neurone contient de l'information sur le stimulus; des pics nerveux ont tendance à survenir à une phase particulière de l'onde (phase locking). Cet effet se dissipe au-delà de 4 à 5 kHz [11], soit à des fréquences supérieures à 1.5 kHz. En ce qui a trait à la sensibilité aux variations de ITD, elle est la plus élevée pour une source située droit devant l'auditeur. La variation d'angle la plus faible pouvant être décelée est alors de 0.8° à 3.3° en fonction des caractéristiques de celle-ci [15]. Le ITD maximal décelable chez l'humain est 800 µs [17]. Une dernière remarque concerne l'utilisation de haut-parleurs. Il est intéressant de noter qu'une différence d'amplitude (ILD) présentée par ceux-ci crée un déphasage des signaux présents au niveau des oreilles de l'auditeur, soit un ITD. Toutefois, ce cas est artificiel, car dans la nature, il est difficile d'imaginer deux sources émettant exactement le même signal à des intensités différentes.

Un deuxième paramètre permettant au système auditif de localiser une source est la différence de l'intensité sonore entre les deux oreilles, soit le ILD. Une telle différence se produit dans la nature, car l'intensité d'une onde acoustique diminue avec le carré de la distance. Toutefois, il s'agit d'une différence minime lorsqu'il s'agit de la distance entre les oreilles. Le facteur prédominant est davantage la tête elle-même, qui agit comme une barrière au son, particulièrement à des fréquences supérieures à 1 kHz. Les ILD sont donc pertinents aux fréquences plus élevées, où la tête oppose un meilleur écran et où le ITD est inefficace. La sensibilité aux ILD peut toutefois s'étendre jusqu'à des fréquences aussi basses que 200 Hz. Elle est d'environ 1 dB pour une onde à 1 kHz et de 0.5 dB entre 2 et 10 kHz [16]. Finalement, ce n'est pas parce qu'un ILD est ignoré pour la localisation qu'il ne sera pas perçu. Il faut plutôt comprendre qu'une différence d'intensité sonore (ILD) en basses fréquences (sous 200 Hz) est peu fréquente dans la nature, car l'atténuation de cette gamme de fréquences par la présence de la tête est négligeable.

Lorsque la position d'une source sonore est ambigüe, deux autres facteurs peuvent intervenir. D'abord, s'il s'agit d'une source dont les caractéristiques sont familières à l'auditeur, le pavillon de l'oreille (ainsi que la tête et tout le corps) forme un filtre en fonction de la position de la source. Il est donc possible pour cet auditeur d'estimer une position approximative. La mesure de ces « filtres » est connue dans la littérature sous le nom de HRTF (head related transfer functions). Un HRTF combine à la fois le ITD, le ILD et les réflexions du signal sur le corps et la modification du signal par le pavillon de l'oreille. Finalement, de légers mouvements de tête, souvent inconscients, font varier tous les facteurs dynamiquement et aident ainsi à résoudre l'ambigüité.

2.2.2 L'EFFET DE PRÉCÉDENCE

Dans des conditions d'écoute normales, le son d'une source atteint nos oreilles en passant par divers parcours. Bien qu'une partie du son arrive par un trajet direct, il demeure une autre grande partie de celui-ci qui atteint les oreilles seulement après une ou plusieurs réflexions. Ces « échos » ont cependant peu d'influence sur notre capacité à juger de la direction d'une source sonore. Ainsi, il est possible de localiser un haut-parleur dans une pièce réverbérante, même dans des cas où l'énergie du son réfléchi est supérieure à l'énergie du son direct [11]. L'effet d'un ITD supérieur aux valeurs normales prédites à partir de la distance entre les deux oreilles est nommé l'effet de précédence, l'effet de Haas ou « la loi du premier front d'onde ».

Moore résume en onze points les divers résultats sur ce sujet [11].

1) Une séquence de deux sons qui parviennent aux oreilles dans un intervalle très court est perçue comme un seul. Le délai maximum varie entre 5 ms pour des clics jusqu'à 40 ms pour des signaux complexes comme une attaque dans la musique ou la parole. Il s'agit d'un effet de fusion ou de suppression d'écho.
2) Si deux sons successifs sont perçus comme étant fusionnés, la direction perçue est déterminée en grande partie par la direction d'arrivée de la première onde.
3) Cet effet ne se produit qu'avec des sons discontinus ou transitoires.
4) Si le ITD augmente dynamiquement, la direction perçue d'une onde fusionnée se déplace jusqu'à un maximum de 7°, de la première onde vers la deuxième.

5) Si l'intervalle est de moins de 1 ms, une direction intermédiaire est perçue.
6) Si le deuxième son est plus intense (10-15dB), l'effet de précédence est annulé.
7) L'effet de précédence peut se produire même si la deuxième onde diffère en contenu spectral (tout en conservant une enveloppe temporelle semblable).
8) L'effet de précédence peut se produire sans qu'il n'y ait de disparité binaurale.
9) L'effet peut prendre du temps à se bâtir. Avec un seul clic et son seul écho de 8 ms, l'effet ne se produit pas. Si cette séquence est répétée quatre fois par seconde, l'effet se produit après quelques clics.
10) L'effet est interrompu brièvement si la séquence change de façon considérable.
11) L'effet de précédence n'implique pas la suppression complète des échos. Il est toujours aisé de faire la différence entre un son avec et sans écho. Cela signifie plutôt que les échos ne sont pas perçus comme des évènements distincts et que la direction d'arrivée des échos est partiellement ou complètement perdue.

2.2.3 LE DÉMASQUAGE BINAURAL

Le seuil de détection d'un signal en présence d'un signal masquant peut être diminué (le signal masqué devient plus facile à détecter) lorsque ces signaux sont présentés d'une certaine façon aux deux oreilles. Les expériences démontrent que ce seuil est identique pour un signal présenté à une oreille ou le même signal présenté de façon identique aux deux oreilles. Toutefois, si on permet des signaux différents aux deux oreilles, le seuil de masquage diminue [13]. Cela signifie que les modèles perceptuels monauraux ne s'étendent pas systématiquement à plusieurs canaux. En particulier, même si un signal est masqué individuellement dans le canal gauche et dans le canal droit, cela ne signifie pas qu'il sera toujours masqué lorsque les deux canaux sont présentés simultanément avec un casque d'écoute [20]. Cette différence dans le seuil de masquage s'appelle le BMLD (binaural masking level difference) ou plus simplement le MLD. Ces effets de masquage ont reçu peu d'attention dans les modèles de masquage [21]. Il existe tout de même une grande quantité de données expérimentales sur le sujet. De plus, il est reconnu que la séparation spatiale d'un signal masqué du signal masquant peut réduire le seuil de masquage jusqu'à 20 dB pour un masquage simultané [22] et jusqu'à 5 dB pour un pré ou post masquage de

moins de 40 ms. Le MLD est l'une des justifications du codage stéréo mid/side (M/S) car il peut tenir compte de cet effet. Dans le cas de tonalités masquées par du bruit à large bande, cet effet est davantage marqué en basses fréquences, sous 1.5kHz, où la localisation est possible par la différence de phase. Le tableau 1 (tableau 7.1 dans [11]) présente des valeurs de MLD pour des signaux masqués en basses fréquences. Pour interpréter ce tableau, il faut savoir que « N » signifie un bruit masquant et que « S » est la tonalité masquée, « 0 » est un signal en phase aux deux oreilles, « π » est un signal inversé à une oreille (phase de 180°), « u » est un signal décorrélé aux deux oreilles et « m » est un signal présenté à une seule oreille. Selon cette nomenclature, N_0S_0 signifie un bruit masquant et une tonalité masquée présentés de façon identique aux deux oreilles (c'est-à-dire le MLD de référence), et N_0S_π signifie un bruit masquant identique (en phase) aux deux oreilles présenté avec une tonalité dont la phase est inversée à l'une des oreilles.

Condition interaural	MLD (dB)
N_0S_0	0
N_uS_π	3
N_uS_0	4
$N_\pi S_m$	6
N_0S_m	9
$N_\pi S_0$	13
N_0S_π	15

Tableau 1 – Valeurs de MLD pour plusieurs cas typiques

Étant donné que les valeurs de MLD peuvent être aussi élevées, il pourrait être avantageux pour un codec stéréo ou multicanal de mettre en œuvre un modèle psychoacoustique qui tient compte de ce phénomène.

2.2.4 LES MODÈLES DE PERCEPTION BINAURAUX

Il existe plusieurs modèles de perception binaurale, allant du plus simple « détecteur de coïncidences » [23] jusqu'aux modèles plus complets présentés par [15] (et précisés dans [22]). D'abord, Jeffress a présenté un modèle simple pour les ITD; une

série d'éléments de retards (ΔT) et de détecteurs de coïncidences de l'activité neurale entre les deux oreilles (CC) permettent de déterminer la direction d'arrivée latérale d'un signal (figure 6). Par exemple, un son très proche de l'oreille gauche activera particulièrement la sortie (+2), là où la corrélation est la plus élevée, ce qui indique au reste du cerveau que le son est situé à la gauche de la tête.

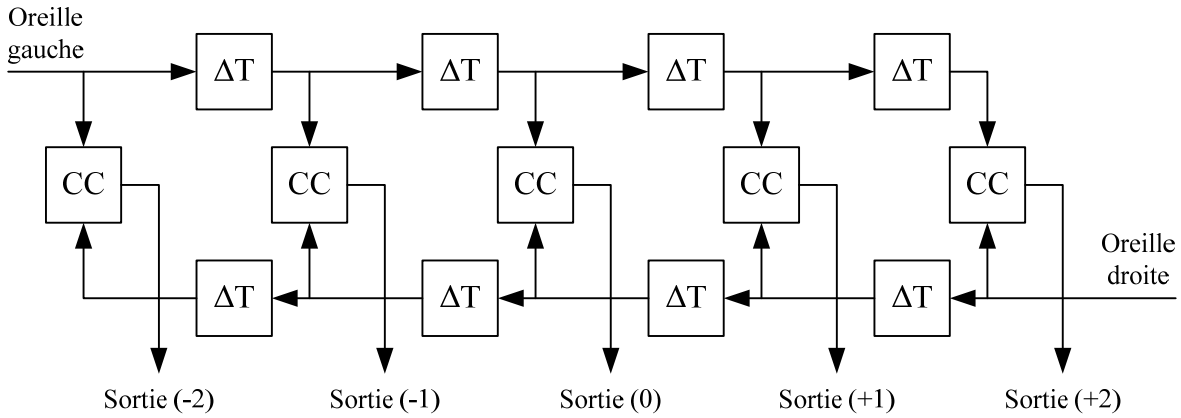

Figure 6 – Modèle du détecteur de coïncidences de Jeffress

Toutefois, il s'agit d'un modèle relativement incomplet. D'après les connaissances actuelles, six éléments sont requis minimalement dans un modèle binaural [15] :

1) Des filtres qui simulent la fonction de transfert de l'oreille externe.
2) Des filtres qui simulent l'effet de l'oreille moyenne (middle ear).
3) Des éléments qui simulent les fonctions de l'oreille interne comme celles de la cochlée; en particulier sa sélectivité en fréquence et la transformation des signaux analogiques en probabilités de déclenchement de neurones.
4) Un élément qui simule l'évaluation des ITD et identifie les signaux cohérents.
5) Un élément qui analyse et tient compte des ILD.
6) Un élément qui analyse et interprète la sortie des items (4) et (5).

Étonnamment, peu de modèles incluent les deux premiers éléments. Par contre, tous les modèles pertinents mettent en œuvre le troisième élément; soit en modélisant les pics nerveux individuels ou tout simplement la probabilité de leur déclenchement. Les éléments 4 à 6 doivent tenir compte des temps de réaction lents du système auditif. Une description plus détaillée de tous ces éléments est présentée dans [15].

3. MODÈLES DE CODAGE STÉRÉO ET MULTICANAL

Cette section présente l'état de l'art des méthodes de codage de plusieurs canaux. Les techniques présentées sont basées sur l'exploitation des redondances entre les canaux, comme le matriçage et la prédiction linéaire, ou sur l'exploitation de la psychoacoustique binaurale, y compris les modèles de perception spatiaux.

3.1 Exploitation des redondances

Cette sous-section couvre les techniques de codage stéréo et multicanal basées sur l'exploitation de la redondance entre les canaux. Deux approches principales sont approfondies; les algorithmes basés sur le matriçage et ceux basés sur la prédiction.

3.1.1 LE MATRIÇAGE ET LE CODAGE MID/SIDE (M/S)

D'abord, le matriçage est une opération qui consiste à transformer un certain nombre de canaux discrets en un certain nombre de canaux cibles. Il peut s'agir de la simple somme de deux canaux, mais le choix d'un nombre de canaux cibles inférieur à celui de départ ne permet pas d'appliquer l'opération inverse sans perte. L'opération est réversible dans les cas respectant la logique suivante :

$$\overline{\mathbf{C}} = \mathbf{A} \times \mathbf{C}, \quad \tilde{\mathbf{C}} = \mathbf{B} \times \overline{\mathbf{C}} \quad \Rightarrow \quad \tilde{\mathbf{C}} = \mathbf{C} \text{ si } \mathbf{A} \times \mathbf{B} = \mathbf{I}. \tag{3.1}$$

Dans le cas spécifique du codage audio multicanal, cette contrainte se traduit par la transformation de deux ou de plusieurs canaux en un même nombre de canaux ayant toutefois des caractéristiques différentes. Ce matriçage est sans perte, car l'objectif principal de cette opération est de réduire la corrélation entre les canaux, augmentant du coup l'efficacité du codage. Un codage perceptuel des canaux mixés peut alors être effectué [24]. Toutefois, sans précaution, le codage suivi de l'opération de matriçage inverse peut générer un bruit de quantification dans un canal autre que celui de la source sonore, où il risque d'être démasqué [25].

Pour sa part, le codage M/S est un cas spécifique de matriçage avec deux canaux :

$$\begin{bmatrix} \mathbf{M} \\ \mathbf{S} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{R} \\ \mathbf{L} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{R} \\ \mathbf{L} \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{M} \\ \mathbf{S} \end{bmatrix}, \quad \text{où } \mathbf{A} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{3.2}$$

où $\mathbf{R}$ et $\mathbf{L}$ sont les canaux de départ, et où $\mathbf{M}$ et $\mathbf{S}$ sont les canaux intermédiaires. Ce cas respecte la logique formulée en (3.1) et $\mathbf{A}\times\mathbf{A}^{\mathrm{T}} = \mathbf{I}$. Dans le cas extrême où les deux canaux $\mathbf{R}$ et $\mathbf{L}$ sont identiques, le canal $\mathbf{M}$ est le signal compatible mono et le signal $\mathbf{S}$ est nul, donc trivial à encoder. Cette technique est utilisée depuis les années 1960 en radio FM pour préserver la compatibilité mono et a été introduite en codage de l'audio au début des années 1990 [26, 27]. Plus récemment, il a été démontré que la technique de matriçage M/S pouvait aider à réduire l'effet du démasquage binaural (MLD) [28]. Le codage M/S est de toute évidence très utile pour éviter les artéfacts lors du codage de signaux presque mono. Toutefois, appliquée systématiquement, cette approche peut entrainer le démasquage du bruit de quantification. Un codeur doit donc choisir d'employer le codage M/S ou non selon les caractéristiques du signal à encoder [29]. De nombreuses améliorations ont donc été apportées depuis les articles originaux de Wall et Johnston pour éviter les problèmes intrinsèques au codage M/S. Par exemple, l'approche M/S est maintenant appliquée de façon facultative dans chaque bande critique, et les critères pour cette décision ont été perfectionnés [20]. Une revue de la littérature fait même ressortir des améliorations relativement récentes à cette technique de codage [30]. Le codage M/S peut aussi être combiné à la technique IS (intensité stéréo) qui sera présentée à la section 3.2.1.

Certains auteurs ont appliqué la transformée de Karhunen-Loève (KLT ou PCA) au codage audio multicanal, définie par les vecteurs propres de la matrice de covariance du signal audio. La propriété de concentration d'énergie de la KLT est employée pour transformer les canaux de départ en canaux complètement décorrélés de puissances décroissantes. Dans ces cas, la matrice $\mathbf{A}$ de l'équation (3.1) est la matrice de la KLT calculée pour une trame à encoder, et elle doit être transmise au décodeur pour la réalisation du matriçage inverse. Il a été démontré que cette technique est efficace lorsqu'il y a un grand nombre de canaux représentant un champ acoustique réel [31]. Une amélioration de cette approche consiste en une version adaptative quantifiée vectoriellement [32]. Une autre variante, conçue spécifiquement pour le codage audio sans perte, emploie une INT-DCT (DCT entière) pour décorréler les canaux [33]. En général, les techniques qui calculent la matrice $\mathbf{A}$ de façon dynamique ne sont efficaces qu'avec plusieurs canaux (autour de 10).

3.1.2 LA PRÉDICTION INTER-CANAL

Intuitivement, on peut imaginer qu'il y a une forte dépendance entre plusieurs canaux audio. Toutefois, cela ne signifie pas pour autant que la corrélation calculée entre les canaux sera élevée. Une étude portant sur les statistiques temporelles de signaux stéréo a démontré que la relation linéaire entre les canaux est très faible (sans tenir compte du passé du signal) [34]. Cependant, dans le domaine fréquentiel, les mêmes auteurs ont démontré que les amplitudes (logarithmiques) des coefficients dans ce domaine sont corrélées (mais pas ceux de la phase) [35]. Dans le cas de cinq canaux audio enregistrés de façon à reproduire un champ sonore réel, la corrélation temporelle est élevée [36], ce qui est le cas d'une grande partie des signaux enregistrés de cette façon. De nombreux algorithmes tentent donc d'exploiter la redondance entre plusieurs canaux. La dépendance entre plusieurs canaux peut être exploitée par une opération de matriçage (pour exploiter un certain degré de corrélation entre les canaux) ou par une prédiction s'effectuant entre les canaux, pour tenir compte des cas où il y a une différence de phase ou un retard temporel. Fuchs [37] présente la technique de base où un prédicteur FIR à court terme est inséré entre deux canaux. Cependant, même s'il existe une corrélation entre les signaux dans le domaine temporel, une prédiction linéaire entre les canaux ne procure pas nécessairement un gain dans le cas du codage perceptuel d'audio pleine bande [36]. Toutefois, la même étude laisse une porte ouverte dans le cas de la prédiction dans la bande basse. Cet aspect est d'autant plus important que les auteurs, qui ont attribué le problème à la performance dans la bande haute, n'ont pas poursuivi l'expérience en mesurant la performance en bande basse combinée à une technique de réplication de bande, ou en large bande avec une prédiction linéaire déformée (warped) [38]. En effet, Härmä propose une technique de codage stéréo basée sur la prédiction linéaire déformée (warped) et la mise en forme temporelle du bruit de quantification (TNS), mais sans conclure avec une solution complète (quantifiée). Dans un article plus récent [39], l'idée de prédiction linéaire (IIR) a été poursuivie et d'autres solutions ont été apportées, sans toutefois être appliquées à la prédiction de plus d'un canal. Dans un cas plus spécifique, la prédiction stéréo a été appliquée avec succès dans un codeur hiérarchique au niveau des deux étages [40]. En

général, la revue de la littérature permet de constater que la prédiction inter-canal est efficace pour le codage à haut débit et sans perte; les résultats à bas débit étant mitigés ou pas suffisamment complets pour en tirer des conclusions intéressantes.

3.2 Exploitation de la psychoacoustique

Cette section étudie les techniques de codage basées principalement sur la psychoacoustique binaurale. D'abord, le codage de l'intensité stéréo (IS), à la base de toutes les autres techniques, est présenté. La sous-section suivante analyse le codage stéréo paramétrique (PS). Ensuite, des approches pour le codage multicanal sont présentées. D'abord, le codage de repères binauraux (BCC) à l'origine du codage multicanal paramétrique est décrit. Finalement, le codage audio spatial « MPEG Surround » est présenté selon les informations disponibles à ce moment (car la technologie est en cours de standardisation durant l'écriture de ce document).

3.2.1 Le codage de l'intensité stéréo (IS)

Cette technique a été introduite par Waal [26] et précisée par la suite par Herre [1], et repose sur le fait que les différences de phase entre les canaux ne sont pas perçues au-delà d'une certaine fréquence. Il s'agit d'un codage paramétrique (et psycho-acoustique) de l'enveloppe temporelle, les paramètres étant un facteur d'échelle (ou un rapport d'énergie) pour chaque canal à l'intérieur de chaque bande critique. Étant paramétrique, ce type de codage est par définition un codage avec pertes. De plus, il n'y a pas d'accord sur la fréquence la plus basse à laquelle la phase peut être négligée. Étant donné la taille moyenne de la tête et la longueur d'onde du son, cette limite serait d'environ 1.5 kHz pour la localisation de sons. Toutefois, la psycho-acoustique monaurale dicte que le verrouillage de phase (phase locking) est un mécanisme actif jusqu'à des fréquences de 4 à 5 kHz [11] et que les phases relatives d'un signal composé d'au moins trois sinusoïdes sont perceptibles dans une large gamme de fréquences [12]. De plus, c'est le codage de la phase à toutes les fréquences qui permet de préserver l'emplacement temporel des attaques dans un signal, correspondant à leur temps d'arrivée relatif (ITD) dans le cas du codage de plusieurs canaux. Parallèlement, le codage IS ne préserve pas les phases relatives

des enveloppes temporelles d'un signal à l'intérieur même d'une trame. Cela signifie qu'un signal modulé en amplitude (avec plus d'un cycle par trame) ne sera pas perçue comme ayant la même rudesse sonore suite à un tel codage. Pour toutes ces raisons, les codeurs modernes comme AC-3 appliquent rarement le mode IS aux fréquences sous 10 kHz. Il s'agit tout de même d'une technique encore utilisée aujourd'hui [20]. Dans le cas de AAC, il est stipulé que la préservation de l'enveloppe basse fréquence n'est pas problématique lorsque l'outil TNS est appliqué avant le codage IS [28]. Finalement, il est à noter que le codage IS peut être facilement étendu à plus de deux canaux en assignant tout simplement des facteurs d'échelle supplémentaires à chaque canal additionnel. Quelques variantes, dont deux basées sur la KLT, ont aussi été publiées [41]. Mais, de façon générale, le codage IS n'est efficace qu'en hautes fréquences. La section suivante couvre le sujet du codage stéréo paramétrique, qui est valable pour toute la bande de fréquences audio.

3.2.2 LE CODAGE STÉRÉO PARAMÉTRIQUE (PS)

Le codage stéréo paramétrique (PS) comprend les techniques qui permettent de coder très efficacement un signal audio stéréo sous forme d'un signal audio mono et de paramètres compacts décrivant l'image stéréo. Ces algorithmes sont basés sur la psychoacoustique binaurale, tandis que le signal audio mono peut être encodé avec n'importe quel codec audio mono conventionnel. Il s'agit d'une extension au codage de l'intensité stéréo (IS), où l'on se préoccupe de l'ambiance stéréophonique en plus de la simple localisation des signaux. De toute évidence, le signal stéréo reproduit au décodeur ne représente pas nécessairement les formes d'ondes originales, mais l'image stéréo sera tout de même perçue comme étant sensiblement identique. Cette sous-section fait une revue du codage stéréo paramétrique tel qu'il a été défini dans les publications et les standards récents [6, 7, 42, 43]. On y trouve la structure générale d'un codec audio intégrant cette approche pour encoder la stéréo, incluant la description de la représentation temps-fréquence requise. Ensuite, la mise en œuvre basée sur la transformée de Fourier rapide (FFT) est discutée en détail, incluant le calcul du canal mono, l'estimation des paramètres spatiaux, leur quantification et finalement, leur synthèse au décodeur.

3.2.2.1 STRUCTURE GÉNÉRALE

La structure générale d'un système de codage PS est présentée à la figure 7. La convention employée dans cette figure et les figures suivantes : une ligne pleine signifie un signal dans le domaine temporel ou fréquentiel, et une ligne pointillée représente des informations quantifiés ou non. Les signaux x_1 et x_2 peuvent être considérés comme les signaux du canal de gauche et du canal de droite respectivement. La figure 7 permet de constater que deux canaux, ici x_1 et x_2, sont d'abord analysés pour en extraire des paramètres représentatifs de l'information spatiale. Lors de cette étape, une version des deux canaux combinés en un seul est aussi produite. Ce canal unique est tout simplement de l'audio mono pouvant être encodé par n'importe quel codec audio. En général, comme décrit en détail à la section 3.2.2.4, la sortie mono n'est pas une somme linéaire des deux entrées, car l'énergie résultante serait alors dépendante du degré de corrélation entre les canaux.

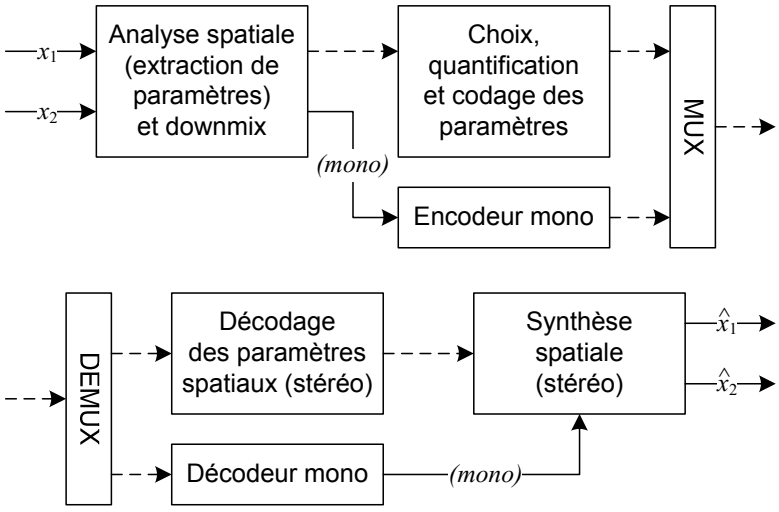

Figure 7 – Encodeur et décodeur stéréo paramétrique

Les données produites par l'encodeur audio monophonique et l'encodage des paramètres spatiaux sont alors transmis au décodeur. Pour sa part, le décodeur doit effectuer le chemin inverse : l'audio mono et les paramètres spatiaux sont décodés, pour ensuite recréer deux canaux qui seront perçus comme ayant des attributs spatiaux et la même sonorité que l'original (dans les limites du codec mono et de la précision de la description spatiale transmise).

3.2.2.2 REPRÉSENTATION TEMPS-FRÉQUENCE

Étant donné que les paramètres spatiaux sont estimés et synthétisés en fonction du temps et des fréquences, l'encodeur et le décodeur doivent mettre en œuvre une telle représentation à l'interne. Donc, la structure de base d'un encodeur et d'un décodeur stéréo paramétrique repose sur un banc de filtres ou bien une transformée. Une contrainte supplémentaire est que la résolution fréquentielle de ce système doit être non linéaire pour approximer la résolution fréquentielle de l'audition, c'est-à-dire proportionnelle aux largeurs des bandes critiques (avec plus ou moins de bandes selon la précision et la largeur de bande désirées). Parallèlement, une résolution temporelle suffisante (de l'ordre de quelques dizaines de millisecondes) doit tout de même être préservée pour tenir compte de la résolution binaurale du système auditif. Par contre, cette résolution doit être de l'ordre des millisecondes à l'emplacement des transitoires pour prendre en compte efficacement l'effet de précédence. De surcroit, le banc de filtres ou la transformée doit être sur-échantillonné pour éviter des artéfacts de reconstruction au décodeur. Finalement, le système sélectionné doit être basé sur une représentation complexe pour faciliter l'analyse et la synthèse de paramètres de phase ou de temps en fonction de la fréquence. Un système qui répond à tous ces critères est un système basé sur des FFT avec recouvrement de 50%. On notera qu'il existe aussi des mises en œuvre à plus faible complexité, basées sur des bancs de filtres QMF (Quadrature Mirror Filter) modulés par des exponentielles complexes [43]. Toutefois, cette approche n'est pas détaillée dans ce document.

3.2.2.3 ANALYSE ET SYNTHÈSE BASÉES SUR LA FFT

La figure 8 illustre un système d'analyse et de synthèse pour la stéréo paramétrique basé sur la FFT. D'abord, les deux signaux d'entrée x_1 et x_2 sont fenêtrés et transformés dans le domaine fréquentiel à l'aide de la FFT. Ensuite, les paramètres spatiaux sont estimés et une version mono appropriée est calculée. Les paramètres sont transmis au décodeur et le canal mono est synthétisé par recouvrement (OLA), et subséquemment encodé par un codec mono conventionnel. Au décodeur, on retrouve pratiquement l'opération inverse. Cependant, si des paramètres de corrélation inter-canal sont aussi transmis, une première étape du décodeur consiste

à appliquer un filtre de décorrélation au signal mono en entrée pour en obtenir une version dont l'allure temporelle est préservée, mais où la corrélation avec le signal d'entrée est pratiquement nulle. Une fois ce canal obtenu, le décodeur doit effectuer la même segmentation temps-fréquence que l'encodeur sur ces deux signaux et ensuite synthétiser des signaux de sortie avec les ratios d'énergie et le ratio de signal direct versus signal décorrélé et les déphasages adéquats.

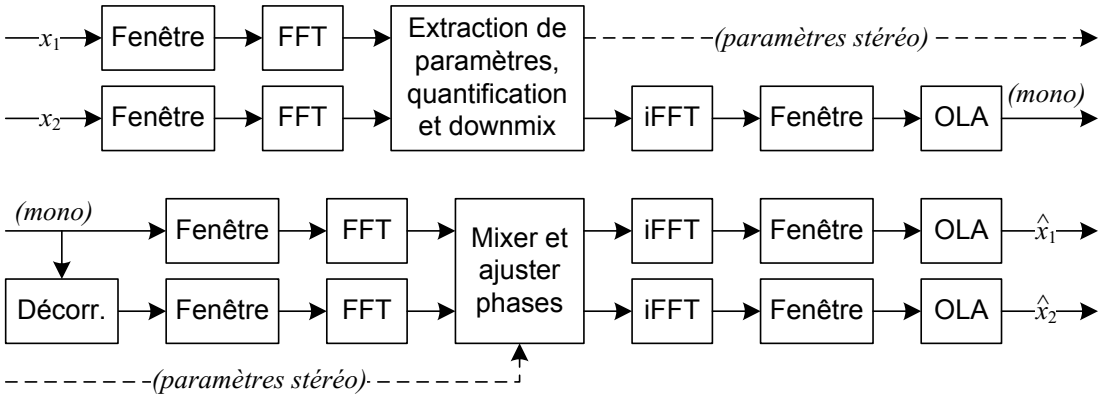

Figure 8 – Détails de l'encodeur et du décodeur stéréo paramétrique

La figure 9 illustre en plus grand détail le fenêtrage des signaux pour l'analyse et la synthèse par recouvrement (OLA). Ici, la taille des fenêtres est fixe, mais la structure peut évidemment être modifiée pour employer des fenêtres plus courtes lorsque des transitoires sont détectés. Dans cette figure, k représente le numéro de la trame en cours, M est la longueur des trames et OS est le facteur de sur-échantillonnage.

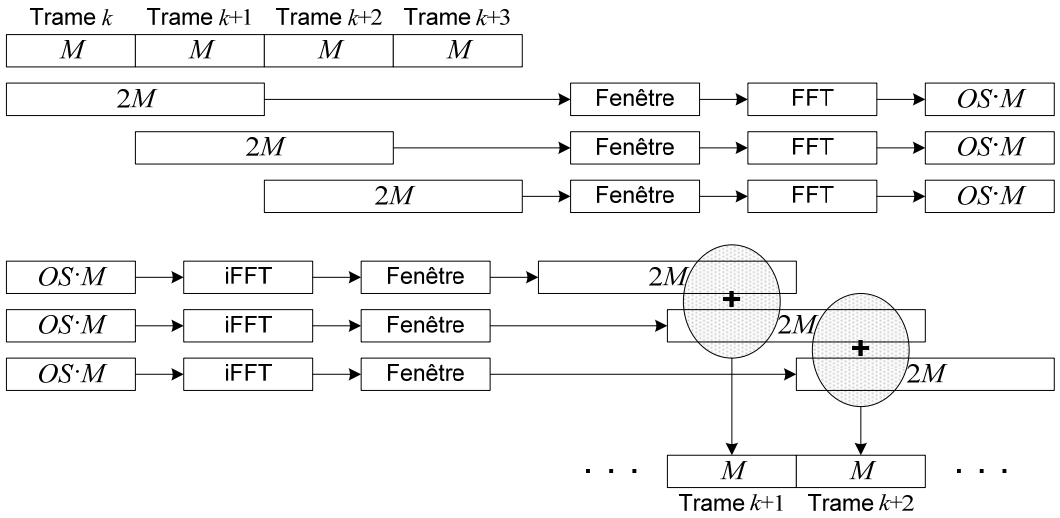

Figure 9 – Technique d'analyse et de synthèse par FFT

L'analyse de la figure 9 permet de constater qu'un délai algorithmique d'un total de $2M$ échantillons est introduit par cette structure, car la trame k+2 doit être disponible au complet avant de pouvoir débuter la synthèse de la trame k+1. Pour sa part, le facteur de sur-échantillonnage OS (de 2 en général) représente l'ajout de zéros à la FFT (zero-padding) pour éviter un repliement temporel suivant la modification des raies spectrales à la synthèse. En ce qui a trait au choix de la fenêtre, pour un recouvrement de 50% (pour $L = 2M$ comme à la figure 9) on peut choisir n'importe quelle fenêtre qui répond aux contraintes exprimées par les équations (3.3) et (3.4) :

$$w(2M-1-n) = w(n), \tag{3.3}$$

$$w^2(n) + w^2(n+M) = 1. \tag{3.4}$$

Ces équations correspondent aux conditions de Princen-Bradley pour la MDCT, mais sont aussi adéquates pour un système d'analyse et de synthèse par FFT. L'équation (3.3) signifie tout simplement que la fenêtre doit être symétrique, ou de façon plus générale, que la fin d'une fenêtre doit correspondre au début de la suivante (dans le cas de fenêtres de transition). L'équation (3.4) concerne le recouvrement : le signal de sortie du système doit avoir conservé une amplitude constante dans le temps après deux applications de la fenêtre (à l'analyse et à la reconstruction). La fenêtre la plus simple qui respecte ces deux contraintes est la fenêtre en sinus, définie par :

$$w[n] = \sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2M}\right]. \tag{3.5}$$

Alternativement, la fenêtre dite « Kaiser-Bessel derived » (KBD) $d[n]$ permet un certain contrôle sur son étalement grâce au paramètre α de la fenêtre de Kaiser $w[n]$:

$$d[n] = \begin{cases} \sqrt{\dfrac{\sum_{j=0}^{n} w[j]}{\sum_{j=0}^{N} w[j]}} & \text{pour } 0 \le n < N \\ \sqrt{\dfrac{\sum_{j=0}^{2N-1-n} w[j]}{\sum_{j=0}^{N} w[j]}} & \text{pour } N \le n < 2N \\ 0 & \text{ailleurs} \end{cases}, \tag{3.6}$$

où :

$$w[n] = \begin{cases} \dfrac{I_0\left(\pi\alpha\sqrt{1-\left(2n/N-1\right)^2}\right)}{I_0\left(\pi\alpha\right)} & \text{pour } 0 \le n \le N \\ 0 & \text{ailleurs.} \end{cases} \tag{3.7}$$

où I_0 est la fonction de Bessel de la première espèce. La fenêtre KBD sera de longueur $2N$, et elle tend vers une fenêtre rectangulaire lorsque α augmente.

3.2.2.4 CALCUL DU MÉLANGE MONO

Tout l'algorithme de stéréo paramétrique repose en premier lieu sur un signal mono adéquat. La façon la plus simple d'obtenir ce signal est tout simplement une somme dans le domaine temporel, exprimée par :

$$s[n] = w_1 x_1[n] + w_2 x_2[n], \tag{3.8}$$

où $s[n]$ est le mélange (down-mix) mono, $x_1[n]$ et $x_2[n]$ sont les canaux gauche et droit, n est l'index du temps, et w_1 et w_2 sont des poids qui déterminent le ratio de $x_1[n]$ et $x_2[n]$ contenus dans le signal de sortie mono. Généralement, une demi-somme, c'est-à-dire $w_1 = w_2 = 0.5$, correspond au cas « normal ». De toute évidence, le même calcul peut être effectué dans le domaine spectral, tel qu'exprimé par :

$$S[k] = w_1[k] X_1[k] + w_2[k] X_2[k], \tag{3.9}$$

où k est l'index de la raie spectrale, et où $S[k]$, $X_1[k]$ et $X_2[k]$ représentent les signaux $s[n]$, $x_1[n]$ et $x_2[n]$ respectivement, dans le domaine spectral. Ici, les poids $w_1[k]$ et $w_2[k]$ pourraient prendre des valeurs dépendantes de la fréquence. Cependant, dans le cas où $w_1[k] = w_2[k] = 0.5$, le même résultat que l'équation (3.8) est obtenu. Dans ces conditions, l'énergie du signal mono dépend de la corrélation et du déphasage entre les canaux en entrée. De plus, cette variation d'énergie est fonction du temps et de la bande de fréquences. La figure 10 illustre deux raies spectrales déphasées X_1 et X_2 dans le plan complexe. Il s'agit d'un exemple où l'énergie d'un mélange passif, dans ce cas-ci une demi-somme, est largement inférieure à la demi-somme des énergies de chacune de ces raies. Pour sa part, le mélange actif S' est obtenu en normalisant l'énergie de S dans chaque sous-bande, de telle façon que son énergie

soit équivalente à la moitié de la somme des énergies de X_1 et de X_2. En termes mathématiques, cela signifie que pour chaque sous-bande b constituée des raies spectrales k_b à k_{b+1}-1, le mélange actif est obtenu par l'équation (3.10) [6] :

$$S'[k] = \gamma[k] S[k], \tag{3.10}$$

où:

$$S[k] = \frac{X_1[k] + X_2[k]}{2}, \tag{3.11}$$

et :

$$\gamma[k] = \sqrt{\sum_{k=k_b}^{k_{b+1}-1} \frac{|X_1[k]|^2 + |X_2[k]|^2}{1/2\,|X_1[k] + X_2[k]|^2}}, \tag{3.12}$$

où $S[k]$ est défini par l'équation (3.11) comme étant la demi-somme de $X_1[k]$ et $X_2[k]$. En pratique, le terme $\gamma[k]$ est souvent limité à une valeur maximale de 2 (+6dB).

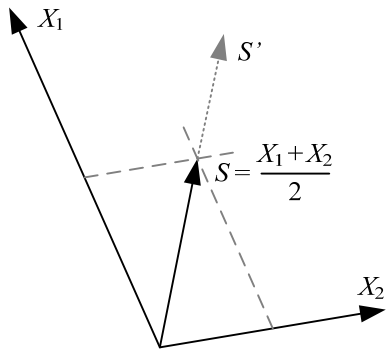

Figure 10 – Mélange passif (S) et actif (S')

Toutefois, la mise en œuvre d'un mélange actif dans le domaine spectral à l'encodeur au lieu d'une simple demi-somme signifie qu'un délai algorithmique de la longueur d'une trame du module de stéréo paramétrique est introduit avant le codec mono. De plus, la complexité d'une transformée inverse supplémentaire est ajoutée. Dans le cas de certains codecs, comme l'Enhanced aacPlus (avec SBR) [10], ces problèmes n'existent pas en pratique, car le module stéréo et le module de réplication de bande (SBR) partagent tous les deux la même représentation temps-fréquence. Toutefois, on remarquera qu'en ce qui concerne le codec AMR-WB+ [8], une innovation qui sera présentée à la section 4.2 permet de contourner ce problème sans ajouter de complexité ou de débit supplémentaire à l'ensemble du codec.

3.2.2.5 DÉFINITION DES PARAMÈTRES

Une fois que l'encodeur a segmenté, fenêtré et transformé chaque canal du signal audio de façon appropriée, les paramètres spatiaux doivent en être estimés. Cette section présente les quatre principaux paramètres qui sont définis pour le codage stéréo paramétrique, bien que certains standards ne les emploient pas tous.

D'abord, le paramètre qui est sans doute le plus important pour la reproduction stéréophonique est celui qui permet de positionner une image fantôme entre les deux haut-parleurs (ou entre les deux oreilles pour l'écoute sur casque). Il représente le ratio d'énergie entre les deux canaux, qui peut être représenté sous la forme d'un « IID » (Interchannel Intensity Difference). Il est défini pour chaque sous-bande b, constitué des raies spectrales k_b à k_{b+1}-1, par l'équation :

$$IID[b] = 10\log_{10}\frac{\sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_1^*[k]}{\sum_{k=k_b}^{k_{b+1}-1} X_2[k]X_2^*[k]} \quad \text{(dB)}, \tag{3.13}$$

où $X_1[k]$ et $X_2[k]$ sont les deux canaux stéréophoniques dans le domaine spectral et l'astérisque (*) représente le conjugué complexe. Cette équation permet de constater que le IID est tout simplement le ratio de l'énergie d'une sous-bande du premier canal sur le deuxième canal, en décibels. Le IID aura donc une valeur de 0dB pour un signal fantôme en plein centre, et peut prendre des valeurs de $-\infty$ à $+\infty$ pour des signaux complètement à droite ou à gauche. Toutefois, en pratique, cette valeur peut être quantifiée à l'intérieur des limites de la perception, c'est-à-dire dans une plage aux environs de ±30dB [12]. La résolution requise pour ce paramètre est de l'ordre du décibel pour des sons en face de l'auditeur (IID près de 0dB), tandis qu'elle est d'environ 5dB pour des sons près d'un côté ou de l'autre. Cela permet de tenir compte de la précision angulaire du système auditif qui est plus précis pour des angles face à l'auditeur. Pour cette raison, il a déjà été proposé de calculer l'angle d'arrivée perçue du signal stéréo (comme un angle de « panning ») au lieu du paramètre IID en prétendant qu'il existe de meilleures données psychométriques pour ceux-ci et que la plage des valeurs possibles est bornée. Toutefois, l'angle d'arrivée réellement perçue est difficile à estimer avec précision, car cela nécessite

un modèle qui dépend au minimum de la fréquence du signal. De plus, chaque angle d'arrivée possible correspond à un unique IID, donc rien n'empêche d'employer des données psychométriques en fonction de l'angle d'arrivée pour concevoir les tables de quantification des IID ayant les caractéristiques désirées. Les angles d'arrivé n'ont donc aucun avantage par rapport au paramètre IID, tout en nécessitant des calculs supplémentaires. Par conséquent, ils n'ont remplacé les ratios de puissance (IID) employés depuis plus d'une décennie [1] dans aucun codec standard. Ceci dit, les réalisations actuelles mettent en œuvre une table de quantification non uniforme, précisément dans le but d'accorder une plus grande précision aux sons directement en face de l'auditeur [6, 7].

Ensuite, un paramètre est requis pour transmettre le déphasage ou le délai entre les canaux par sous-bandes. En général, on considère les paramètres de phase au lieu de paramètres de délai entre les canaux, car ils sont plus faciles à synthétiser avec la précision requise, tout en minimisant l'introduction d'artéfacts. Le paramètre de phase est justifié par le fait que notre système auditif est sensible aux différences de phase entre les oreilles jusqu'à 1.5kHz et le « phase locking » est un phénomène psychoacoustique qui opère au niveau des neurones jusqu'à 5kHz environ. Donc même si la phase au dessus de 1.5kHz est inutile à la localisation, une différence audible peut être perçue dans la phase seule à des fréquences plus élevées. Le « phase locking » est aussi une des justifications pour transmettre la phase absolue en plus de la différence de phase entre les canaux. L'expression de la phase requiert donc deux paramètres : un pour le déphasage entre les canaux (IPD) et un autre pour la phase absolue par rapport au signal mono (OPD). La phase absolue est aussi nécessaire pour minimiser les artéfacts de reconstruction lors de l'étape de recouvrement (OLA). La figure 11 illustre ces deux paramètres dans le plan complexe par rapport à une raie spectrale des canaux X_1 et X_2, et de leur demi-somme S. Mathématiquement, les paramètres IPD (Interchannel Phase Difference) et OPD (Overall Phase Difference) sont définis pour chaque sous-bande b par les équations suivantes, où « $\angle$ » représente la tangente inverse sur quatre quadrants ($\pm\pi$) de la partie imaginaire divisée par la partie réelle :

$$IPD[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right), \tag{3.14}$$

$$OPD[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] S^*[k]\right). \tag{3.15}$$

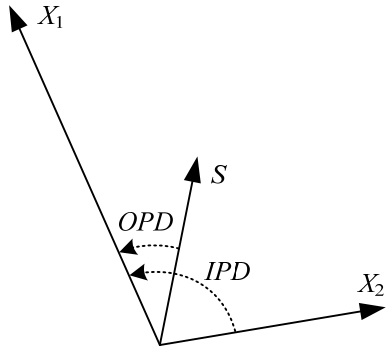

Figure 11 – Paramètres de phase *IPD* et *OPD*

Les paramètres de phase *IPD* et *OPD* ont donc une plage valide de $\pm\pi$ par définition.

Toutefois, pour capturer l'ambiance stéréophonique du signal original, même avec des paramètres de phase, un paramètre représentant la corrélation entre les canaux pour chaque sous-bande est nécessaire. Ce paramètre représente la corrélation entre les canaux, ou le paramètre *IC* (Interchannel Coherence). Dans les systèmes à bas débit, ce paramètre est plus efficace que les paramètres de phase pour obtenir une bonne qualité à bas débit [7]. Le paramètre *IC* est défini de deux façons différentes, selon qu'il s'agit d'une bande dont les paramètres de phase (*IPD* et *OPD*) sont aussi transmis ou non. Dans le cas où la phase est transmise, ou dans une version où la phase est ignorée pour les hautes fréquences [6], le paramètre *IC* est la corrélation entre les canaux dans une sous-bande, une fois que le déphasage (exprimé par le paramètre *IPD*) est « ignoré » ou « retiré » en prenant la valeur absolue de l'inter-corrélation :

$$IC[b] = \frac{\left|\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right|}{\sqrt{\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_1^*[k]\right)\left(\sum_{k=k_b}^{k_{b+1}-1} X_2[k] X_2^*[k]\right)}}. \tag{3.16}$$

Par contre, dans le cas où la phase n'est pas transmise, on définit IC_2 comme suit :

$$IC_2[b] = \frac{\mathrm{Re}\left\{\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right\}}{\sqrt{\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_1^*[k]\right)\left(\sum_{k=k_b}^{k_{b+1}-1} X_2[k] X_2^*[k]\right)}}. \quad (3.17)$$

De façon plus intuitive, si on suppose un sinus dans un canal et un cosinus dans l'autre, le paramètre IC vaut 1 lorsque les paramètres de phase sont transmis, mais le paramètre IC_2 vaut 0 pour une bande dont les paramètres de phase ne sont pas transmis. Parallèlement, pour un sinus avec un IPD de $\pi/2$, le IC vaudrait toujours 1, mais le IC_2 vaudrait la racine de ½. En fait, ceci n'est rien de plus qu'un exemple pour comprendre la nuance entre IC et IC_2, car on peut facilement démontrer par l'équation (3.18) que le paramètre IC vaut toujours 1 lorsqu'il n'y a qu'une seule raie spectrale dans une sous-bande. Dans ce cas hypothétique, $k_b = k_{b+1}$-1 et trois paramètres suffisent pour définir la phase de chaque raie et leur amplitude relative, ce qui explique d'ailleurs pourquoi ce paramètre n'est pas présent dans la figure 11.

$$\begin{aligned} IC[k_b] &= \frac{\left|X_1[k_b] X_2^*[k_b]\right|}{\sqrt{\left(X_1[k_b] X_1^*[k_b]\right)\left(X_2[k_b] X_2^*[k_b]\right)}} \\ &= \frac{\sqrt{\left(X_1[k_b] X_2^*[k_b]\right)\left(X_1[k_b] X_2^*[k_b]\right)^*}}{\sqrt{X_1[k_b] X_1^*[k_b] X_2[k_b] X_2^*[k_b]}} = 1 \end{aligned} \quad (3.18)$$

La plage des valeurs possibles pour le paramètre IC est de 0 à 1 par définition. En ce qui a trait à IC_2, la plage des valeurs possibles s'étale de -1 à 1. Par contre, cette plage est parfois limitée explicitement aux valeurs positives. De plus, une plus grande résolution est requise pour les valeurs près de 1.

3.2.2.6 SEGMENTATION TEMPS-FRÉQUENCE ET QUANTIFICATION DES PARAMÈTRES

Pour atteindre la plus grande qualité au débit le plus faible possible, la quantification de chacun des paramètres spatiaux doit être basée sur la résolution et les caractéristiques du système auditif. Mais d'abord, il faut établir quelle est la résolution appropriée pour la segmentation temps-fréquence. Ceci déterminera la longueur des trames et le nombre de sous-bandes (bandes critiques). Selon la vitesse de réaction du système auditif aux changements au niveau binaural, les paramètres spatiaux

devraient être mis à jour à une vitesse de l'ordre des dizaines de millisecondes (10-50ms). Dans le cas d'audio échantillonné à 48kHz, cela signifie que des trames de 1024 ou de 2048 échantillons peuvent être employées. Toutefois, pour éviter des problèmes de repliement temporel lors de la synthèse, la transformée employée a normalement le double d'échantillons, ce qui signifie une transformée de Fourier (FFT) de 2048 ou de 4096 échantillons. Par contre, l'effet de précédence a préséance lors d'un signal transitoire, ce qui implique un temps de réaction beaucoup plus rapide, typiquement inférieur à dix millisecondes. Une solution qui permet de tenir compte des signaux transitoires est de les détecter et de leur appliquer une fenêtre d'analyse et de synthèse plus courte. Parallèlement, le choix du nombre de bandes critiques est aussi très important. Les mises en œuvre actuelles, comme AAC, divisent un spectre large bande (20kHz) en 10, 20 ou 34 bandes critiques selon l'échelle ERB (équation (2.9)). Cela correspond à l'acuité auditive qui est d'environ 24 à 25 bandes critiques pour une telle largeur de bande.

Une fois la segmentation temps-fréquence déterminée, le débit final ne dépend que du choix des paramètres à transmettre et de leur quantification. D'abord, les paramètres de phase, lorsque transmis, ne sont nécessaires que pour les fréquences sous 1.5kHz environ. Cette fréquence représente environ la moitié des bandes critiques dans un codec large bande. Selon le standard actuel, les paramètres de phase IPD et OPD sont chacun représentés à l'aide d'un quantificateur uniforme à 3 bits, ce qui correspond au vecteur suivant :

$$\mathbf{IPDs} = \mathbf{OPDs} = \left[0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}, \frac{4\pi}{4}, \frac{5\pi}{4}, \frac{6\pi}{4}, \frac{7\pi}{4}\right] \tag{3.19}$$

En ce qui a trait au paramètre de corrélation inter-canal IC, il est aussi quantifié sur 3 bits selon les standards actuels. Cependant, cette quantification n'est pas uniforme, car notre perception de la diffusion des sons est plus précise pour des signaux corrélés. Le quantificateur pour IC est défini par le vecteur suivant :

$$\mathbf{IC2s} = \left[1, 0.937, 0.84118, 0.60092, 0.36764, 0, -0.589, -1\right] \tag{3.20}$$

Bien que cela ne soit pas réalisé dans le standard de stéréo paramétrique actuel, on remarque qu'un quantificateur **ICs** plus efficace peut être conçu pour les paramètres IC obtenus avec l'équation (3.16) ($IC[b]$), car les valeurs négatives sont impossibles. Seule l'équation (3.17) ($IC_2[b]$) peut engendrer des valeurs de corrélation négatives. Cette optimisation a été employée pour les expériences présentées au chapitre 5.

Le paramètre d'intensité IID est quantifié sur 4 ou 5 bits dans les standards actuels. Toutefois, les mêmes standards imposent toujours un choix ou l'autre pour toutes les bandes de fréquences d'une trame donnée. En aucun cas un choix de quantificateur n'est fait en fonction d'informations disponibles au décodeur comme l'énergie dans la bande mono ou l'importance de la plage de fréquences concernée. Par contre, la conception de ces quantificateurs tient compte de la plus grande sensibilité du système auditif pour les sons directement en face l'auditeur. Ces quantificateurs, de 4 et 5 bits respectivement, sont définis par les vecteurs suivants :

$$\mathbf{IIDs} = \left[-25,-18,-14,-10,-7,-4,-2,0,2,4,7,10,14,18,25\right], \tag{3.21}$$

$$\mathbf{IIDs} = \begin{bmatrix} -50,-45,-40,-35,-30,-25,-22,-19,-16,-13,-10,-8, \\ -6,-4,-2,0,2,4,6,8,10,13,16,19,22,25,30,35,40,45,50 \end{bmatrix}. \tag{3.22}$$

Puis, étant donné qu'un codage entropique (de Huffman) des index de paramètres spatiaux est effectué par la suite, le standard définit aussi la possibilité d'effectuer soit un codage absolu, soit un codage différentiel-temporel de ceux-ci. Les index de paramètres peuvent donc être exprimés de façon absolue ou par rapport à leur valeur précédente dans le temps. Les valeurs finales obtenues demeurent les valeurs exprimées dans les vecteurs de quantification ci-dessus, car il s'agit des index et non des valeurs des paramètres qui sont encodés de façon différentielle. Dans le cas des paramètres de phase IPD et OPD, cette approche permet de réaliser un codage modulo-différentiel (autour de 2π). Le codage différentiel-temporel permet, en théorie, de toujours obtenir une entropie égale ou inférieure au codage absolu. Toutefois, le gain est faible (15% pour les paramètres de phase), voir nul pour les paramètres IID et IC [7]. Notez que des tables de Huffman différentes sont utilisées pour l'encodage efficace des paramètres, selon qu'ils sont définis de façon absolue ou différentielle.

3.2.2.7 SYNTHÈSE DES PARAMÈTRES

Le décodeur stéréo a la tâche de synthétiser deux signaux de sortie à partir du canal mono à son entrée, ceux-ci ayant les caractéristiques décrites par les paramètres spatiaux. La figure 8 (page 30) illustre en détail les étapes de ce décodeur.

D'abord, pour être en mesure de synthétiser des signaux de sortie avec un degré de corrélation variable, un signal orthogonal doit être produit. Ce signal doit avoir une allure temporelle très proche de l'original, tout en étant incohérent au niveau de la structure fine de la forme d'onde. De préférence, la technique choisie sera un système linéaire à temps invariant (LTI) avec une allure passe-tout. Plusieurs techniques existent pour réaliser cette étape, la décorrélation de signaux étant pratiquement un domaine de recherche en soi. Cela dit, la technique la plus simple et la moins couteuse est un simple délai d'un nombre d'échantillons suffisant pour produire un effet de filtre en peigne dans toutes les bandes auditives (sans quoi la largeur de l'image stéréo sera réduite). Toutefois, l'effet de filtre en peigne en soi est indésirable, sans compter qu'un délai trop long peut introduire des effets d'écho lorsqu'il y a des transitoires dans le signal. Il est donc préférable de concevoir un filtre pour la décorrélation qui produit un délai qui varie en fréquence, sans introduire la structure harmonique du filtre en peigne. Un exemple de filtre passe-tout ayant un délai qui augmente linéairement dans le temps est une rampe de fréquence linéaire (chirp). Un filtre ayant une telle réponse impulsionnelle peut causer un artéfact lors de transitoires. Un meilleur choix de filtre qui répond aux critères énoncés ci-dessus est un filtre constitué d'un filtre de Schroeder [7]. Sa réponse impulsionnelle $h_d[n]$ est définie par l'équation suivante où $N_s = 640$ et $0 \le n < N_s$:

$$h_d[n] = \frac{2}{N_s} \sum_{k=0}^{N_s/2} \cos\left(\frac{2\pi}{N_s} k\left(k + n - 1\right) \right). \tag{3.23}$$

Pour des raisons de complexité, dans les réalisations basées sur un banc de filtres QMF, c'est un ensemble de cascades de filtres passe-tout à réponse impulsionnelle infinie (RII) qui est normalement employé. Dans le cas du standard Enhanced aacPlus [10], des cascades de trois (3) filtres passe-tout IIR sont définies pour

chacune des bandes basses et un simple délai de type $h[n] = z^{-D}$ est utilisé pour les bandes hautes, car le délai D requis à ces fréquences est très court (il est soit 1 ou 14 dans le standard actuel). Finalement, dans le standard Enhanced aacPlus, sous prétexte de réduire certains artéfacts, l'énergie du signal décorrélé est atténuée lorsque des signaux transitoires sont détectés dans le canal mono.

Suite à la décorrélation, le décodeur effectue le même découpage temps-fréquence que l'encodeur, soit un fenêtrage et une transformée de Fourier (FFT), tel qu'illustré à la figure 8. Le décodeur doit alors créer un signal ayant les caractéristiques spatiales décrites par les paramètres transmis. Il existe deux solutions ou « procédures » dans le standard pour le mixage des canaux de sortie, connus sous le nom de R_a et R_b. La procédure R_a est utilisée dans le standard du 3GPP, et la procédure R_b est surtout employée lorsque les paramètres de phase sont aussi transmis (le paramètre IC est toujours compris entre 0 et 1 par définition). Dans les deux cas, les signaux de sortie y_1 et y_2 (figure 8) sont obtenus dans le domaine fréquentiel (Y_1 et Y_2) pour chaque sous-bande b par mixage du signal direct $S[k]$ et du signal décorrélé $S_d[k]$ avec une matrice $\mathbf{R}_b$ de la façon suivante :

$$\begin{bmatrix} Y_1[k] \\ Y_2[k] \end{bmatrix} = \mathbf{R}_b[b] \begin{bmatrix} S[k] \\ S_d[k] \end{bmatrix}, \tag{3.24}$$

où $\mathbf{R}_b[b]$ correspond a l'une des deux procédures de mixage, soit $\mathbf{R}_A[b]$ ou $\mathbf{R}_B[b]$:

$$\mathbf{R}_A[b] = \mathbf{P}[b]\mathbf{A}_A[b]\mathbf{V}_A[b], \tag{3.25}$$

$$\mathbf{R}_B[b] = \sqrt{2}\,\mathbf{P}[b]\mathbf{A}_B[b]\mathbf{V}_B[b]. \tag{3.26}$$

Dans la procédure R_a, basée sur $\mathbf{R}_A[b]$, la matrice d'amplitude $\mathbf{A}_A[b]$ est définie par :

$$\mathbf{A}_A[b] = \begin{bmatrix} c_1[b] & 0 \\ 0 & c_2[b] \end{bmatrix}, \tag{3.27}$$

où $c_2[b]$ et $c_1[b]$ sont définis par ces équations :

$$c_1[b] = \sqrt{\frac{2c^2[b]}{1+c^2[b]}}$$
$$c_2[b] = \sqrt{\frac{2}{1+c^2[b]}}\ , \tag{3.28}$$

où $c[b]$ est le paramètre d'intensité *IID* transmis pour la bande b donnée.

Ensuite, la matrice de rotation $\mathbf{V}_A[b]$ qui doit mixer les proportions adéquates du signal direct versus le signal décorrélé est définie par l'équation suivante :

$$\mathbf{V}_A[b] = \begin{bmatrix} \cos\left(\beta[b] + \alpha_A[b]\right) & \sin\left(\beta[b] + \alpha_A[b]\right) \\ \cos\left(\beta[b] - \alpha_A[b]\right) & \sin\left(\beta[b] - \alpha_A[b]\right) \end{bmatrix}, \tag{3.29}$$

où les termes $\alpha_A[b]$ et $\beta[b]$ sont définis à leur tour par les équations suivantes :

$$\alpha_A[b] = \frac{1}{2}\arccos\left(\rho[b]\right), \tag{3.30}$$

$$\beta[b] = \alpha_A[b]\frac{c_1[b] - c_2[b]}{\sqrt{2}}, \tag{3.31}$$

avec $c_2[b]$ et $c_1[b]$ définis par l'équation (3.28) et $\rho[b]$, le paramètre de corrélation *IC* transmis pour la bande b donnée. Pour éviter des problèmes de stabilité lorsque cette procédure de mixage est employée, le paramètre de corrélation $\rho[b]$ est limité à une valeur minimale de 0.05 (ce qui empêche l'utilisation de valeurs négatives qui surviennent lorsque les paramètres de phase ne sont pas transmis).

En ce qui a trait à la procédure de mixage R_b, basé sur la matrice $\mathbf{R}_B[b]$, les termes $\mathbf{A}_B[b]$ et $\mathbf{V}_B[b]$ sont définis par les équations suivantes :

$$\mathbf{A}_B[b] = \begin{bmatrix} \cos\left(\alpha[b]\right) & -\sin\left(\alpha[b]\right) \\ \sin\left(\alpha[b]\right) & \cos\left(\alpha[b]\right) \end{bmatrix}, \tag{3.32}$$

et :

$$\mathbf{V}_B[b] = \begin{bmatrix} \cos\left(\gamma[b]\right) & 0 \\ 0 & \sin\left(\gamma[b]\right) \end{bmatrix}, \tag{3.33}$$

où $\alpha_B[b]$ et $\gamma[b]$ sont définis par les équations suivantes :

$$\alpha_B[b] = \frac{1}{2}\arccos\left(\frac{2c[b]\rho[b]}{c^2[b]-1}\right), \tag{3.34}$$

$$\begin{aligned}\mu[b] &= 1+\frac{4\rho^2[b]-4}{\left(c[b]+c^{-1}[b]\right)^2} \quad , \\ \gamma[b] &= \arctan\left(\sqrt{\frac{1-\sqrt{\mu[b]}}{1+\sqrt{\mu[b]}}}\right) \quad .\end{aligned} \tag{3.35}$$

Pour les deux procédures de mixage, la matrice $\mathbf{P}[b]$ qui exprime le déphasage dans chaque sous-bande b est définie par l'équation suivante :

$$\mathbf{P}[b] = \begin{bmatrix} e^{j\varphi_1[b]} & 0 \\ 0 & e^{j\varphi_2[b]} \end{bmatrix}, \tag{3.36}$$

avec les phases des canaux gauche et droit, $\varphi_1[b]$ et $\varphi_2[b]$ respectivement, définies dans les équations qui suivent en fonction des paramètres de phase « adoucis » (filtrés) dans le temps $\varphi_{opd}[b]$ et $\varphi_{ipd}[b]$:

$$\begin{aligned}\varphi_1[b] &= \varphi_{opd}[b] \\ \varphi_2[b] &= \varphi_{opd}[b] - \varphi_{ipd}[b] \quad ,\end{aligned} \tag{3.37}$$

où $\varphi_{opd}[b]$ et $\varphi_{ipd}[b]$ sont à leur tour définis par les paramètres de phases *IPD* et *OPD* transmis, soit $opd[b,n]$ et $ipd[b,n]$ respectivement, pour chaque sous-bande b, où n est l'index dans le temps de la trame en cours :

$$\begin{aligned}\varphi_{opd}[b] &= \angle\left\{\frac{1}{4}e^{j\cdot opd[b,n-2]} + \frac{1}{2}e^{j\cdot opd[b,n-1]} + e^{j\cdot opd[b,n]}\right\} \\ \varphi_{ipd}[b] &= \angle\left\{\frac{1}{4}e^{j\cdot ipd[b,n-2]} + \frac{1}{2}e^{j\cdot ipd[b,n-1]} + e^{j\cdot ipd[b,n]}\right\}\end{aligned} \tag{3.38}$$

et « $\angle$ » représentant la tangente inverse sur quatre quadrants ($\pm\pi$) de la partie imaginaire divisée par la partie réelle.

Finalement, le standard prévoit aussi la possibilité pour le décodeur d'interpoler les paramètres stéréo sur plusieurs trames lorsque ceux-ci varient lentement.

3.2.3 LE CODAGE MULTICANAL PAR REPÈRES BINAURAUX

Le codage paramétrique multicanal de type « codage des repères binauraux » (binaural cue coding, ou BCC) a été introduit récemment [3]. Ces techniques vont plus loin que le codage de l'intensité stéréo, car d'autres paramètres importants (du point de vue de la psychoacoustique binaurale) sont mis en œuvre. Il peut s'agir, entre autres, de la phase, du délai inter-canal ou de la corrélation inter-canal. Il est donc possible d'obtenir de meilleures performances avec le BCC qu'avec le codage de l'intensité seule [44]. Toutefois, l'analyse, la quantification, la synthèse et l'intégration des divers paramètres binauraux dans les codeurs actuels font toujours partie de la recherche active. Ces dernières années, Breebaart a conçu un système basé sur trois catégories de paramètres, soit l'intensité, la phase et la corrélation inter-canal; et où il n'existe aucun paramètre explicite pour la différence temporelle entre les canaux (ITD) [42]. Le paramètre de l'intensité est le logarithme du rapport de puissance entre les deux canaux en entrée. Les paramètres de phase sont calculés sur la portion du signal sous 2kHz; il s'agit de la différence de phase entre les deux canaux et la phase absolue. Cette dernière est la phase qui se retrouve entre un des deux canaux et le canal mono. Elle est requise, car la différence de phase n'est pas nécessairement répartie de façon égale par rapport au signal mono. Ensuite, le paramètre de corrélation est calculé et normalisé par la puissance des signaux et il est synthétisé en pondérant le canal mono avec une version parfaitement décorrélée de ce dernier, obtenu par un filtre spécial. Faller propose d'inclure aussi le paramètre ITD (sous forme de délai de groupe), jusqu'à 1.5 kHz, surtout pour l'écoute avec un casque. Toutefois, dans les cas où le débit total alloué permet déjà un codage transparent ou presque transparent, le codage paramétrique BCC n'obtient pas de meilleures performances que les techniques classiques comme celles employées dans un codec MP3 stéréo [45].

3.2.4 LE CODAGE MULTICANAL MPEG SURROUND

Tout d'abord, on précisera que le codage audio multicanal MPEG Surround [46, 47], précédemment connu sous le nom de « Spatial Audio Coding », est présentement en cours de standardisation au moment où ce document est rédigé. Cela dit, la plus grande différence entre le codage MPEG Surround et le codage BCC (Binaural Cue Coding) présenté plus tôt est fort probablement sa plus grande compatibilité avec la stéréo. Contrairement au codage BCC qui repose sur un mélange mono, le codage MPEG Surround a été conçu dès le départ pour être en mesure de synthétiser un signal ambiophonique (de type 5.1 ou autre) à partir d'un signal stéréo et des paramètres spatiaux de l'ordre de 16 kbps et plus. Le standard prévoit aussi la possibilité d'employer un mélange artistique au lieu d'une version calculée par l'encodeur. Toutefois, les premières expériences, présentées le 22 mai 2006 à l'AES à Paris [48], démontraient une diminution considérable de la qualité sonore de la synthèse ambiophonique avec l'emploi d'un tel mélange. Évidemment, ce système permet aussi la mise en œuvre d'un mélange compatible « Dolby Surround » pour supporter les systèmes ambiophoniques basés sur le matriçage. Un schéma illustrant les principes de base d'un codec audio stéréo intégrant le codage MPEG Surround pour l'ambiophonie à N canaux est présenté ci-dessous.

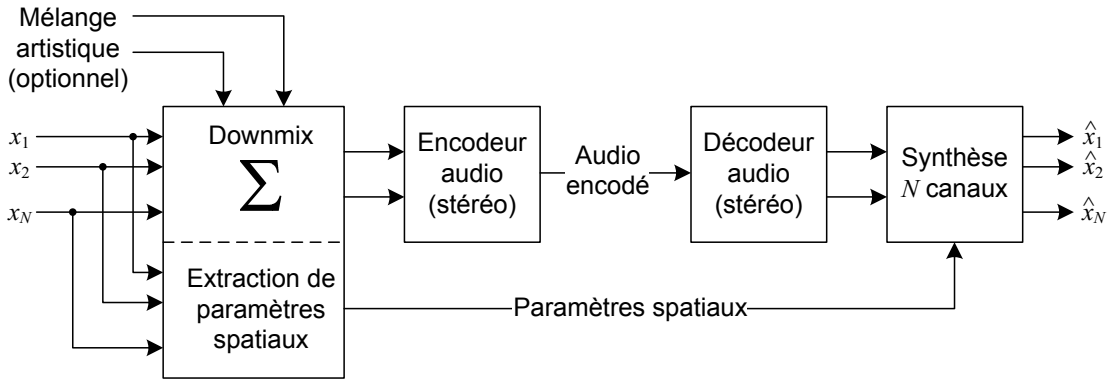

Figure 12 – Schéma-bloc d'un codec stéréo intégrant le MPEG Surround

Ce système permet d'intégrer l'ambiophonie à des codecs existants comme le MP3, car une version stéréo est toujours disponible et l'information spatiale peut être ignorée sans problème. Par contre, pour les systèmes ambiophoniques ayant le décodeur adéquat, une version à N canaux est disponible.

En détail, la structure de haut niveau de l'analyse et de la synthèse des paramètres est semblable à celle d'un codec stéréo (figure 8). En fait, le codage MPEG Surround combine les technologies développées pour le codage stéréo paramétrique et le codage BCC. Toutefois, en pratique, le standard est basé sur un banc de filtres QMF compatible avec celui utilisé pour l'extension de bande dans un codec AAC. De cette façon, la complexité globale d'un codec AAC avec MPEG Surround est minimisée.

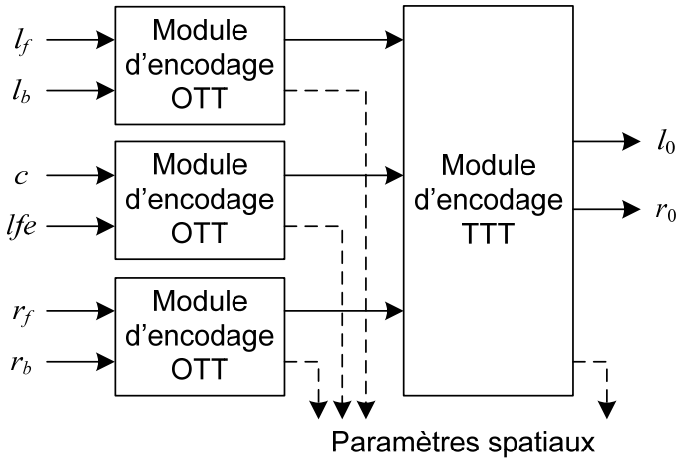

Figure 13 – Schéma des blocs de base d'un encodeur « 5.1 »

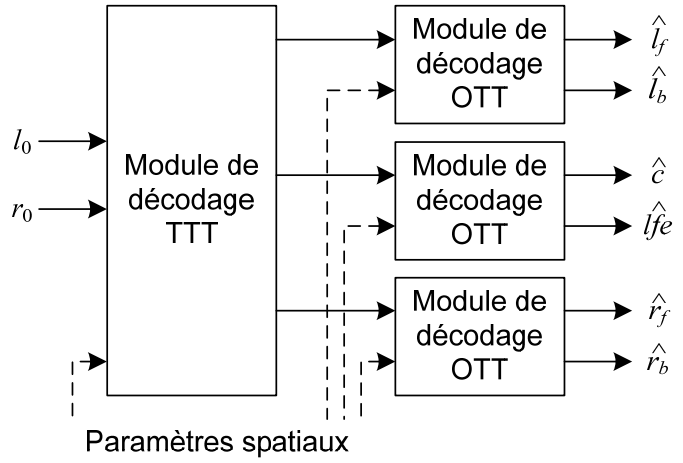

Figure 14 – Schéma des blocs de base d'un décodeur « 5.1 »

Plus précisément, le standard MPEG Surround est basé sur des blocs « stéréo » OTT (one to two) et des blocs TTT (two to three) qui permettent de réaliser plusieurs configurations de canaux, ainsi que des combinaisons intermédiaires. Par exemple, un système pourrait supporter l'ambiophonie 5.1 et ajouter deux modules OTT pour

obtenir de l'ambiophonie 7.1 (à partir des canaux arrière l_b et r_b). La figure 13 présente un encodeur 5.1 MPEG Surround basé sur de tels blocs, et la figure 14 présente le décodeur qui y est associé. Les désignations standards l_f, l_b, c, lfe, r_f et r_b correspondent respectivement aux canaux gauche-devant, gauche-arrière, centre, sous-graves, droit-avant et droit-arrière. Les canaux l_0 et r_0 représentent le signal stéréo qui sera encodé et transmis avec les paramètres spatiaux.

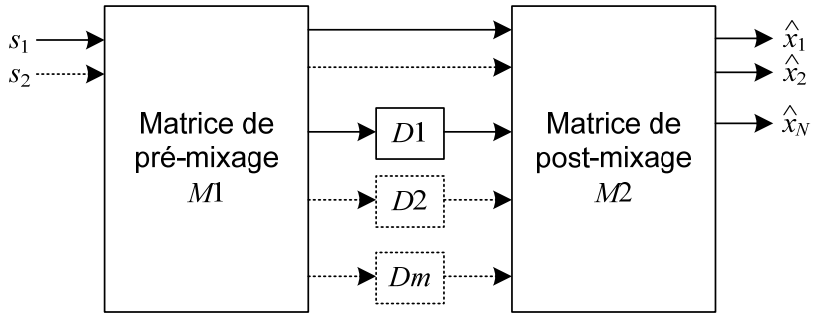

Figure 15 – Schéma-bloc de la synthèse des paramètres de corrélation

Toutefois, le décodeur MPEG Surround n'est pas mis en œuvre exactement comme il est représenté à la figure 14, car le module de décorrélation y serait imbriqué à plusieurs reprises (en série). La figure 15 illustre comment le système est réellement mis en œuvre. En particulier, les filtres de décorrélation Dn sont en parallèle pour éviter la décorrélation du même signal à plusieurs reprises. Cela permet d'éviter l'augmentation du délai algorithmique et l'étalement temporel. En ce qui a trait aux matrices de mixage $M1$ et $M2$, elles sont calculées de façon à obtenir les mêmes ratios de gain et degrés de corrélation que la version théorique présentée à la figure 14, où les modules Dn sont conçus pour produire des signaux décorrélés de l'entrée et décorrélés entre eux (il s'agit de filtres différents).

D'autres modules sont aussi en développement pour ce standard. Par ailleurs, il est prévu de lisser les paramètres spatiaux, de façon adaptative, en fonction de leur vitesse de variation. En ce qui a trait aux signaux décorrélés, un outil de post-traitement temporel (TP) a pour but de réajuster l'enveloppe temporelle pour mieux respecter celle du signal de départ. Une variante est la mise en forme de l'enveloppe temporelle (TES), inspirée du TNS. Toutefois, cette alternative introduit des artéfacts lorsque la mise en œuvre est basée sur des bancs de filtres QMF.

4. AMÉLIORATIONS AU CODAGE STÉRÉO PARAMÉTRIQUE

Suite à l'étude de l'état de l'art en codage stéréo paramétrique, ce chapitre propose deux améliorations importantes au standard actuel, qui ont aussi été présentées à l'AES [9]. En premier lieu, la section 4.1 démontre la présence d'une redondance d'information dans les paramètres stéréo actuels, et explique comment elle peut être exploitée pour transmettre la phase avec un seul paramètre (*IPD*) au lieu des deux normalement employés (*IPD* et *OPD*). Cette approche permet ainsi de réduire le débit numérique, avec une qualité sonore, une complexité et un délai algorithmiques équivalents. En second lieu, la section 4.2 démontre comment l'information contenue dans les paramètres stéréo peut aussi servir à compenser, au décodeur, l'énergie « annulée » lors de la simple somme des signaux *gauche* et *droit* plus ou moins en phase. Dans le cas d'un encodeur mono qui ne partage pas la même représentation temps-fréquence que le module de stéréo paramétrique, cette approche permet d'obtenir des performances équivalentes à la compensation à l'encodeur sans augmenter le débit numérique ou la complexité et le délai algorithmiques. Finalement, le chapitre 5 démontrera que la quantification des paramètres stéréo a un effet nul ou négligeable sur la performance des techniques présentées dans ce chapitre.

4.1 Réduction du débit requis pour transmettre l'information de phase

Tel que présenté à la section 3.2.2, deux paramètres sont requis pour transmettre l'information de phase, soit la phase entre les deux canaux (*IPD*) et la phase absolue par rapport au canal mono (*OPD*). Lorsque la phase est transmise, les paramètres de ratio d'intensité (*IID*) et de corrélation (*IC*) entre les canaux sont aussi transmis. Cette section démontre qu'il existe une redondance entre ces quatre paramètres, et que l'exploitation de cette redondance permet d'éliminer le paramètre *OPD* sans perdre de qualité. D'abord, voici un rappel des équations (3.13) à (3.16) présentées à la section 3.2.2.5 :

$$IID[b] = 10\log_{10}\frac{\sum_{k=k_b}^{k_{b+1}-1} X_1[k]X_1^*[k]}{\sum_{k=k_b}^{k_{b+1}-1} X_2[k]X_2^*[k]} \quad \text{(dB)}, \tag{4.1}$$

$$IC[b] = \frac{\left|\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right|}{\sqrt{\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_1^*[k]\right)\left(\sum_{k=k_b}^{k_{b+1}-1} X_2[k] X_2^*[k]\right)}}, \tag{4.2}$$

$$IPD[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right), \tag{4.3}$$

$$OPD[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k] S^*[k]\right), \tag{4.4}$$

où:

$$S[k] = \frac{X_1[k] + X_2[k]}{2}. \tag{4.5}$$

On peut définir les substitutions suivantes :

$$\alpha_1[b] = \sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_1^*[k], \tag{4.6}$$

$$\alpha_2[b] = \sum_{k=k_b}^{k_{b+1}-1} X_2[k] X_2^*[k], \tag{4.7}$$

$$\gamma_1[b] = \mathrm{Re}\left\{\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right\}, \tag{4.8}$$

$$\gamma_2[b] = \mathrm{Im}\left\{\sum_{k=k_b}^{k_{b+1}-1} X_1[k] X_2^*[k]\right\}. \tag{4.9}$$

Alors, on peut réécrire les équations (4.1) à (4.4), où « $\angle$ » représente la tangente inverse sur quatre quadrants ($\pm\pi$) de la partie imaginaire divisée par la partie réelle :

$$IID[b] = 10\log_{10}\frac{\alpha_1[b]}{\alpha_2[b]}, \tag{4.10}$$

$$IPD[b] = \angle\left(\gamma_1[b] + j\,\gamma_2[b]\right), \tag{4.11}$$

$$OPD[b] = \angle\left(\alpha_1[b] + \gamma_1[b] + j\,\gamma_2[b]\right), \tag{4.12}$$

$$IC[b] = \sqrt{\frac{\gamma_1^2[b] + \gamma_2^2[b]}{\alpha_1[b]\alpha_2[b]}}. \tag{4.13}$$

Ensuite, à la main ou à l'aide d'un outil de calcul algébrique comme le TI-92+, on obtient la solution suivante pour OPD en fonction des paramètres IID, IPD et IC :

$$OPD[b] = \angle\left(c[b] + IC[b]e^{j\,IPD[b]}\right), \tag{4.14}$$

où

$$c[b] = 10^{IID[b]/20}. \tag{4.15}$$

L'équation (4.14) permet donc de calculer la valeur du paramètre OPD au décodeur, sans que celui-ci soit transmis explicitement dans le train binaire. Le chapitre 5 démontre que l'estimation du paramètre OPD à partir des trois autres paramètres quantifiés procure une performance équivalente à l'approche qui consiste à transmettre un paramètre OPD distinct (qui est aussi quantifié).

Une fois cette équation trouvée, il est relativement facile de démontrer sa validité en prouvant qu'elle est équivalente à l'équation (4.4) en procédant par le chemin inverse. D'abord, il s'agit de remplacer le terme $IID[b]$ dans l'équation (4.15) avec la substitution définie à l'équation (4.10) :

$$c[b] = \sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}}. \tag{4.16}$$

Puis, les termes $c[b]$, $IC[b]$ et $IPD[b]$ de l'équation (4.14) sont remplacés par les substitutions définies aux équations (4.16), (4.13) et (4.11) :

$$OPD[b] = \angle\left(\sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}} + \sqrt{\frac{\gamma_1^2[b] + \gamma_2^2[b]}{\alpha_1[b]\alpha_2[b]}}e^{j\angle(\gamma_1[b] + j\gamma_2[b])}\right). \tag{4.17}$$

Ensuite, le terme exponentiel complexe de l'équation (4.17) représentant un point sur le cercle unité ayant l'angle de $\gamma_1[b] + j\gamma_2[b]$ est réécrit de façon équivalente, comme étant un vecteur normalisé par son module :

$$e^{j\angle(\gamma_1[b] + j\gamma_2[b])} = \frac{\gamma_2[b] + j\gamma_2[b]}{\sqrt{\gamma_1^2[b] + \gamma_2^2[b]}}, \tag{4.18}$$

ce qui permet alors de simplifier l'écriture de l'équation (4.17) :

$$OPD[b] = \angle\left(\sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}} + \sqrt{\frac{\cancel{\gamma_1^2[b] + \gamma_2^2[b]}}{\alpha_1[b]\alpha_2[b]}} \frac{\gamma_2[b] + j\gamma_2[b]}{\sqrt{\cancel{\gamma_1^2[b] + \gamma_2^2[b]}}}\right). \tag{4.19}$$

$$OPD[b] = \angle\left(\sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}} + \frac{\gamma_2[b] + j\gamma_2[b]}{\sqrt{\alpha_1[b]\alpha_2[b]}}\right) \tag{4.20}$$

Maintenant, l'opérateur « $\angle$ » représentant la tangente inverse sur quatre quadrants ($\pm\pi$) de la partie imaginaire divisée par la partie réelle sera réécrit en utilisant la notation atan2 (car $\tan^{-1}$ et atan sont définis pour seulement deux quadrants, c'est-à-dire $\pm\pi/2$). L'équation (4.20) devient donc :

$$OPD[b] = \text{atan2}\left(\frac{\frac{\gamma_2[b]}{\sqrt{\alpha_1[b]\alpha_2[b]}}}{\sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}} + \frac{\gamma_2[b]}{\sqrt{\alpha_1[b]\alpha_2[b]}}}\right), \tag{4.21}$$

$$OPD[b] = \text{atan2}\left(\frac{\gamma_2[b]}{\sqrt{\frac{\alpha_1[b]}{\cancel{\alpha_2[b]}}}\sqrt{\alpha_1[b]\cancel{\alpha_2[b]}} + \gamma_1[b]}\right). \tag{4.22}$$

Finalement, la preuve est complétée avec l'équation (4.23), qui correspond en tout point à l'équation (4.12) réécrite avec l'opérateur atan2 :

$$OPD[b] = \text{atan2}\left(\frac{\gamma_2[b]}{\alpha_1[b] + \gamma_1[b]}\right) \tag{4.23}$$

Cela démontre que le calcul du paramètre *OPD* au décodeur à partir des paramètres *IID*, *IC* et *IPD* donne un résultat exact (identique au calcul standard) en absence de quantification. Le chapitre 5 démontre que l'estimation du paramètre *OPD* à partir des paramètres *IID*, *IC* et *IPD* quantifiés procure des performances équivalentes à celles obtenues avec la transmission d'un paramètre *OPD* quantifié.

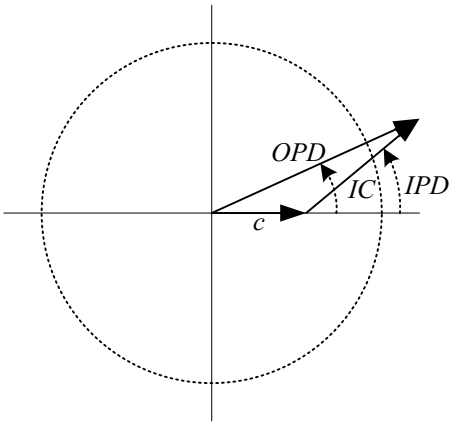

Figure 16 – Illustration du lien entre les paramètres stéréo

La figure 16 illustre l'équation (4.14) mettant ainsi en lumière le lien qui unit les quatre paramètres stéréo. Cette figure permet aussi de trouver plus facilement des équations qui permettraient de calculer un autre paramètre (en fonction des trois autres). Par exemple, en employant la loi des sinus et l'illustration ci-dessus, on trouve aisément la relation suivante pour IC :

$$IC = \frac{c \cdot \sin\left(OPD\right)}{\sin\left(IPD - OPD\right)}. \tag{4.24}$$

L'équation (4.24) peut s'avérer utile dans le cadre du standard actuel de stéréo paramétrique, du fait qu'il existe un bit *enable_icc* pour signal la présence ou non du paramètre IC [10]. Toutefois, le paramètre IC est normalement quantifié avec plus de précision que les deux paramètres de phase. Une analyse serait donc nécessaire pour évaluer si la précision obtenue sur le paramètre IC est suffisante. Quant à l'équation (4.14), le standard actuel prévoit un seul bit *enable_ipdopd* pour signaler la présence des deux paramètres de phase. Cela signifie que cette équation ne peut pas être employée dans le cadre du standard actuel, car il n'existe aucune façon d'être compatible avec les encodeurs et les décodeurs déjà déployés.

En ce qui a trait à l'estimation du paramètre IPD à partir des autres paramètres stéréo, un cas particulier est rencontré. Deux solutions peuvent exister, étant donné que le module du vecteur formé par l'angle OPD est inconnu. Ce cas particulier est illustré à la figure 17.

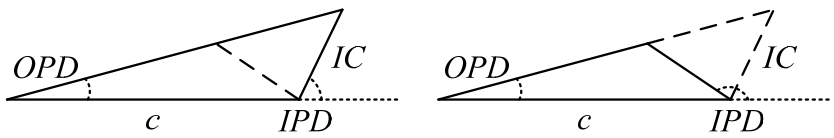

Figure 17 – Illustration des deux solutions possibles pour le paramètre *IPD*

4.2 Compensation d'énergie au décodeur

De toute évidence, le standard de stéréo paramétrique repose sur un mélange mono. Celui-ci peut être réalisé simplement de façon passive en calculant la demi-somme des deux canaux. Par contre, cette approche ne préserve pas l'énergie acoustique des signaux originaux. En effet, l'énergie de la demi-somme de deux canaux dans une bande de fréquence donnée dépend de leur degré de corrélation. Par exemple, la demi-somme de deux sinusoïdes identiques donne cette même sinusoïde (l'équation (4.25)), tandis que la demi-somme de deux sinusoïdes, dont l'une est déphasée de 90° par rapport à l'autre (tel qu'un sinus et un cosinus, exprimé par l'équation (4.26)), produit un signal avec une énergie 6dB plus faible :

$$\frac{1}{2\pi}\int_0^{2\pi}\left\{\frac{\sin(\theta)+\sin(\theta)}{2}\right\}^2 = \frac{1}{2\pi}\int_0^{2\pi}\left\{\sin(\theta)\right\}^2 = \frac{1}{2}, \tag{4.25}$$

$$\frac{1}{2\pi}\int_0^{2\pi}\left\{\frac{\sin(\theta)+\cos(\theta)}{2}\right\}^2 = \frac{1}{2\pi}\int_0^{2\pi}\left\{\frac{\sqrt{2}}{2}\sin\left(\theta+\frac{\pi}{4}\right)\right\}^2 = \frac{1}{4}. \tag{4.26}$$

Dans cette optique, le décodeur stéréo paramétrique ne pourra pas produire deux signaux de sortie ayant la même énergie que le signal original, car par définition les opérations effectuées produisent deux canaux dont l'énergie est égale à celle du canal mono en entrée. Pour ces raisons, l'idéal est d'effectuer un mélange actif qui normalisera l'énergie dans chaque bande critique pour qu'elle soit équivalente à la somme des énergies des deux canaux en entrée (section 3.2.2.4). Toutefois, cette opération, effectuée dans le domaine temps-fréquence du module de stéréo paramétrique, implique une transformée inverse lorsque le codec mono traite le signal dans un autre domaine, ainsi qu'un délai algorithmique supplémentaire associé à ces opérations. Pour ces raisons, certains codecs n'effectuent pas le

mélange actif. En revanche, l'innovation présentée ici permet de compenser le gain du signal mono au décodeur pour obtenir les mêmes performances qu'un mélange actif à l'encodeur. Étant donné que le signal est déjà dans le domaine temps-fréquence approprié au décodeur, cette approche n'augmente pas la complexité et le délai algorithmiques du codec. L'approche proposée ici emploie les paramètres stéréo déjà transmis au décodeur pour chaque bloc temps-fréquence pour calculer et compenser l'énergie « perdue » à l'encodeur lors de la simple demi-somme des signaux originaux.

Pour arriver à cette solution, il faut d'abord définir $E_{st}[b]$, soit l'énergie totale du signal stéréo; c'est-à-dire la somme des énergies des deux canaux :

$$E_{st}[b] = \sum_{k=k_b}^{k_{b+1}-1} \left(X_1[k] X_1^*[k] + X_2[k] X_2^*[k] \right), \tag{4.27}$$

et $E_m[b]$, l'énergie du mélange passif (l'énergie de la demi-somme) :

$$E_m[b] = 2 \sum_{k=k_b}^{k_{b+1}-1} S[k] S^*[k] = \frac{1}{2} \sum_{k=k_b}^{k_{b+1}-1} \left(X_1[k] + X_2[k] \right) \left(X_1^*[k] + X_2^*[k] \right). \tag{4.28}$$

Ensuite, l'écriture de ces équations peut être simplifiée à l'aide des substitutions définies par les équations (4.6) à (4.9) :

$$E_{st}[b] = \alpha_1[b] + \alpha_2[b], \tag{4.29}$$

$$E_m[b] = \frac{1}{2} \left(\alpha_1[b] + \alpha_2[b] + 2\gamma_1[b] \right). \tag{4.30}$$

Maintenant, $\lambda[b]$ est défini comme étant le ratio de l'énergie du signal stéréo $E_{st}[b]$ sur l'énergie de son mélange passif $E_m[b]$:

$$\lambda[b] = \frac{E_{st}[b]}{E_m[b]} = \frac{2\alpha_1[b] + 2\alpha_2[b]}{\alpha_1[b] + \alpha_2[b] + 2\gamma_1[b]}. \tag{4.31}$$

Ensuite, à la main ou à l'aide d'un outil de calcul algébrique comme le TI-92+, l'équation (4.31) est résolue en fonction des termes $c[b]$, $IC[b]$ et $IPD[b]$ définis par les équations (4.16), (4.13) et (4.11) :

$$\lambda[b] = \frac{2 + 2c^2[b]}{1 + c^2[b] + 2\,c[b]\,IC[b]\cos\left(IPD[b]\right)}. \tag{4.32}$$

Le facteur $\lambda[b]$ est calculé au décodeur pour chaque bande critique et il est appliqué sur le signal mono synthétisé $S[k]$ pour obtenir un signal mono $S'[k]$ ayant la même énergie acoustique que le signal stéréo original :

$$S'[k] = \sqrt{\lambda[b]}\,S[k]. \tag{4.33}$$

Cette approche peut aussi s'appliquer lorsque le paramètre de phase n'est pas transmis. Dans ce cas, le paramètre $IID[b]$ ne change pas, mais la corrélation inter-canal est alors définie par l'équation (3.17), reprise ici :

$$IC_2[b] = \frac{\mathrm{Re}\left\{\sum_{k=k_b}^{k_{b+1}-1} X_1[k]\,X_2^*[k]\right\}}{\sqrt{\left(\sum_{k=k_b}^{k_{b+1}-1} X_1[k]\,X_1^*[k]\right)\left(\sum_{k=k_b}^{k_{b+1}-1} X_2[k]\,X_2^*[k]\right)}}. \tag{4.34}$$

Cette équation peut être simplifiée avec les substitutions définies par les équations (4.6) à (4.8) :

$$IC_2[b] = \frac{\gamma_1[b]}{\sqrt{\alpha_1[b]\,\alpha_2[b]}}. \tag{4.35}$$

L'équation (4.31) est alors résolue à l'aide des équations (4.35) et (4.16) :

$$\lambda_2[b] = \frac{2 + 2c^2[b]}{1 + c^2[b] + 2\,c[b]\,IC_2[b]}. \tag{4.36}$$

Enfin, la validité des équations (4.32) et (4.36) peut être prouvée facilement, ce qui démontre leur équivalence à l'équation (4.31). D'abord, le terme $\cos(IPD[b])$ de l'équation (4.32) peut être réécrit de la façon suivante :

$$\cos\left(IPD[b]\right) = \cos\left(\angle\left(\gamma_1[b] + j\,\gamma_2[b]\right)\right) = \frac{\gamma_1[b]}{\sqrt{\gamma_1^2[b] + \gamma_2^2[b]}}. \tag{4.37}$$

L'équation (4.32) peut alors être reformulée en remplaçant chaque terme avec les substitutions définies plus tôt. L'équation obtenue est équivalente à l'équation (4.31) :

$$\lambda[b] = \frac{2 + 2\frac{\alpha_1[b]}{\alpha_2[b]}}{1 + \frac{\alpha_1[b]}{\alpha_2[b]} + 2\sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}}\sqrt{\frac{\gamma_1^2[b] + \gamma_2^2[b]}{\alpha_1[b]\alpha_2[b]}}\frac{\gamma_1[b]}{\sqrt{\gamma_1^2[b] + \gamma_2^2[b]}}}. \tag{4.38}$$

La même logique peut être appliquée à l'équation (4.36) :

$$\lambda_2[b] = \frac{2 + 2\frac{\alpha_1[b]}{\alpha_2[b]}}{1 + \frac{\alpha_1[b]}{\alpha_2[b]} + 2\sqrt{\frac{\alpha_1[b]}{\alpha_2[b]}}\frac{\gamma_1[b]}{\sqrt{\alpha_1[b]\alpha_2[b]}}}. \tag{4.39}$$

Cela démontre que le calcul de la compensation du gain au décodeur à partir des paramètres *IID*, *IC* et *IPD*, ou *IID* et IC_2, donne un résultat exact (identique à la compensation à l'encodeur) en absence de quantification. Par la suite, le chapitre 5 démontrera que l'erreur introduite par l'estimation du gain de compensation au décodeur à partir des paramètres stéréo quantifiés est négligeable.

5. ÉVALUATION DES PERFORMANCES

Ce chapitre présente une évaluation des performances des algorithmes proposés au chapitre précédent. Les équations qui y sont présentées sont exactes en absence de quantification des paramètres (*IID*, *IC* et *IPD*), mais elles sont en fait des estimations lorsqu'elles sont appliquées au décodeur sur des paramètres quantifiés.

5.1 Évaluation objective de l'estimation au décodeur du paramètre *OPD*

L'équation (4.14) proposée pour estimer le paramètre *OPD* à partir des trois autres paramètres stéréo est mathématiquement équivalente à l'équation (4.4) employée pour calculer le paramètre *OPD* dans un encodeur standard. Toutefois, l'objectif ici est d'appliquer l'équation (4.14) au décodeur à partir des paramètres *IID*, *IC* et *IPD* quantifiés. Dans cette optique, le but de cette analyse est d'étudier si la qualité obtenue avec cette approche est au moins équivalente à celle obtenue avec la transmission d'un paramètre *OPD* quantifié. Cette sous-section effectue donc une simulation de ces deux approches (*OPD* quantifié et transmis versus *OPD* estimé à partir de paramètres quantifiés) pour réaliser une analyse objective de leur performance. Ces simulations mettent en œuvre des quantificateurs standards [7], soit **IIDs**, **IC2s**, **IPDs** et **OPDs**, définis par les équations (3.22), (3.20) et (3.19).

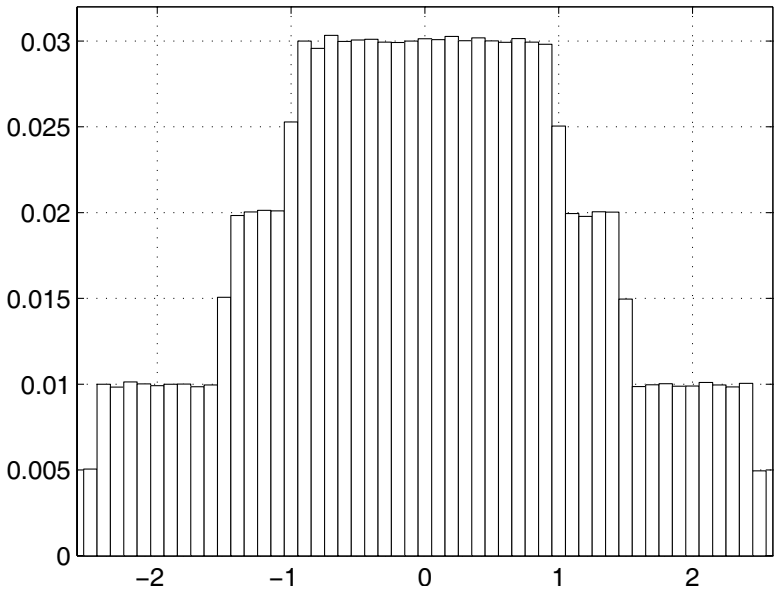

Figure 18 – Distribution de l'erreur de quantification du paramètre *IID*

Pour ces simulations, les paramètres IID, IC et IPD sont des valeurs aléatoires avec une distribution uniforme et le paramètre OPD est calculé en fonction de ceux-ci. La distribution de l'erreur de quantification pour le paramètre IID est illustrée avec la figure 18, qui montre bien que les pas du quantificateur **IIDs** ne sont pas uniformes.

Lorsque la phase (IPD) est transmise, le paramètre de corrélation IC ne peut prendre que des valeurs positives (entre 0 et 1). Dans ce cas, le quantificateur **IC2s** défini par l'équation (3.20) peut être remplacé par un quantificateur **ICs** plus efficace, défini ci-dessous par l'équation (4.40). La distribution de l'erreur de quantification pour ce paramètre est illustrée à la figure 19.

$$\mathbf{ICs} = \left[0, 0.248, 0.464, 0.645, 0.792, 0.902, 0.973, 1\right]. \tag{4.40}$$

Encore une fois, les valeurs générées pour cette simulation sont aléatoires avec une distribution uniforme. Un quantificateur non uniforme est employé, car il peut être conçu en fonction de la psychoacoustique et non en fonction de la distribution des valeurs à quantifier. Dans ce cas, les variations de corrélation minimales perceptibles sont employées [7, 49]. En ce qui a trait au quantificateur **IC2s**, une distribution d'erreur de forme semblable est obtenue, avec l'exception que l'amplitude de cette erreur est légèrement plus élevée.

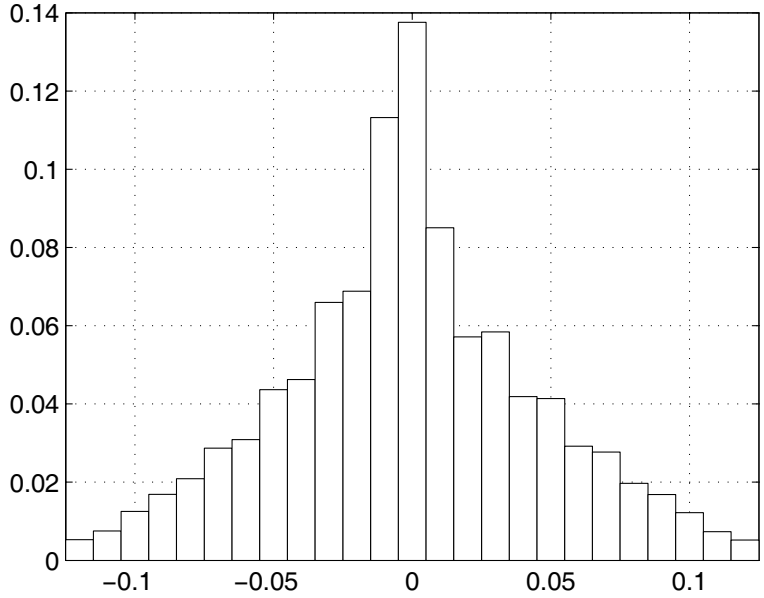

Figure 19 – Distribution de l'erreur de quantification du paramètre IC

Ensuite, la figure 20 montre la distribution de l'erreur de quantification du paramètre IPD. Étant donné qu'il s'agit d'une distribution uniforme quantifiée avec des pas uniformes de grandeur $\pi/4$, la distribution de l'erreur de quantification obtenue par simulation est bien entendu une distribution uniforme entre -$\pi/8$ et $\pi/8$.

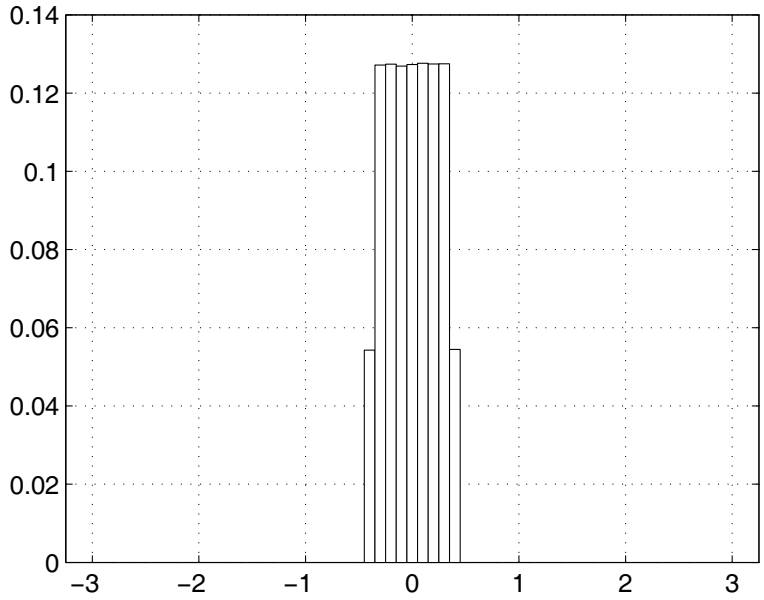

Figure 20 – Distribution de l'erreur de quantification du paramètre IPD

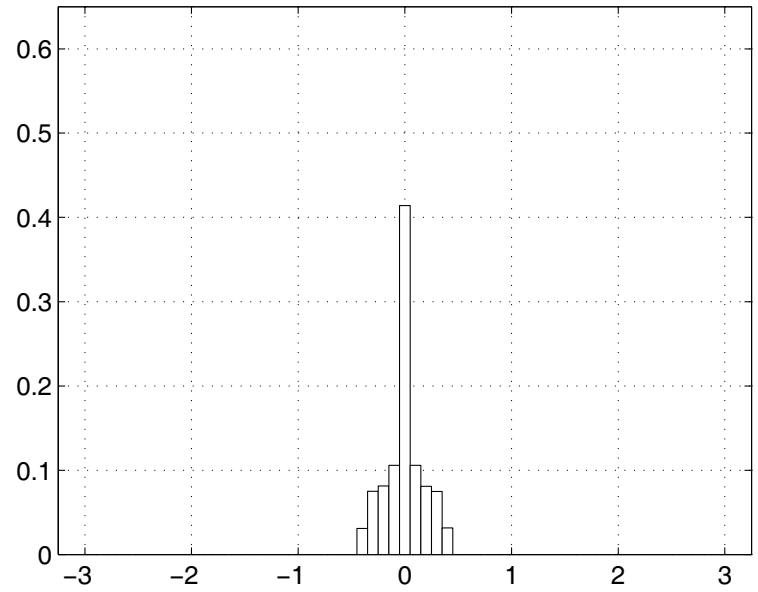

Figure 21 – Distribution de l'erreur de quantification du paramètre OPD

Puis, l'équation (4.14) est employée pour calculer la valeur du paramètre *OPD* correspondant aux combinaisons aléatoires de paramètres *IID*, *IC* et *IPD*. Cette façon de procéder permet d'obtenir une distribution de paramètres *OPD* plus réaliste qu'une simple distribution uniforme. Ces valeurs de paramètre *OPD* sont alors quantifiées pour obtenir la distribution d'erreur présentée à la figure 21.

À l'opposé, la figure 22 illustre la distribution de l'erreur d'estimation du paramètre *OPD* lorsque l'équation (4.14) est appliquée à partir des paramètres *IID*, *IC* et *IPD* préalablement quantifiés.

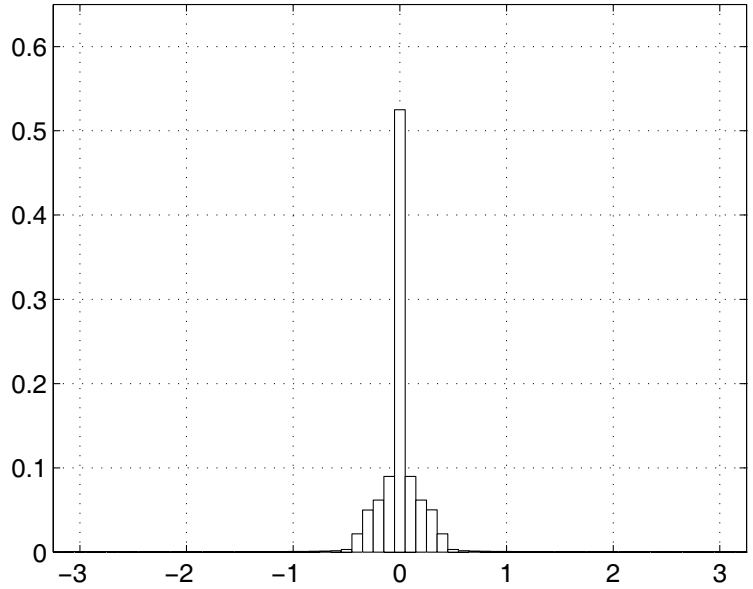

Figure 22 – Distribution de l'erreur d'estimation du paramètre *OPD* (**IPDs**)

Dans un deuxième temps, étant donné que les erreurs crêtes (outliers) sont légèrement plus élevées que lorsqu'un paramètre *OPD* quantifié est transmis, l'un des trois bits épargnés par l'omission de celui-ci est attribué à la quantification du paramètre *IPD*. Cela a aussi pour effet de réduire de moitié la distribution de l'erreur de quantification de ce paramètre (qui se trouve maintenant entre $-\pi/16$ et $\pi/16$). Ce nouveau quantificateur, **IPD2s**, est défini par l'équation (4.41) :

$$\mathbf{IPD2s} = \left[0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}, \pi, \ldots \right.$$
$$\left. \frac{9\pi}{8}, \frac{10\pi}{8}, \frac{11\pi}{8}, \frac{12\pi}{8}, \frac{13\pi}{8}, \frac{14\pi}{8}, \frac{15\pi}{8} \right] . \tag{4.41}$$

La figure 23 démontre que le résultat obtenu est meilleur qu'avec un paramètre *OPD* quantifié et transmis, même si le nombre total de bits a été réduit; sans compter que la précision du paramètre *IPD* a été doublée grâce au bit supplémentaire.

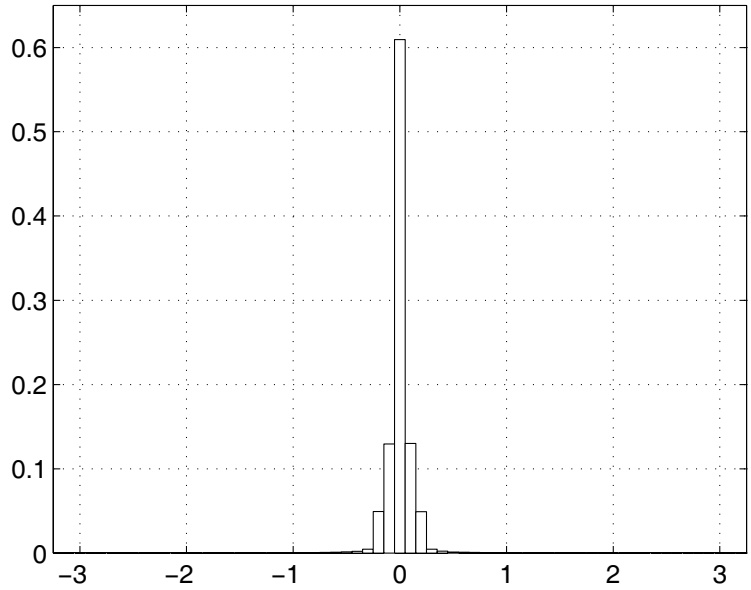

Figure 23 – Distribution de l'erreur d'estimation du paramètre *OPD* (**IPDs**)

Comme mise en évidence par la figure 11 (voir page 36), la phase d'une bande critique d'un canal (gauche) dépend directement du paramètre *OPD*, tandis que la phase de l'autre canal (droit) dépend de la somme des paramètres *IPD* et *OPD*.

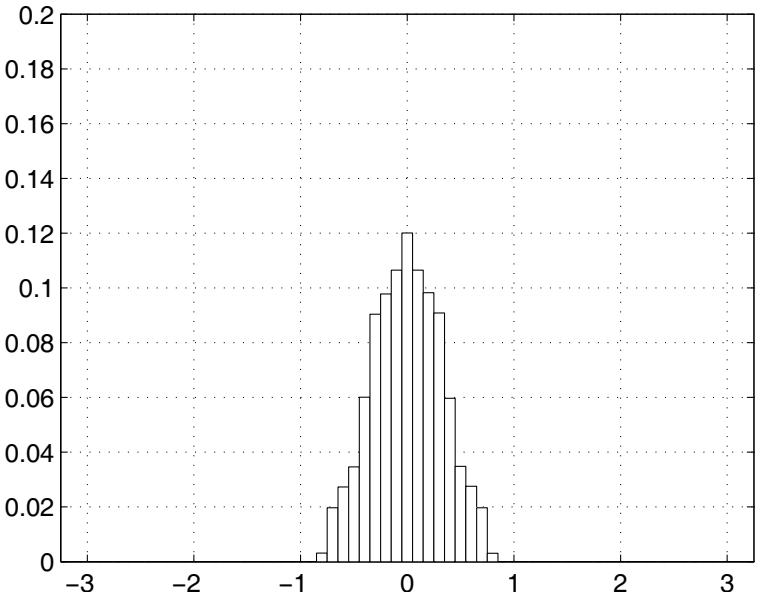

Figure 24 – Distribution de l'erreur de la somme *IPD* et *OPD* quantifiés (**IPDs**)

Donc, en plus de la distribution de l'erreur du paramètre *OPD*, on s'intéresse aussi à celle de la somme des paramètres *IPD* et *OPD*. La figure 24 montre cette distribution d'erreur lorsque les paramètres *IPD* et *OPD* sont tous les deux quantifiés. Ensuite, la figure 25 illustre la distribution de cette erreur lorsque le paramètre *OPD* est estimé à partir des trois autres paramètres (*IID*, *IC* et *IPD*) quantifiés, avec l'équation (4.14).

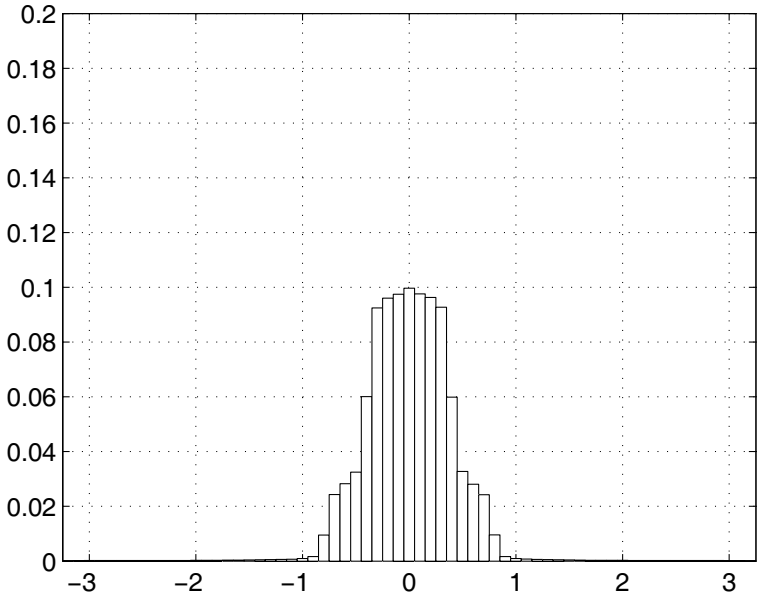

Figure 25 – Distribution de l'erreur de la somme *IPD* quantifié et *OPD* estimé (**IPDs**)

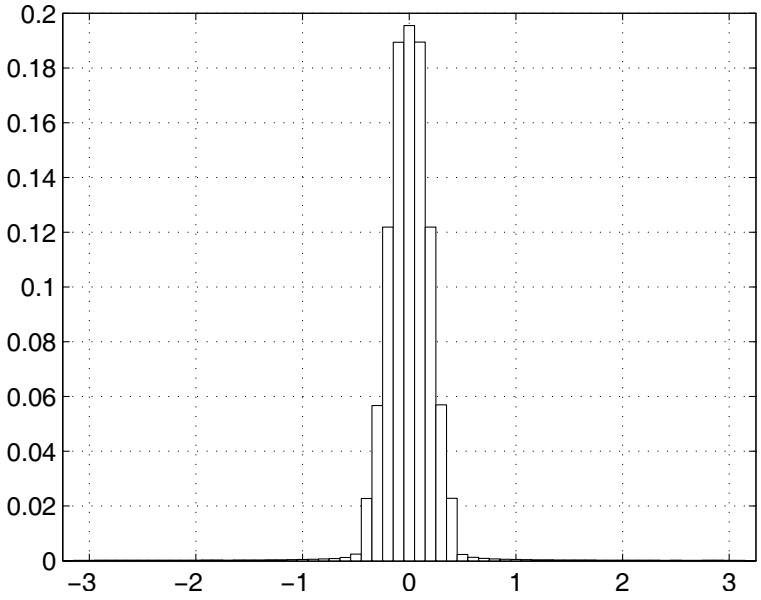

Figure 26 – Distribution de l'erreur de la somme *IPD* quantifié et *OPD* estimé (**IPD2s**)

Encore une fois, étant donné que le résultat obtenu n'est pas équivalent en tout point de vue à celui obtenu lorsque le paramètre OPD est quantifié, le quantificateur amélioré **IPD2s** décrit par l'équation (4.41) est mis en œuvre pour obtenir les résultats présentés à la figure 26. Ce dernier affiche des performances supérieures à celles obtenues avec un paramètre OPD quantifié.

Les distributions d'erreur présentées ci-dessus permettent d'observer l'étendue et l'amplitude des erreurs, mais on s'intéresse aussi à l'erreur quadratique moyenne. Le tableau 2 présente l'erreur RMS de l'angle OPD pour les différentes approches, tandis que le tableau 3 présente l'erreur RMS pour la somme des paramètres IPD et OPD. Étant donné que l'estimation est plus sensible aux signaux en inversion de phase, et que les enregistrements professionnels sont réalisés de façon à pratiquement éviter ces cas (par ailleurs pour conserver la compatibilité mono et FM) des simulations ont aussi été réalisées avec le paramètre IPD restreint à des angles entre $-\pi/2$ et $\pi/2$.

Quantification de $\angle X_1$: OPD	Erreur
OPD quantifié	0.176 rad
OPD estimé avec **IPDs**, IPD entre $-\pi$ et π	0.280 rad
OPD estimé avec **IPD2s**, IPD entre $-\pi$ et π	0.237 rad
OPD estimé avec **IPDs**, IPD entre $-\pi/2$ et $\pi/2$	0.169 rad
OPD estimé avec **IPD2s**, IPD entre $-\pi/2$ et $\pi/2$	0.120 rad

Tableau 2 – Erreurs RMS pour la phase du canal X_1

Quantification de $\angle X_2$: $OPD + IPD$	Erreur
OPD quantifié	0.324 rad
OPD estimé avec **IPDs**, IPD entre $-\pi$ et π	0.416 rad
OPD estimé avec **IPD2s**, IPD entre $-\pi$ et π	0.283 rad
OPD estimé avec **IPDs**, IPD entre $-\pi/2$ et $\pi/2$	0.351 rad
OPD estimé avec **IPD2s**, IPD entre $-\pi/2$ et $\pi/2$	0.195 rad

Tableau 3 – Erreurs RMS pour la phase du canal X_2

Les tableaux ci-dessus permettent de constater que les résultats obtenus pour les angles $\angle X_1$ et $\angle X_2$ avec l'estimation du paramètre OPD sont comparables à ceux obtenus avec la transmission de celui-ci, particulièrement lorsque le quantificateur **IPD2s** est employé et que l'audio est compatible mono (sans composantes en inverse de phase). Dans ce cas, l'approche proposée performe mieux que la transmission du paramètre OPD, tout en économisant sur le débit total. Cette approche permet aussi de réduire l'erreur RMS maximale des deux canaux (gauche et droit), surtout lorsque le quantificateur **IPD2s** est employé.

En somme, une analyse objective de la performance de l'approche proposée pour réduire le débit numérique d'une extension stéréo paramétrique démontre que les performances obtenues seront au moins aussi bonnes que celles réalisées avec l'approche actuelle où le paramètre OPD est quantifié et transmis. Les résultats d'une évaluation subjective présentée à la section 5.3 confirment que la qualité obtenue avec l'approche proposée est équivalente, tout en ayant un débit numérique inférieur.

5.2 Évaluation objective de la compensation d'énergie au décodeur

La compensation d'énergie par sous-bandes réalisée au décodeur, à partir des équations (4.32) et (4.36), est mathématiquement équivalente à la compensation à l'encodeur en absence de quantification des paramètres stéréo. Cette section étudie les distributions de l'erreur sur ces facteurs de gain lorsque cette compensation s'effectue sur des paramètres quantifiés. Si ces erreurs sont en dessous de 1 dB, on peut conclure que leur effet sera négligeable sur la qualité de la synthèse audio.

La figure 27 présente la distribution de l'erreur (en dB) sur le gain de compensation, obtenue lorsque l'équation (4.32) est appliquée aux paramètres stéréo quantifiés IID, IC et IPD. Cette simulation a été effectuée avec les quantificateurs **IIDs**, **ICs** et **IPD2s** définis auparavant par les équations (3.22), (3.20) et (4.41).

Pour sa part, la figure 28 présente la distribution de l'erreur (en dB) sur le gain de compensation, obtenue lorsque l'équation (4.36) est appliquée aux paramètres stéréo quantifiés IID et IC_2. Cette simulation a été effectuée avec les quantificateurs **IIDs** et **ICs** définis par les équations (3.22) et (3.20).

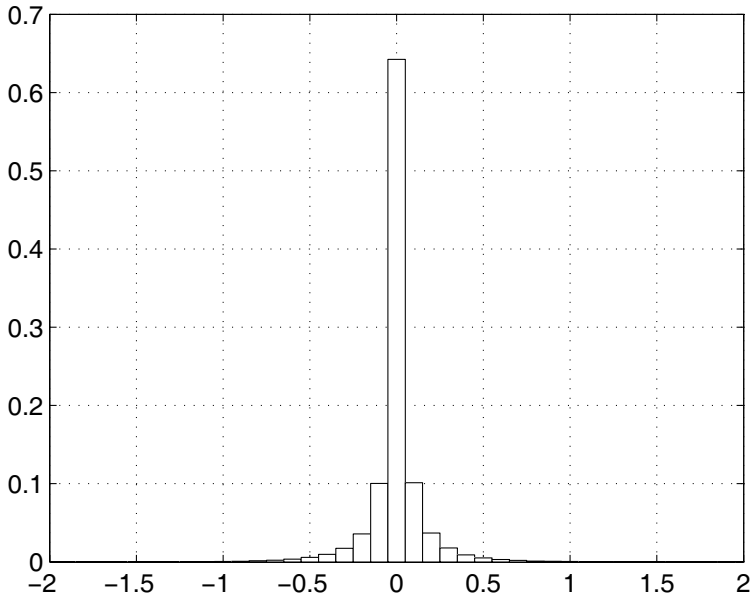

Figure 27 – Distribution de l'erreur (en dB) sur le gain de compensation (IPD)

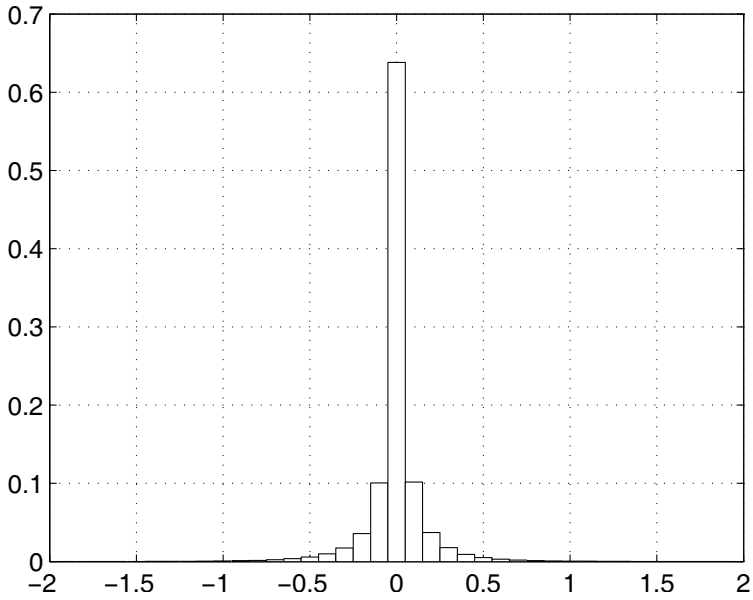

Figure 28 – Distribution de l'erreur (en dB) sur le gain de compensation (IC_2)

Étant donné que les erreurs d'estimation obtenues sont bien en dessous de la barre d'un décibel, on peut les qualifier sans risque de négligeables. Dans le contexte où cette technique est employée parce que la compensation d'énergie à l'encodeur est impraticable, l'évaluation subjective présentée à la section 5.3 démontre clairement un gain de qualité important avec l'emploi d'un facteur de gain estimé au décodeur, tel que proposé, par rapport au cas où rien n'est fait.

5.3 Évaluation subjective des variantes de stéréo paramétrique

Cette section présente une évaluation subjective des techniques proposées dans le chapitre 4. Étant donné que des techniques paramétriques sont à valider, cette évaluation est réalisée en employant la méthodologie MUSHRA, conçue pour des codecs audio ayant une qualité jugée intermédiaire [50]. Ce test est réalisé à l'aide d'un casque d'écoute et non de haut-parleurs, car il s'agit du cas le plus critique et le plus reconnu pour la stéréophonie [7, 51]. De surcroit, l'utilisation d'un casque d'écoute permet d'éliminer l'effet de la salle d'écoute sur le champ sonore qui atteint les oreilles de l'auditeur. Plus précisément, l'équipement utilisé pour les tests était principalement constitué d'un casque d'écoute Beyerdynamic DT 770 et d'une carte de son M-Audio Audiophile 192 avec amplificateur Rotel RA-971 Mk II ou d'une carte de son Creative Sound Blaster Audigy 2 avec un amplificateur pour écouteurs Rega Ear (modèle 2006). Une capture d'écran de l'interface graphique du logiciel MUSRA pour Windows mis au point au sein du laboratoire (GRPA) est présentée ci-dessous :

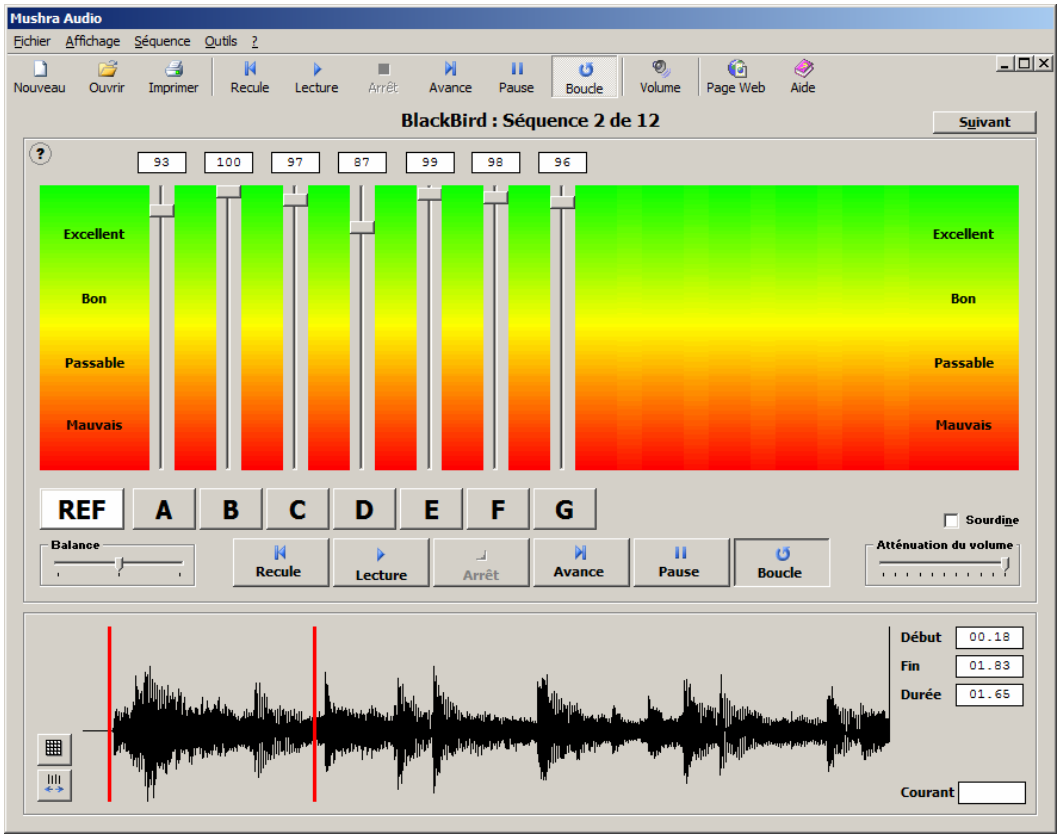

Figure 29 – Capture d'écran du logiciel pour tests d'écoute MUSHRA

Ce logiciel permet d'écouter un extrait en boucle, sélectionné entre deux curseurs (début et fin). Il est possible de passer à n'importe quel moment entre la référence et les 7 autres items avec une seule touche de clavier ou un seul clic de souris, et ce, sans artéfact (clic) sonore. La référence peut être écoutée autant de fois que désiré, et une version identique de celle-ci est cachée parmi les choix A à G. Cette référence cachée permet de valider la capacité d'un auditeur à réellement entendre des différences entre les échantillons et permet d'éliminer ceux qui réduiraient la précision des résultats finaux. La tâche de l'auditeur est donc d'évaluer la qualité des items A à G en fonction de la référence (REF). Idéalement, un seul des items se verra attribuer la note parfaite de 100, car un des items est une copie de la référence.

Par ailleurs, un échantillon de référence (anchor) est aussi requis dans un test MUSHRA pour éviter que de « bons » items se voient attribuer une trop mauvaise note. Cet échantillon doit donc être de qualité inférieure aux autres items à évaluer. Une version à bande réduite (passe-bas) de l'originale est souvent employée, mais elle force l'auditeur à choisir entre deux types d'artéfacts différents (ici, bande limitée versus stéréophonie). Pour éviter ce problème, l'échantillon de référence employé dans ce test est une version codée par intensité seulement (l'item annoté IID).

Ci-dessous, le tableau 4 présente les acronymes utilisés pour décrire les résultats obtenus (figure 30). Par exemple, l'acronyme « IPD_CE » signifie une version de stéréo paramétrique qui transmet trois paramètres, soit *IID*, *IC* et *IPD*, et qui estime la valeur du paramètre *OPD* au décodeur. Ce dernier effectue aussi une compensation d'énergie (CE) au décodeur calculée à partir des trois mêmes paramètres stéréo.

Référence	Version originale
IID	Paramètre d'intensité *IID* seulement
IC	Paramètre d'intensité *IID* et de corrélation *IC*
IPD	Trois paramètres *IID*, *IC* et *IPD* (*OPD* estimé)
OPD	Quatre paramètres quantifiés et transmis
*_CE	Compensation d'énergie au décodeur

Tableau 4 – Légende des acronymes employés dans la figure 30

Étant donné que l'objectif de cette étude est d'évaluer la performance de diverses alternatives au niveau de la stéréo paramétrique, le signal issue de la somme des deux canaux originaux n'est pas encodé et décodé par un codec audio monophonique avant d'être présenté au décodeur stéréo.

Puis, les tests d'écoute ont eu lieu avec 12 auditeurs, incluant l'auteur. Il y avait 12 séquences critiques à évaluer, composées de musique et d'extraits de film (incluant de la parole). Enfin, la compilation des résultats obtenus lors de ce test est présentée à l'aide de la figure 30, incluant des marqueurs pour l'intervalle de confiance à 95%.

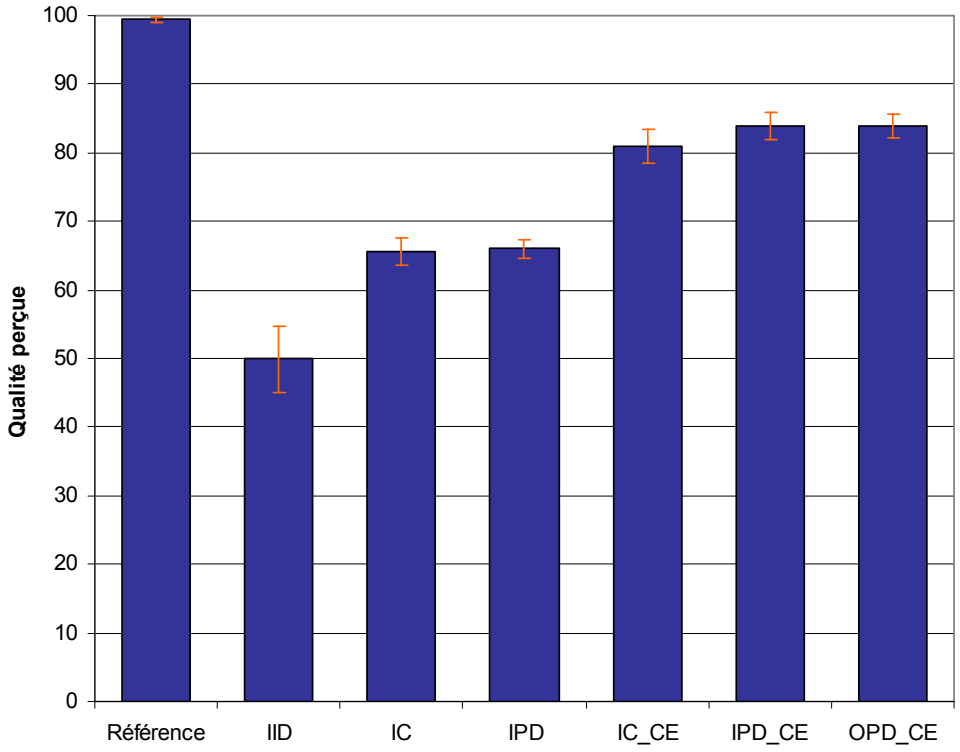

Figure 30 – Évaluation subjective de diverses variantes de stéréo paramétrique

D'abord, on constate que même l'état de l'art en codage stéréo paramétrique n'atteint pas la transparence sur des extraits sonores critiques. Cela dit, le premier résultat remarquable est l'effet de la compensation d'énergie. Bien que la version employée ici, au décodeur, n'ait pas été comparée à la compensation d'énergie à l'encodeur, on peut aisément conclure que la compensation d'énergie dans un codec stéréo paramétrique est très importante. La performance du codec « IC_CE » est préférable à celle de la version « IC », et la performance du codec « IPD_CE » est aussi

supérieure à celle de la version « IPD ». Donc, dans les cas où il s'avère impraticable d'appliquer la compensation d'énergie à l'encodeur, il est tout de même possible de parvenir à des gains de qualité remarquables en appliquant la compensation de gain proposée pour le décodeur à la section 4.2.

Parallèlement, la très faible différence entre les résultats pour les codecs « IPD_CE » et « OPD_CE » démontre que l'estimation du paramètre *OPD* au décodeur à partir des trois autres paramètres quantifiés, tel que présenté à la section 4.1, procure une qualité équivalente à un codec où le paramètre *OPD* est quantifié et transmis.

Finalement, on remarque peu de gain de performance lorsque le paramètre de phase est employé. Le paramètre de corrélation inter-canal *IC* est celui qui permet réellement d'obtenir une qualité sonore acceptable avec la stéréo paramétrique. Cela explique fort bien pourquoi les paramètres de phase ne sont pas employés dans la version du codec aacPlus du 3GPP [10]. D'un autre côté, le débit perdu par la redondance entre les paramètres de phase *IPD* et *OPD* explique aussi pourquoi le coût de transmettre la phase a été jugé trop élevé. Ce choix pourrait donc être réévalué en tenant compte de la diminution de débit obtenu grâce à la mise en œuvre des contributions présentées dans ce mémoire.

6. CONCLUSION

Ce mémoire a placé l'accent sur le codage audio paramétrique à bas débit de plus d'un canal audio. Les approches existantes, développées avant et durant l'avancement de ces travaux, ont été étudiées en détail. D'ailleurs, cette étude a mené à la proposition de deux améliorations significatives à l'état de l'art.

6.1 Résumé du mémoire

Faisant suite au chapitre d'introduction, une revue des notions de psychoacoustique pertinentes aux modèles paramétriques présentés a été effectuée au chapitre 2. Ensuite, au chapitre 3, les techniques de codage stéréo et multicanal ont été présentées, avec un accent particulier sur le codage stéréo paramétrique. D'ailleurs, une étude approfondie du standard MPEG pour la stéréo paramétrique [6, 7] a mené à des propositions pour l'amélioration de celui-ci, présentées dans ce mémoire au chapitre 4, ainsi qu'à la 120^{e} convention de l'AES à Paris [9]. Ces contributions sont significatives du fait qu'elles s'appliquent aux deux codecs standardisés qui sont reconnus comme étant l'état de l'art, soit l'AMR-WB+ et l'Enhanced aacPlus.

Plus précisément, les améliorations proposées permettent d'intégrer la stéréo paramétrique à un codec comme l'AMR-WB+ avec un gain de qualité et une réduction du débit requis par rapport à la mise en œuvre décrite dans le standard, tout en minimisant la complexité et le délai algorithmiques ajoutés. En particulier, il est démontré que l'information requise pour transmettre l'information correspondant à la phase des signaux dans les deux canaux peut être réduite par rapport au standard sans effet négatif sur la qualité sonore. Parallèlement, la compensation d'énergie normalement réalisée à l'encodeur peut maintenant être effectuée au décodeur sans perte de qualité et sans ajout d'information supplémentaire, permettant ainsi une mise en œuvre plus efficace de la stéréo paramétrique dans bien des cas de figure.

Par la suite, le chapitre 5 a présenté les résultats d'analyses objectives et de tests subjectifs qui permettent de valider la performance des améliorations proposées pour la stéréo paramétrique. L'analyse objective effectuée sur la méthode de réduction de débit pour la communication de l'information de phase a d'ailleurs mené à une

optimisation (ou compromis) supplémentaire, où environ un tiers des bits économisés sont réassignés à un autre usage pour ainsi réduire les valeurs extrêmes des erreurs (outliers). Sans analyse objective, la seule réalisation de tests subjectifs aurait démontré l'efficacité des méthodes proposées, mais n'aurait pas permis d'identifier de telles optimisations. Manifestement, il a été démontré que les améliorations proposées dans ce mémoire sont très avantageuses pour une intégration de la stéréo paramétrique autour de codecs modernes comme l'AMR-WB+.

6.2 Directions futures

Il est possible d'identifier au moins deux axes de recherche pour faire suite aux technologies représentant actuellement l'état de l'art. D'abord, la convergence et la mobilité grandissante des applications sonores signifient que les technologies comme le MPEG Surround devront pouvoir s'adapter à des situations variées, comme la simulation sur casque d'écoute. Ensuite, il est aussi primordial d'avoir une approche efficace pour faire progresser la qualité sonore des technologies paramétriques vers la transparence, que ce soit par des approches paramétriques supplémentaires ou par des techniques efficaces pour coder le « résidu » de l'information spatiale.

Parallèlement, une technologie totalement différente de MPEG Surround pourrait réellement encoder l'information spatiale au lieu de « simplement » représenter d'une façon compacte (basée sur la psychoacoustique) des canaux audio discrets. Cette technologie aurait probablement l'avantage de faire abstraction du nombre de canaux à la source et à la reproduction. Par exemple, de l'audio ambiophonique enregistré sur 5 canaux pourrait être reproduit aussi aisément sur un système « 7.1 » à la maison que sur un casque d'écoute stéréo sur la route.

BIBLIOGRAPHIE

[1] J. Herre, K. Brandenburg, et D. Lederer, "Intensity Stereo Coding," au *96th Audio Engineering Society Convention*, Amsterdam, The Netherlands, 1994.

[2] J. Breebaart, S. v. d. Par, et A. Kohlrausch, "Binaural Processing Model Based on Contralateral Inhibition. I. Model Structure," *The Journal of the Acoustical Society of America,* vol. 110, pp. 1074-1088, 2001.

[3] C. Faller et F. Baumgarte, "Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression," au *112th Audio Engineering Society Convention*, München, Germany, 2002.

[4] "Call for Proposals on Spatial Audio Coding," *ISO/IEC JTC1/SC29/WG11 (MPEG), doc. N6455,* Munich, 2004.

[5] E. Schuijers, J. Breebaart, H. Purnhagen, et J. Engdegard, "Low Complexity Parametric Stereo Coding," au *116th Audio Engineering Society Convention*, Berlin, Germany, 2004.

[6] "Enhanced aacPlus General Audio Codec; Encoder specification; Parametric stereo part," *3GPP TS 26.405*, 2004.

[7] J. Breebaart, S. v. d. Par, A. Kohlrausch, et E. Schuijers, "Parametric Coding of Stereo Audio," *EURASIP Journal on Applied Signal Processing,* pp. 1305-1322, 2005.

[8] "Extended AMR Wideband codec; Transcoding functions," *3GPP TS 26.290*, 2004.

[9] J. Lapierre et R. Lefebvre, "On Improving Parametric Stereo Audio Coding," au *120th Audio Engineering Society Convention*, Paris, France, 2006.

[10] "Enhanced aacPlus General Audio Codec; General Description," *3GPP TS 26.401*, 2004.

[11] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. Amsterdam; Boston: Academic Press, 2003.

[12] E. Zwicker et H. Fastl, *Psychoacoustics: Facts and Models*, 2nd updated ed. Berlin; New York: Springer, 1999.

[13] W. A. Yost, *Fundamentals of Hearing: An Introduction*, 4th ed. San Diego, California: Academic Press, 2000.

[14] C. C. Wier, W. Jesteadt, et D. M. Green, "Frequency Discrimination as a Function of Frequency and Sensation Level," *The Journal of the Acoustical Society of America,* vol. 61, pp. 178-184, 1977.

[15] J. Blauert, *Spatial Hearing: the Psychophysics of Human Sound Localization*, Rev. ed. Cambridge, Mass.: MIT Press, 1997.

[16] S. Carlile, *Virtual Auditory Space: Generation and Applications*. New York; Berlin: Springer-Verlag, 1996.

[17] R. H. Gilkey et T. R. Anderson, *Binaural and Spatial Hearing in Real and Virtual Environments*. Mahwah, N.J.: Lawrence Erlbaum Associates, 1997.

[18] B. C. J. Moore, "Controversies and Mysteries in Spatial Hearing," au *16th Audio Engineering Society International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, 1999, pp. 249–256.

[19] R. O. Duda, C. Avendano, et V. R. Algazi, "An Adaptable Ellipsoidal Head Model for the Interaural Time Difference," au *IEEE ICASSP*, Phoenix, AZ, USA, 1999, pp. 965-968.

[20] J. D. Johnston, J. Herre, M. Davis, et U. Gbur, "MPEG 2 NBC Audio–Stereo and Multichannel Coding Methods," au *101st Audio Engineering Society Convention*, Los Angeles, CA, USA, 1996.

[21] C. R. Cave, "Perceptual Modelling for Low-Rate Audio Coding," *M.Eng. Thesis*, McGill University, *Department of Electrical & Computer Engineering*, Montréal, QC, Canada: McGill University, 86 p., 2002.

[22] D. J. M. Robinson, "Perceptual Model for Assessment of Coded Audio," *Ph.D. Thesis*, University of Essex, *Department of Electronic Systems Engineering*, Colchester, Essex, UK: University of Essex, 321 p., 2002.

[23] L. A. Jeffress, "A Place Theory of Sound Localization," *Journal of Comparative and Physiological Psychology,* vol. 41, pp. 35–39, 1948.

[24] S. H. Nielsen, G. Stoll, et L. v. d. Kerkhof, "Perceptual Coding of Matrixed Audio Signals," au *96th Audio Engineering Society Convention*, Amsterdam, The Netherlands, 1994.

[25] W. R. T. Ten Kate, "Compatibility Matrixing of Multichannel Bit-Rate Reduced Audio Signals," au *96th Audio Engineering Society Convention*, Amsterdam, The Netherlands, 1994.

[26] R. G. v. d. Waal et R. N. J. Veldhuis, "Subband Coding of Stereophonic Digital Audio Signals," au *IEEE ICASSP*, Toronto, ON, CA, 1991, pp. 3601-3604.

[27] J. D. Johnston et A. J. Ferreira, "Sum-difference Stereo Transform Coding," au *IEEE ICASSP*, San Francisco, CA, USA, 1992, pp. 569-572.

[28] J. Herre et J. D. Johnston, "Exploiting Both Time and Frequency Structure in a System That Uses an Analysis/Synthesis Filterbank with High Frequency Resolution," au *103rd Audio Engineering Society Convention*, New York, NY, USA, 1997.

[29] J. Herre, E. Eberlein, et K.-H. Brandenburg, "Combined Stereo Coding," au *93rd Audio Engineering Society Convention*, San Franscisco, CA, USA, 1992.

[30] C.-M. Liu, W.-C. Lee, et Y.-H. Hsiao, "M/S Coding Based on Allocation Entropy," au *Int. Conf. on Digital Audio Effects (DAFX-03)*, London, UK, 2003.

[31] S. Torres-Guijarro, J. A. B. Álava, L. I. Ortiz-Berenguer, et F. J. Casajús-Quirós, "Multichannel Audio Decorrelation for Coding," au *6th Int. Conf. on Digital Audio Effects (DAFX-03)*, London, UK, 2003.

[32] D. Yang, H. Ai, C. Kyriakakis, et C.-C. J. Kuo, "An Inter-Channel Redundancy Removal Approach for High-Quality Multichannel Audio Compression," au *109th Audio Engineering Society Convention*, Los Angeles, CA, USA, 2000.

[33] Y. Wang, L. Yaroslavsky, M. Vilermo, et M. Vaananen, "A Multichannel Audio Coding Algorithm for Inter-Channel Redundancy Removal," au *110th Audio Engineering Society Convention*, Amsterdam, The Netherlands, 2001.

[34] D. Bauer et D. Seitzer, "Statistical Properties of High Quality Stereo Signals in the Time Domain," au *IEEE ICASSP*, Glasgow, Scotland, 1989, pp. 2045-2048.

[35] D. Bauer et D. Seitzer, "Frequency Domain Statistics of High Quality Stereo Signals," au *86th Audio Engineering Society Convention*, Hamburg, Germany, 1989.

[36] S.-S. Kuo et J. Johnston, "A Study of Why Cross Channel Prediction is Not Applicable to Perceptual Audio Coding," au *111st Audio Engineering Society Convention*, New York, NY, USA, 2001.

[37] H. Fuchs, "Improving Joint Stereo Audio Coding by Adaptive Inter-Channel Prediction," au *IEEE ASSP Workshop*, New Paltz, NY, USA, 1993.

[38] A. Härmä et U. K. Laine, "A Comparison of Warped and Conventional Linear Predictive Coding," *IEEE Transaction on Speech and Audio Processing,* vol. 9, pp. 579-588, 2001.

[39] A. C. d. Brinker, V. Voitishchuk, et S. J. L. v. Eijndhoven, "IIR-Based Pure Linear Prediction," *IEEE Transaction on Speech and Audio Processing,* vol. 12, pp. 68-75, 2004.

[40] A. Aggarwal, S. L. Regunathan, et K. Rose, "Optimal Prediction in Scalable Coding of Stereophonic Audio," au *109th Audio Engineering Society Convention*, Los Angeles, CA, USA, 2000.

[41] C.-M. Liu, W.-C. Lee, et S. Y. Juang, "Design of the Coupling Schemes for the Dolby AC-3 Coder in Stereo Coding," au *IEEE International Conference on Consumer Electronics*, Los Angeles, CA, USA, 1998, pp. 328-329.

[42] J. Breebaart, S. v. d. Par, A. Kohlrausch, et E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates," au *116th Audio Engineering Society Convention*, Berlin, Germany, 2004.

[43] H. Purnhagen, "Low Complexity Parametric Stereo Coding in MPEG-4," au *7th Int. Conf. on Digital Audio Effects (DAFX-04)*, Naples, Italy, 2004.

[44] F. Baumgarte et C. Faller, "Why Binaural Cue Coding is Better than Intensity Stereo Coding," au *112th Audio Engineering Society Convention*, München, Germany, 2002.

[45] C. Faller et F. Baumgarte, "Binaural Cue Coding - Part II : Schemes and Applications," *IEEE Transactions on Speech and Audio Processing,* vol. 11, pp. 520-531, 2003.

[46] S. Quackenbush et J. Herre, "MPEG Surround; What's New with MPEG?," *IEEE Multimedia,* vol. 12, pp. 18-23, 2005.

[47] J. Breebaart, J. Herre, C. Faller, J. Rödén, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K. Kjorling, et W. Oomen, "MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status," au *119th Audio Engineering Society Convention*, New York, NY, USA, 2005.

[48] J. Herre, S. Church, M. Dietz, C. Faller, F. Lott, W. Oomen, et H. Purnhagen, "Workshop W9 - MPEG Surround - Recent Progress in Parametric Coding of Multichannel Audio," au *120th Audio Engineering Society Convention*, Paris, France, 2006.

[49] J. F. Culling, H. S. Colburn, et M. Spurchise, "Interaural Correlation Sensitivity," *Journal of the Acoustical Society of America,* vol. 110, pp. 1020-1029, 2001.

[50] "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)," *ITU-R Recommendation BS. 1543-1,* Geneva, Switzerland, 2001.

[51] D. Schobben et S. v. d. Par, "The Effect of Room Acoustics on MP3 Audio Quality Evaluation," au *117th Audio Engineering Society Convention*, San Francisco, CA, USA, 2004, p. 6.