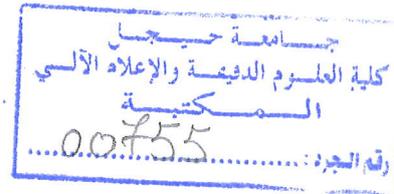


République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Seddik Benyahia de Jijel  
Faculté des Sciences Exactes et informatique  
Département d'Informatique

1  
2

inf. LM, 07/19



*Mémoire de fin d'étude*  
*pour l'obtention du diplôme Master*  
*de Recherche en Informatique*  
Option : *Informatique Légale et Multimédia*

Thème

Evaluation de la qualité des données  
audiovisuelles

Présentées par :  
AYACHE Nihad.  
ABDOU Amina.

Encadrée par :  
BOUDJERIDA Fatima



Promotion : 2019.

## *\* Remerciements \**

Tout d'abord , je tiens à remercier en premier lieu le Dieu le tout puissant et miséricordieux, qui m'a donné la force et la patience d'accomplir ce Modeste travail.

Un sincère et honnête merci à nos parents et nos frères et soeurs pour leur soutien indéfectible qu'ils savent nous l'apporter tout au long de nos études et en particulier pendant cette mémoire.

Ce travail est réalisé pour obtenir le diplôme de master , spécialité Informatique légale multimédia au département de l'informatique , université de Jijel.

Nous avons témoigné ici notre respectueuse reconnaissance et remerciement très sincèrement à Mme F.BOUDJRIDA enseignante au département de l'informatique à l'Université de Jijel pour l'intérêt qu'elle a apporté à notre travail, et pour les conseils qu'elle nous a donné et pour sa patience au cours d'encadrement.

Nous remercions vivement les membres de jury.

Nous remercions également tous les enseignants du département de l'informatique et spécialement les enseignants qui apportent cette spécialité à l'Université de Jijel.

Nous remercions tous nos collègues et amis qui ont partagés deux années agréable , avec une ambiance éducative inoubliable.

Nous remercions tous les membres de département de l'informatique qui font pendant cinq années nous informer et guider concernant nos intérêts préoccupations et administrative.

Enfin, nous remercions tous ceux qui ont participé dans la réalisation de ce travail de près ou de loin.

Merci pour tous...

## *\* Dédicaces \**

*Je dédie ce modeste travail à :*

A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir . Grâce à toi papa j'ai appris le sens du travail et de la responsabilité. Je voudrais te remercier pour ton amour , ta générosité , ta compréhension... Ton soutien fut une lumière dans tout mon parcours.

*\* Mon père Azzaddine que j'adore \**

A la lumière de mes jours , la source de mes efforts, la flamme de mon coeur , ma vie et mon bonheur . Aucune dédicace ne saurait exprimer l'amour l'estime et le respect que j'ai toujours eu pour toi.

*\* Ma mère Ghania je t'aime \**

A tous les moments d'enfance passés avec toi mon frère, en gage de ma profonde estime pour l'aide que tu m'as apporté . Tu m'as soutenu , réconforté et encouragé . Puissent nos liens fraternels se consolider et se pérenniser encore plus.

*\* Mon cher frère Bilal \**

Tu as toujours offert soutien et réconfort , j'exprime envers toi une profonde admiration , reconnaissance et attachement inconditionnels . Que Dieu le tout puissant te protège et te réserve un bon avenir.

*\* Mon adorable frère Mohamed Noor \**

En témoignage de mon affection fraternelle, de ma profonde tendresse et reconnaissance , je vous souhaite une vie pleine de bonheur et de succès et que Dieu , le tout puissant , vous protège et vous garde à vous mes chères et adorable soeurs .

*\* Soumia la prunelle de mes yeux , Amina la douce  
au coeur si grand , Assia mon idole \**

Avoir un neveu est le plus beau cadeau qu'une soeur puisse vous faire. Tes petites mains , ton envie de parcourir le monde, ton enthousiasme, tes sourires, tes yeux brillants sont incomparables . Tu as apporté beaucoup de bonheur à notre famille . Je t'aime

*\* Mon neveu Iyad "Didou" \**

Aucun langage ne saurait exprimer mon respect et ma considération pour votre soutien et encouragements. Je vous dédie ce travail en reconnaissance de l'amour que vous m'offrez quotidiennement et votre bonté exceptionnelle . Que Dieu le Tout Puissant vous garde et vous procure santé et bonheur.

*\* A toute la famille Ayache et Benayade \**

**A mon binôme Amina** qui je lui considère comme beaucoup plus une soeur que comme mon amie et toute la famille **ABDOU**.

A mes amies **Rima , Yousra , Lyticia , Ilhem , Houdhoud , Selma** Je ne peux trouver les mots justes et sincères pour vous exprimer mon affection et mes pensées , vous êtes pour moi des soeurs et des amies sur qui je peux compter . En témoignage de l'amitié qui nous unit et des souvenirs de tous les moments que nous avons passés ensemble , je vous dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur .

Et à tous ceux qui ont contribué de près ou de loin pour que ce travail soit possible , je vous dis merci.

*\* Nihad \**

*\* Dédicaces \**

Du profond de mon coeur je dédie ce travail à tous ceux qui me sont chers :

*\* A mon cher père Abdeslam \**

Ce travail a dédié a mon père , qui ma toujours poussé et motivé dans mes études . Pour le lourd sacrifice afin de voir sa famille unie et heureuse , il a donné sa belle vie . Pour tous les moments où tu n'as jamais épargné le moindre effort pour nous aider et nous encourage.

Merci à vous d'être toujours présent pour faire mon bonheur.

*\* A ma chère mère Fakia \**

Pour les heures pleines de bonheur et de tendresse , pour les chaudes larmes qui hantent ma mémoire . Pour l'ampleur de tes sacrifices.

Pour toi ma chère maman qui mis l'avenir de tes enfants en premiers et les jouissances de la vie terrestre en dernier . Vous êtes ce que j'ai de plus chère au monde et je vous jure qu'aucun mot, ni expression ne saurais exprimer tous mon amour et toute ma gratitude . Merci de faire l'importance pour moi.

*\* A mon adorable soeur Houda \**

Nulle dédicace ne saurait exprimer mon amour et ma profonde affection.

Aux moments inoubliable que nous avons passé ensemble , tu as toujours été là et où il faut et quand il faut.

Les phrases me manquent pour t'exprimer ma grande connaissance et mon admiration profonde.

Que Dieu le tous puissant te protège et te réserve un bon avenir.

*\* A mes frères Salah et Nadjmeddine \**

Veillez trouvez dans ce travaille un modeste témoignage de mon admiration et de toute ma gratitude , de mon affection la plus sincère et de mon attachement le plus profond.

*\* A ma famille Abdou et Ayache \**

A mes chère tantes , mes oncles et mes cousins et cousines . Votre confiance en moi , vos encouragement, vos prière sont ce qui m'a poussé et me pousse toujours à suivre la voie de l'excellence , à rêvé et à réaliser mes rêves.

*\* A mes amies \**

**Nihad** qui je lui considère comme beaucoup plus une soeur que comme mon amie , **Rima** a qui ma connaissances dépasse la porte des mots , **Youssra** la sage qui ne cesse de me pardonner, **Kounayza** qui ma toujours comblé d'amour et d'affection , **Bachira** et **Latifa** mon idole , **Ilham** et **Selma** mes confidentes , **Faiza** et **Kenza** mes deux petite soeurs . Je vous souhaite tout un avenir plein de succès.

Sous peine de ne pas mentionner une personne, ce travail est dédicace a tous les gens qui m'ont encouragé.

*\* Amina \**

# Table des matières

<b>Table des matières</b>	<b>1</b>
<b>Liste des tableaux</b>	<b>3</b>
<b>Table des figures</b>	<b>5</b>
<b>Liste des abréviations</b>	<b>6</b>
<b>Introduction générale</b>	<b>8</b>
<b>1 Audiovisuel et la qualité perçus</b>	<b>10</b>
1.1 Définition audiovisuel . . . . .	11
1.2 Les domaines de l'utilisation de l'audiovisuel . . . . .	11
1.2.1 Dans le domaine de l'enseignement/apprentissage . . . . .	11
1.2.2 Dans le domaine de l'autorité . . . . .	11
1.2.3 Dans le domaine de l'art . . . . .	11
1.3 Les composants de base de l'audiovisuel . . . . .	12
1.3.1 La vidéo . . . . .	12
1.3.1.1 Différents types de dégradations . . . . .	12
1.3.2 L'audio . . . . .	15
1.3.2.1 Caractéristiques du l'audio . . . . .	16
1.3.2.2 Types de formats audio . . . . .	17
1.4 La qualité . . . . .	18
1.4.1 La qualité de l'expérience (QoE) . . . . .	18
1.5 Evaluation de la qualité . . . . .	19
1.5.1 Evaluation de la qualité subjective . . . . .	19
1.5.2 Evaluation de la qualité objective . . . . .	19
<b>2 Evaluation de la qualité multimédia (video , audio , audiovisuel)</b>	<b>20</b>
2.1 Les types d'évaluation de la qualité . . . . .	20
2.2 Evaluation subjective . . . . .	20
2.2.1 Evaluation subjective de la vidéo . . . . .	21
2.2.1.1 Facteurs influents . . . . .	21

2.2.1.2	Protocoles . . . . .	24
2.2.1.3	Analyse des résultats . . . . .	27
2.2.2	Evaluation subjective de l'audio . . . . .	28
2.2.2.1	La méthode MUSHRA :MULTiple Stimuli with Hidden Reference and Anchor . . . . .	28
2.2.3	Evaluation subjective de l'audiovisuel . . . . .	29
2.2.3.1	Méthode ACR : ABSOLUTE CATEGORY RATING . . . . .	29
2.2.3.2	MÉTHODE DCR : DEGRADATION CATEGORY RATING . . . . .	30
2.2.3.3	METHODE PC : PAIR COMPARISON . . . . .	31
2.2.3.4	Comparaison entre les différentes méthodes de l'évaluation de la qualité audiovisuelle . . . . .	32
2.3	Evaluation objective . . . . .	32
2.3.1	Evaluation objective de la vidéo . . . . .	32
2.3.1.1	Métriques à référence complète (FR) . . . . .	32
2.3.1.2	Métriques sans référence (NR) : . . . . .	36
2.3.1.3	Métriques à référence réduite (RR) : . . . . .	38
2.3.2	Evaluation objective de l'audio : . . . . .	39
2.3.2.1	Méthode intrusive : . . . . .	39
2.3.2.2	Méthode non intrusive : . . . . .	42
2.3.2.3	Méthodes Paramétriques . . . . .	43
2.3.3	Evaluation objective de l'audiovisuel : . . . . .	43
<b>3</b>	<b>Evaluation objective de la qualité audiovisuel</b> . . . . .	<b>46</b>
3.1	Les méthodes d'évaluation objectives de la qualité perçue audiovisuel : . . . . .	46
3.1.1	L'approche de la fusion des deux modalités : . . . . .	46
3.1.2	L'approche de prévision de la qualité audiovisuelle : . . . . .	48
3.1.3	L'approche d'apprentissage automatique : . . . . .	49
<b>4</b>	<b>La méthode choisie , les résultats , les discussions</b> . . . . .	<b>54</b>
4.1	Environnement de travail . . . . .	54
4.1.1	Langage . . . . .	54
4.1.2	Caractéristique de la plateforme . . . . .	54
4.2	La base de donnée utilisé : . . . . .	55
4.2.1	L'ensemble de données sur la qualité audiovisuelle de l'INRS : . . . . .	55
4.3	Description de notre méthode : . . . . .	57
4.4	Analyse et interprétation des résultats . . . . .	60
	<b>Conclusion générale</b> . . . . .	<b>65</b>
	<b>Bibliographie</b> . . . . .	<b>67</b>

# Liste des tableaux

2.1	Exemple d'échelle de notation . . . . .	25
2.2	Echelle d'évaluation de qualité video . . . . .	27
4.1	Métrique de référence complète[57] . . . . .	57
4.2	Comparaison entre les méthodes de la qualité audiovisuel et notre méthode .	64

# Table des figures

1.1	Exemple du l'effet de bloc	12
1.2	Exemple de flou	13
1.3	Exemple de bruit	13
1.4	Exemple de GOP	15
1.5	L'évolution de l'amplitude sonore dans le temps	16
1.6	Les types de fréquence	17
2.1	Présentation des séquences d'évaluation	22
2.2	test d'ordonnancement	22
2.3	test de choix forcé	23
2.4	test Best/Worst	23
2.5	Test d'Ishihara	24
2.6	Schéma d'un système de DSCQS	25
2.7	Schéma d'un système de DSIS	26
2.8	Schéma d'un système de SSCQE	26
2.9	Exemple d'un affichage sur ordinateur utilisé pour un test MUSHRA	28
2.10	Echelle d'évaluation de qualité à 9 et 5 niveaux	29
2.11	Chronogramme de la méthode ACR	30
2.12	Chronogramme de la méthode DCR	30
2.13	Echelle de dégradation à cinq niveaux	31
2.14	Chronogramme de la méthode PC	31
2.15	Schéma de métrique à référence complète	33
2.16	Schéma de métrique sans référence	36
2.17	Schéma de métrique à référence réduite	38
2.18	Principe de la PEAQ	40
3.1	Modèle d'estimation de la qualité multimédia de base	47
4.1	Dossier MOS	55
4.2	Fichier FileNameWithMos	56
4.3	Fichier Base Dataset	56
4.4	Dossier Subject Details	57
4.5	Résultats obtenues de PSNR et SSIM pour un vidéo de la base INRS	60

---

4.6	Résultats obtenues de PSNR et SSIM pour un audio de la base INRS . . . . .	60
4.7	Histogramme de a0 . . . . .	61
4.8	Histogramme de a1 . . . . .	61
4.9	Histogramme de a2 . . . . .	62
4.10	Histogramme de a3 . . . . .	62
4.11	Aperçus sur le fichier text des résultats obtenue . . . . .	63

# Liste des abréviations

<b>QoE</b>	Quality of Experience
<b>MOS</b>	Mean Opinion Score
<b>AV</b>	Audio Visuel
<b>JPEG</b>	Joint Photographic Experts Group
<b>MPEG</b>	Moving Picture Experts Group
<b>GOP</b>	Group of pictures
<b>AVI</b>	Audio Video Interleave
<b>MP4</b>	MPEG Layer-4 Audio
<b>ASF</b>	Advanced Systems Format
<b>AC3</b>	Audio Coding 3
<b>MP3</b>	MPEG-1/2 Audio Layer 3
<b>RAW</b>	Real Audio Wrapper
<b>BWF</b>	Broadcast Wave Format
<b>CAF</b>	Core audio format
<b>QA</b>	Qualité Audio
<b>QV</b>	Qualité Vidéo
<b>UIT</b>	Union Internationale des Télécommunications
<b>DSCQS</b>	Double Stimulus Continuous Quality Scale
<b>DSIS</b>	Double Stimulus Impairment Scale
<b>EQM</b>	Erreur Quadratique Moyenne
<b>PSNR</b>	Peak Signal to Noise Ratio
<b>MUSHRA</b>	MUltiple Stimuli with Hidden Reference and Anchor
<b>ACR</b>	ABSOLUTE CATEGORY RATING
<b>DCR</b>	DEGRADATION CATEGORY RATING
<b>PC</b>	PAIR COMPARISON
<b>SSCQE</b>	SINGLE-STIMULUS CONTINUOUS QUALITY EVALUATION

- WAV** Waveform Audio
- SSIM** Structural SIMilarity
- SSIM** Structural SIMilarity
- SVH** Système Visuel Humain
- PESQ** Perceptual Evaluation of Speech Quality
- PEAQ** Perceptual Evaluation of Audio Quality
- NSS** Natural Scene Statistics
- PCC** Pearson Coefficient Corrélation
- BIQI** Blind Image Quality Index
- SCC** Spearman Corrélation Coefficient
- DIVIINE** Distortion Identification-based Image Verity and INtegrity Evaluation

# Introduction générale

L'évolution récente des systèmes de communication numériques (3G et 4G, par exemple) a entraîné une explosion de services et d'applications multimédias, tels que la télévision IP, le multimédia mobile sur smartphones, les réseaux sociaux (Facebook, par exemple), le multimédia immersif et les jeux de réalité virtuelle, vidéoconférence et présentations multimédias éducatives, pour n'en nommer que quelques-unes. Ces applications multimédias font désormais partie intégrante (si ce n'est indispensable) de la vie quotidienne et devraient connaître une croissance exponentielle supplémentaire. Les fournisseurs de services multimédias élaborent diverses techniques pour offrir une meilleure qualité d'expérience (QoE), qui est de plus en plus demandée par les utilisateurs finaux. Par conséquent, l'opinion de l'homme sur la qualité est essentielle à la conception et au déploiement de tous les réseaux et services multimédias actuels et futurs .

L'audio et la vidéo sont deux modalités de base dans la plupart des applications multimédia. Malgré les progrès récents, les signaux audiovisuels souffrent de dégradations dues à la fois au codage source avec pertes et à la transmission sur des canaux sujets aux erreurs, entraînant ainsi une dégradation de la qualité du signal multimédia . Par exemple , de la qualité audiovisuel , un échantillon vidéo reçu par l'utilisateur final peut posséder une large plage de qualité en raison de différentes erreurs de transmission ou de rendu. Une qualité estimée avec précision des signaux audiovisuels transmis peut contribuer de manière considérable aux services multimédia et aux réseaux de communication. En fait, l'évaluation de la qualité des signaux numériques est l'un des problèmes fondamentaux et complexes du traitement multimédia et de ses situations pratiques, telles que l'évaluation et la mise en oeuvre de processus, l'optimisation du codage et du décodage, les tests et la surveillance (par exemple en transmission et en fabrication des sites). De plus, comment évaluer la qualité audio et vidéo joue un rôle central dans la conception de la plupart (sinon de la totalité) des services, algorithmes et systèmes multimédias . L'acquisition , la synthèse , l'amélioration , la compression , le filigranage , le stockage, la récupération, la restitution, le rendu et la présentation de signaux (par exemple, l'affichage sur un appareil mobile) sont des exemples de dépendance technologique vis-à-vis de l'évaluation de la qualité audiovisuelle.

Dans ce mémoire nous nous intéressons à l'évaluation de la qualité des données audiovisuelle qui sont recommandés par l'Union Intentionnelle de Télécommunication .

☞ Dans ce contexte, nous avons organisé notre mémoire en quatre parties principales, la première partie présente une vue panoramique sur l'audiovisuel et sa qualité, ainsi qu'une perspective détaillée sur la vidéo et l'audio, ensuite nous avons abordé l'évaluation de la qualité multimédia (vidéo, audio et audiovisuel), puis nous avons choisi une méthode pour la qualité audiovisuelle et nous terminons par une comparaison entre l'évaluation subjective et objective. Plus précisément, ce travail est composé des chapitres suivants :

- ✓ **Le premier chapitre** est consacré à une introduction sur l'audiovisuel tout en passant par ces composants qui sont la vidéo et l'audio, et pour mieux assimiler ce thème nous allons parler de l'évaluation subjective et l'évaluation objective.
- ✓ **Le deuxième chapitre** est dédié à l'évaluation de la qualité vidéo, la qualité audio et la qualité audiovisuelle, où on s'est penché sur les différentes méthodes de l'évaluation subjective et les différentes méthodes de l'évaluation objectives.
- ✓ **Le troisième chapitre** est destiné à l'évaluation objective de la qualité audiovisuelle et les approches utilisées pour cette dernière.
- ✓ **Le quatrième chapitre** est consacré à la description de la méthode choisie pour l'évaluation de la qualité audiovisuelle on prêter attention aux différentes étapes que nous devons suivre pour obtenir une évaluation, ainsi qu'une comparaison entre les résultats obtenus (évaluation objective) et les résultats qui ont déjà été réalisés (évaluation subjective).

Enfin, nous terminons par une conclusion dans laquelle nous dressons un bilan et une synthèse du travail effectué, et nous discutons un ensemble de perspectives qui peuvent être considérées comme des directions de recherche future.

## Chapitre 1

# Audiovisuel et la qualité perçus

## Introduction

Au cours des années 1920 à 1940, le disque 78 tours, le cinéma noir et blanc (muet puis parlant) et la radiodiffusion sonore ouvrent la voie de ce que l'on appellera plus tard l'audiovisuel.

Les années 1950 à 1970 voient se développer le disque microsillon, le cinéma et la photo en couleurs, la radiodiffusion sonore en modulation de fréquence monophonique puis stéréophonique, la télévision en noir et blanc puis en couleurs, et la bande magnétique audio en bobines puis en cassettes. Le transistor permet de développer des appareils autonomes à piles.

Les années 1970 à 1980 sonnent le glas du cinéma d'amateur sur film (16 mm, 8 mm et super-8), définitivement remplacé par le magnétoscope. Le circuit intégré permet de réaliser économiquement des dispositifs complexes : c'est le début de l'informatique grand public.

Les années 1980 à 1990 sont surtout marquées par la télévision par satellites, le développement des techniques audio numériques avec le disque compact (CD en abrégé), et la miniaturisation des caméras et magnétoscopes désormais fusionnés dans le caméscope.

Les années 1990 s'orientent vers une diversification des formats numériques audio sur disques et bandes magnétiques, et une extension du format CD à l'image (CD vidéo, CD photo) et aux informations de toutes sortes (CD-ROM, CD-I).

Désormais, on peut associer l'ordinateur à des analyseurs d'images, des micros ou des caméras, des mémoires, des imprimantes, des modulateurs de transmission, des écrans, des haut-parleurs, etc., sans oublier les logiciels et les liaisons avec des banques de données. Ainsi secondé, il devient appareil de saisie de l'information, enregistreur, machine de traitement, machine de création ou d'aide à la création, appareil de diffusion de tous messages, audiovisuels et autres, en bref : machine à communiquer. Les supports traditionnels (papier,

disque, bande) peuvent être remplacés par un support unique et universel, la mémoire de l'ordinateur, ce que tente d'exprimer le vocable « multimédia ». Le seul facteur de limitation à ces techniques réside pour le moment dans l'énorme capacité de mémoire exigée par les images . [1]

## **1.1 Définition audiovisuel**

Le signal audiovisuel est doublement composé d'un signal vidéo et d'un signal audio, donc audiovisuel sert à désigner tout ce qui est relatif à l'image et/ou au son. Les fichiers audiovisuels s'agit de toutes les formes d'enregistrement du son et/ou des images animées et/ou des images fixes .

## **1.2 Les domaines de l'utilisation de l'audiovisuel**

### **1.2.1 Dans le domaine de l'enseignement/apprentissage**

L'emploi des supports audiovisuels a plusieurs avantages pour l'enseignement/apprentissage . Nous pouvons en citer les suivants :

- Le professeur possède la liberté et la responsabilité pour organiser le contenu audiovisuel.
- L'enseignant a la possibilité de présenter ce contenu avec des moyens didactiques et pédagogiques appropriés.
- L'audiovisuel facilite la mémorisation. [2]

### **1.2.2 Dans le domaine de l'autorité**

La possibilité d'enregistrer des images et du son a naturellement attiré l'attention des autorités militaires . À partir de la fin du XIXe siècle , l'audiovisuel sert à la fois pour les opérations de renseignement militaire et pour la propagande.

### **1.2.3 Dans le domaine de l'art**

L'art audiovisuel regroupe l'ensemble des pratiques artistiques de l'audiovisuel : diaporama , art vidéo.

## 1.3 Les composants de base de l'audiovisuel

### 1.3.1 La vidéo

Une vidéo est une succession des frames à une certaine cadence, qui peut être caractérisée par trois principaux paramètres :

- Résolution en luminance : détermine le nombre de nuances ou de couleurs possibles pour un pixel. Celle-ci est généralement de 8 bits pour les niveaux de gris et de 24 bits pour les séquences en couleurs.
- Résolution spatiale : définit le nombre de lignes et de colonnes de la matrice de pixels.
- Résolution temporelle : est le nombre d'images par seconde.

#### 1.3.1.1 Différents types de dégradations

##### Effets de bloc

Dans la compression des images (video), les blocs sont encodés indépendamment les uns des autres. Après le processus de quantification, chaque bloc sera représenté par un nombre limité de coefficients non nuls. La dé quantification permet de choisir les valeurs des coefficients parmi un ensemble limité de valeurs. Ce type de codage engendre une discontinuité à la frontière entre deux blocs et apparaît comme des contours verticaux et horizontaux. Cette discontinuité est facilement remarquable par l'oeil humain surtout dans le cas d'une dégradation de couleurs. Cette dégradation présente une principale source d'erreur appelée effet de bloc. [3]



FIGURE 1.1 – Exemple du l'effet de bloc

##### Flou

Le lissage des contours présente une solution efficace au problème des effets de blocs. Cependant, cette technique affecte considérablement la qualité de l'image (vidéo) et

provoque une diminution de contraste et l'image devient floue. Cet artefact peut être provoqué aussi par le processus de compression d'image (vidéo) à cause de l'utilisation d'un filtrage des hautes fréquences. Ce filtrage conduit à une perte de netteté qui se traduit par un flou.

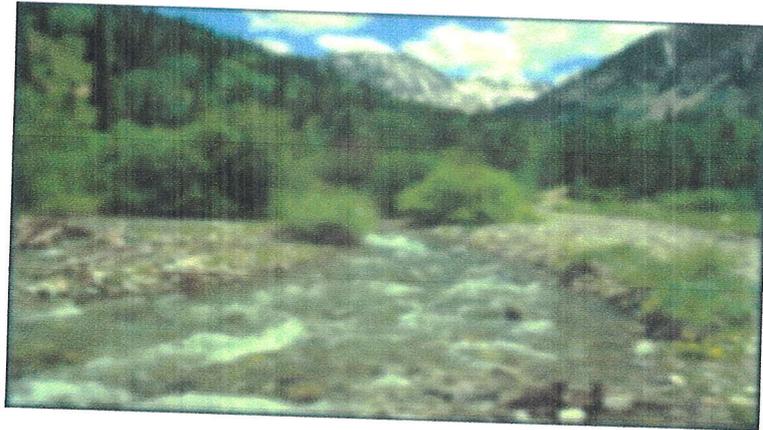


FIGURE 1.2 – Exemple de flou

### Bruit

Dans le cas d'une transmission d'image (ou vidéo), on peut définir le bruit [4] comme étant une dégradation dans l'image (ou vidéo), provoquée par une perturbation externe. Si une image (ou vidéo) est envoyée par voie électronique d'un endroit à un autre, via le satellite ou la transmission sans fil, ou par le câble réseau, on peut s'attendre à des erreurs se produisant dans le signal de l'image (vidéo). Ces erreurs apparaissent sur la vidéo de sortie de différentes manières en fonction du type de la perturbation dans le signal.



FIGURE 1.3 – Exemple de bruit

## La compression vidéo

Quel que soit le dispositif que vous utilisez , carte d'acquisition ou caméscope numérique, dans la plupart des cas , la numérisation de votre vidéo entraîne également sa compression. La compression est nécessaire compte tenu du volume important de données que représente une vidéo dans son état initial.[5]

L'objectif de la compression est de réduire le volume de données tout en conservant une bonne qualité d'image.

Pour réaliser la compression/décompression d'un fichier vidéo, l'ordinateur a besoin d'un logiciel appelé Codec et le fichier vidéo contient des frames , du son et du texte (méta-données) placés dans un conteneur.

**codec** : Un codec est un algorithme de compression / décompression d'un signal audiovisuel numérique [6] , les formats de ces codecs :

**MPEG** : Format de fichier audio et vidéo utilisant la compression avec perte. Utilisé pour la diffusion continu sur Internet (le streaming). Dans une séquence vidéo codée en MPEG on trouve 3 types de codage des images :

Image I : en Intra.

Image P (prédite , forward prediction ) : par rapport à une image passée .

Image B (bidirectionnelle , forward prediction et backward prediction) : par rapport à une image passée et une image future. Le codeur transmet la moyenne des erreurs et transmet les 2 vecteurs.

A l'intérieur d'une image P : on trouve des macroblocs I ou P.

A l'intérieur d'une image B : on trouve des macroblocs I , P ou B.

Pour mieux comprendre c'est trois types de codage des images on trouve le terme GOP .

**GOP** : Définit l'enchaînement des images I, P et B dans le flux vidéo , commence toujours par une image I , il a 2 paramètres M et N noté  $GOP(M,N)$  , où M est la distance entre deux images P et N la distance entre deux images I.

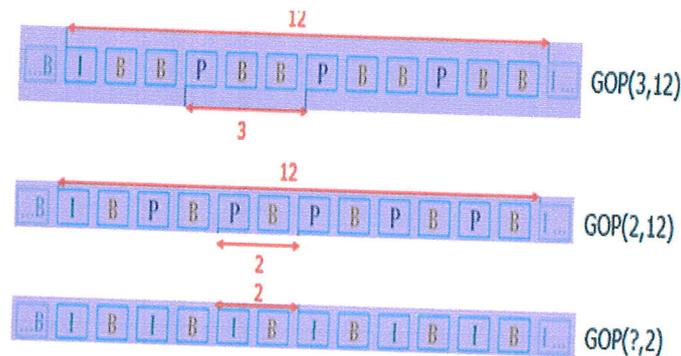


FIGURE 1.4 – Exemple de GOP

**MPEG-1** : Première norme audio et vidéo utilisé plus tard pour les Vidéo CD. Ce format offre une résolution à l'écran de 352 x 240 pixels à 30 images par seconde ou de 352 x 288 à 25 images par seconde avec un débit d'environ 1,5 Mbit/s.

**MPEG-2** : Norme applicable au codage de l'audio et la vidéo , ainsi que leur transport pour la télévision numérique et pour les vidéo disques DVD.

**Le conteneur** : Les conteneurs est un format de fichier qui contient les flux audio et vidéo , l'information codec et les méta-données , les formats de ces conteneurs sont :

**AVI** : Le format AVI a été conçu en novembre 1992 par Microsoft. Il est principalement utilisé pour le stockage et la transmission de films , émissions télévisées et autres vidéos. La popularité des fichiers AVI est due à leur excellente compatibilité avec divers systèmes d'exploitation. En effet, le format AVI n'est pas seulement lisible sous Windows , mais également sous Mac , Linux et Unix , et ce avec la plupart des lecteurs multimédia.

**3GP** : Est un conteneur vidéo destiné au téléphone mobile de troisième génération (3G). C'est une version simplifiée du MP4. Les fichiers 3GP ont pour extension .3gp ou .3g4.

**ASF** : Les fichiers ASF peuvent aussi contenir des métadonnées telles que le nom de l'artiste, le titre, l'album, l'année, le genre, le réalisateur, au début de son histoire, les fichiers de ce format se présentaient avec l'extension .asf , elle a été depuis remplacée par les extensions .wma et .wmv.

### 1.3.2 L'audio

L'audio est une onde produite par la vibration mécanique d'un support fluide ou solide et propagée grâce à l'élasticité du milieu environnant sous forme d'ondes longitudinales. Par extension physiologique, l'audio désigne la sensation auditive à laquelle cette vibration est

susceptible de donner naissance.[7]

### 1.3.2.1 Caractéristiques du l'audio

#### L'amplitude

La première caractéristique d'un audio est son amplitude. Appelée aussi intensité ou volume sonore, c'est l'expression de la pression de l'air qui se mesure en décibels (dB). 0 dB correspond au minimum que l'oreille humaine puisse percevoir (seuil d'audibilité).[7]

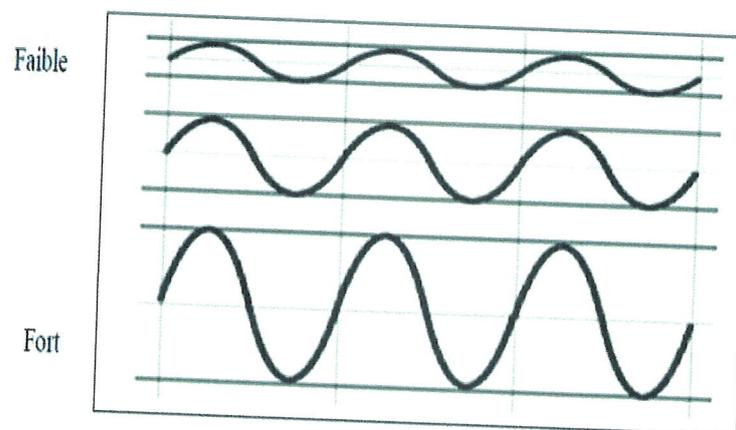


FIGURE 1.5 – L'évolution de l'amplitude sonore dans le temps

#### La Fréquence :

La fréquence, exprimée en Hertz (Hz), est le nombre de répétition d'une période par seconde. Plus elle est élevée et plus le son paraîtra «aiguë», à l'inverse, il paraîtra «grave». En musique, la fréquence définit donc la hauteur d'un audio, soit, la note. (Ex : la note « LA » correspond à 440Hz, soit 440 vibration en une seconde).

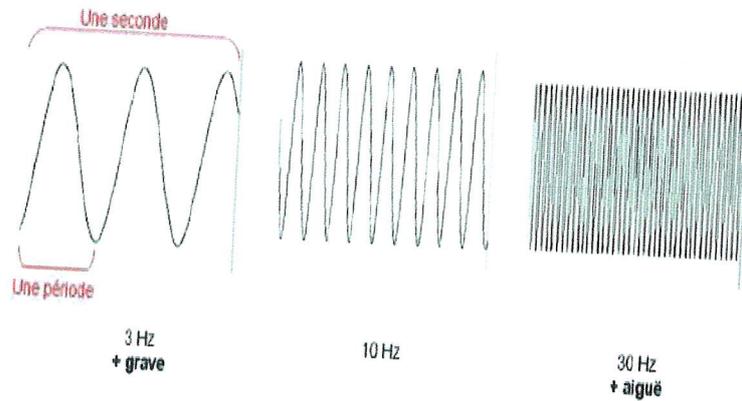


FIGURE 1.6 – Les types de fréquence

### 1.3.2.2 Types de formats audio

#### Les formats audio compressés avec perte

**AC3** : La compression AC3 permet d'utiliser jusqu'à 6 canaux sonores indépendants avec un taux d'échantillonnage de 32, 44,1 ou 48 kHz et avec un taux de transfert allant de 32 à 640 kbit/s. Le Dolby Digital utilise ce principe de codage, c'est pourquoi on le désigne souvent sous ce nom. Format très courant dans les DVD.

**MP3** : MP3 est l'abréviation de MPEG-1/2 Audio Layer 3, La couche (Layer) III est la couche la plus complexe. Elle est dédiée à des applications nécessitant des débits faibles (128 kbit/s) d'où une adhésion très rapide du monde Internet à ce format de compression.

**MP3PRO** : Le format mp3PRO, fruit de la collaboration entre Thomson Multimédia et l'Institut Fraunhofer, combine l'algorithme MP3 et un système améliorant la qualité des fichiers comprimés appelé (en)SBR pour Spectral Bandwidth Replication.

#### Les formats audio compressés sans perte :

**RAW** : RAW est un format audio utilisé pour représenter les données de son en modulation d'impulsion codée sans en-tête ni métadonnées.

**BWF** : Le BWF est un format audio standard permet de stocker des méta-données dans le fichier. Il s'agit du format d'enregistrement usuel utilisé dans de nombreuses stations de travail audio professionnel de la télévision et du cinéma. Les Fichiers BWF incluent une référence standardisée Timestamp qui permet et facilite la synchronisation avec un élément d'image distincte.

**CAF** : Le CAF a été développé par Apple pour s'affranchir des limitations de conteneur audio plus ancien comme le AIFF ou le WAV. Il est compatible avec le système Mac OS X d'Apple depuis la version 10.3 et est lisible par Quicktime 7 .

Dans un contexte fortement concurrentiel, un des principaux enjeux pour les acteurs de l'offre de services audiovisuels (AV) est de garantir une qualité d'expérience (QoE -Quality of Experience-) optimale à l'utilisateur.

## 1.4 La qualité

La qualité est généralement utilisée dans l'optique d'une ingénierie, car elle est un critère essentiel pour évaluer les systèmes, les services ou les applications au cours des phases de conception et d'exploitation. Fondamentalement, la qualité est le résultat d'un jugement humain basé sur divers critères. La qualité perçue, et plus largement la QoE, devient un élément clé qu'il faut par conséquent savoir mesurer.[8]

### 1.4.1 La qualité de l'expérience (QoE)

La QoE est une mesure du jugement personnel de l'utilisateur selon son expérience vécue, sur la qualité globale du service fourni par les opérateurs et fournisseurs de services Internet. En effet, la notion de l'expérience utilisateur a été introduite pour la première fois par le Dr Donald Norman, évoquant l'importance de la conception d'un service centré utilisateur [9]. Gulliver et Ghinea [10] décomposent la QoE en trois composantes : l'assimilation, le jugement et la satisfaction.

La qualité d'assimilation est une mesure de la clarté du contenu d'un point de vue informatif. Le jugement de qualité reflète la qualité de présentation. La satisfaction indique le degré d'appréciation globale de l'utilisateur.

Pour évaluer la qualité audiovisuelle il existe essentiellement deux catégories d'évaluation à savoir les méthodes subjectives qui impliquent des observateurs humains pour évaluer la qualité des contenus multimédias et des méthodes objectives qui calculent la qualité automatiquement à l'aide de modèles mathématiques .

## 1.5 Evaluation de la qualité

### 1.5.1 Evaluation de la qualité subjective

Afin de mesurer de manière fiable la qualité perceptuelle par les systèmes auditifs et/ou visuels humains, les tests subjectifs sont effectués lorsque des groupes d'observateurs humains formés ou naïfs fournissent des cotes de qualité [11]. Cette procédure d'évaluation est connue comme évaluation de la qualité subjective qui vise à quantifier la gamme d'opinions que les utilisateurs expriment quand ils voient / entendre le contenu numérique et elle est généralement effectuée dans un environnement bien contrôlé à l'aide de recommandations normalisées.

Les modèles subjectifs d'évaluation de la qualité des multimédias les plus performants et les plus utilisés sont normalisés par l'Union Internationale des Télécommunications (UIT). Cet organisme est chargé de la normalisation et de la planification des télécommunications dans le monde. Elle établit les normes de ce secteur et diffuse toutes les informations techniques nécessaires pour permettre l'exploitation des services mondiaux de télécommunications. D'autres modèles sont aussi proposés par des laboratoires universitaires ou encore par des sociétés privées, mais ils ne sont pas validés par les normes de l'UIT.

### 1.5.2 Evaluation de la qualité objective

Bien que l'évaluation de la qualité subjective fournit des indices fiables de la qualité de la perception humaine, il ne peut pas être appliquée dans l'évaluation de la qualité en temps réel en service. Ainsi, les méthodes d'évaluation de la qualité objective ont été mises au point pour remplacer le panneau humain par un modèle de calcul pour prédire les résultats d'un test subjectif. À savoir, l'objectif de l'évaluation objective de la qualité est d'estimer automatiquement les valeurs MOS, qui sont aussi proches que possible des scores de qualité obtenus à partir de l'évaluation de la qualité subjective. [12]

## Conclusion

Dans ce chapitre nous avons abordé certaines définitions associées à l'audiovisuel notamment la vidéo et l'audio tout en passant par sa qualité qui contient la qualité subjective et objective et qui seront des points essentiels dans la suite de notre travail.

Dans le chapitre suivant nous verrons les différentes métriques d'évaluation de la qualité audio (QA), la qualité vidéo (QV) ainsi que la qualité audiovisuelle (QAV).

## Chapitre 2

# Evaluation de la qualité multimédia (video , audio , audiovisuel)

## Introduction

Actuellement, les méthodes utilisées pour mesurer la qualité perçue par l'utilisateur sont généralement celles recommandés par l'Union Internationale des Télécommunications (UIT) dont l'objectif est de proposer un cadre commun d'évaluation aux différents laboratoires, instituts ou entreprises cherchant à évaluer la qualité perçue de service de restitution audio et/ou vidéo (téléphonie, TV, etc.). Ces méthodes reposent principalement sur la collecte de notes subjectives recueillies sur des échelles de qualité, après visualisation et/ou écoute de séquences traitées par le service ou la technologie à évaluer.

Dans ce chapitre nous avons présenté des notion de base sur l'évaluation de la qualité vidéo , la qualité audio et la qualité audiovisuel . Cette qualité contient les métriques subjectives et les métriques objective qui discuter au long du chapitre .

## 2.1 Les types d'évaluation de la qualité

Il existe deux types d'évaluation , subjective et objective et elles se changent de familles à une autre. La vidéo a des méthodes propre à lui même ainsi que l'audio et audiovisuel.

## 2.2 Evaluation subjective

Dans la plupart des applications visuelles, l'oeil humain est le terminal final qui va recevoir et traiter l'information. Pour cela l'évaluation subjective est considérée comme étant la meilleure manière de mesure de la qualité de vidéo puisqu'elle est basée sur les appréciations des êtres humains.

## 2.2.1 Evaluation subjective de la vidéo

- Cette manière requiert l'utilisation d'observateurs humains qui doivent évaluer la qualité des vidéos qui leur sont présentées durant des expériences appelées "tests d'évaluation subjective de qualité vidéo".

- Pour augmenter la précision des notes de qualités, certains facteurs doivent être respectés pendant la réalisation des tests. Cependant, dans les mêmes conditions, les observateurs peuvent juger très différemment la même vidéo car leurs décisions dépendent de la qualité attendue par chaque observateur.

### 2.2.1.1 Facteurs influents

Certains facteurs sont pris en compte pour s'assurer de la fiabilité de l'évaluation subjective. Ces derniers peuvent influencer le jugement des observateurs, les plus pertinents sont :

**Distance d'observation :** La distance d'observation affecte la visibilité d'un stimulus. Cette distance est fixée entre 4 à 6 fois la hauteur de l'image de projection et qu'il faut conserver tout au long du test.

**Le calibrage de l'écran :** Il est nécessaire de calibrer le support d'affichage qu'est l'écran afin qu'il fonctionne dans des conditions optimales. Sans cette calibration, les couleurs affichées peuvent être différentes du stimulus d'origine. [13]

**Conditions de visualisation :** L'environnement et l'éclairage de la salle affecte effectivement la perception des stimuli (séquence de test). Un éclairage élevé risque de troubler l'observateur et ainsi pourrir sa perception. En revanche, la couleur de fond de la salle affecte l'apparence de certaines couleurs.

**Séance d'évaluation :** Une séance ne doit en aucun cas dépasser une demi-heure, car l'observateur commence à présenter des signes de fatigue et/ou d'adaptation, et son jugement ne sera plus fiable. Au début de la première séance, environ un certain nombre de "fausses présentations" doivent être introduites afin de stabiliser l'opinion de l'observateur. Les données résultant de ces présentations ne sont pas considérées pour le résultat final du test. [14]

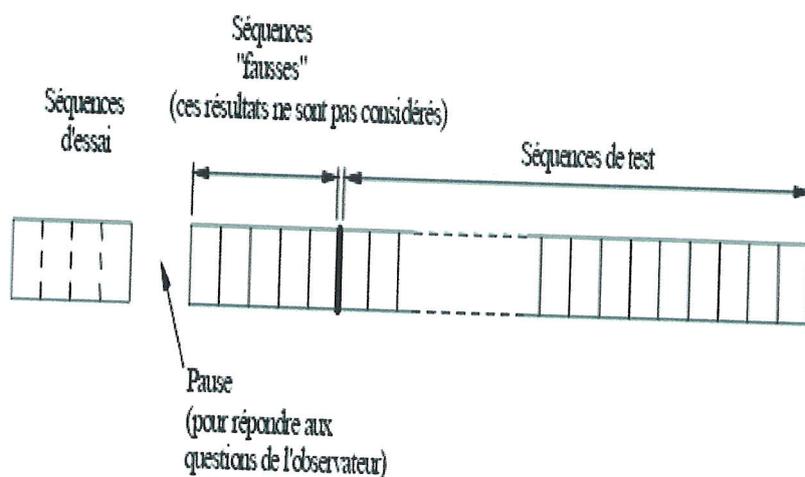


FIGURE 2.1 – Présentation des séquences d'évaluation

**Facteurs psychologiques :** Un test d'évaluation psychophysique permet de calculer la sensibilité de l'observateur par rapport à un environnement normalisé et à une tâche précise. Il existe différents types de tests psychophysiques proposés dans les normes internationales. Ceux-ci peuvent être répartis génériquement en deux classes :

**Tests comparatifs :** Deux ou plusieurs images (vidéo) sont présentées à l'observateur lequel doit les comparer les unes aux autres.



FIGURE 2.2 – test d'ordonnement

Le test d'ordonnement permet d'effectuer un classement des images (vidéo) fournies à l'observateur de la meilleure à la plus mauvaise. Ainsi, chaque image (vidéo) sera notée en fonction de son choix et de la configuration retenue (1 à 8 pour la configuration de la figure 2.2).



FIGURE 2.3 – test de choix forcé

Le test de choix forcé (Figure 2.3) permet de comparer plusieurs images (vidéo) deux à deux (avec ou sans référence). Une variante de ce test a été baptisée test Best/Worst. Ce dernier test (Figure 2.4) permet de faire sortir à partir d'un groupe d'images (vidéo), la meilleure et la plus mauvaise d'un point de vue qualitatif.

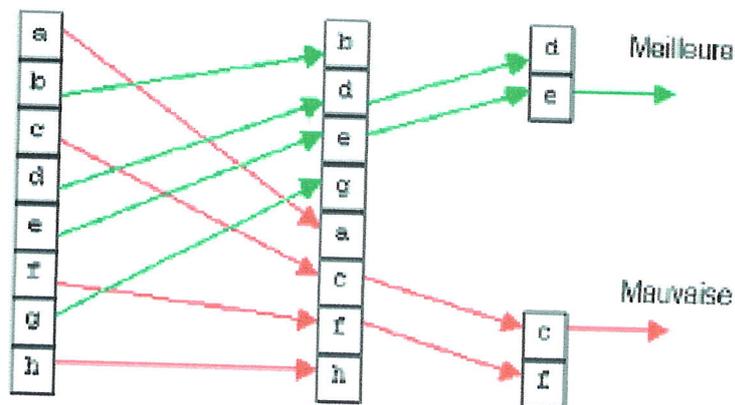


FIGURE 2.4 – test Best/Worst

**Tests de mesure absolue :** l'observateur doit attribuer une note de qualité à une image (vidéo) présentée seule (sans référence).

Le but de ces tests n'est pas de déterminer un seuil de sensibilité, mais de noter la qualité d'une vidéo. Il est demandé à l'observateur de quantifier numériquement la qualité de l'image affichée à l'écran. Ce procédé est très largement utilisé pour valider une métrique de qualité.

**Les observateurs :** Il est recommandé d'avoir un panel d'observateurs le plus large possible, au moins quinze individus. Ils peuvent être experts ou novices, en ce qui concerne le thème de la campagne d'évaluation.

Avant chaque séance, les observateurs seront sélectionnés pour leur acuité visuelle normale

ou rendue normale par correction et leur vision normale des couleurs. L'observateur doit avoir ainsi une acuité visuelle de 10/10 pour les deux yeux avec ou sans correction. Pour la vision normale des couleurs, le test consiste à détecter, grâce à des planches comme celles données par la figure 2.5, un défaut au niveau de la vision des couleurs.

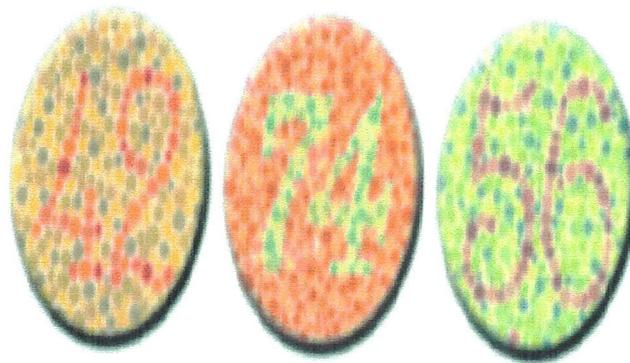


FIGURE 2.5 – Test d'Ishihara

Le panel d'observateurs choisi dépend des types de traitement à évaluer et de la nature de l'étude à mener. Il faut donc prendre en compte, lors de la sélection de ce panel, plusieurs critères comme l'âge, l'origine socioculturelle, professionnelle, le sexe etc.

### 2.2.1.2 Protocoles

Il y a essentiellement trois grandes familles communes d'évaluation subjective définies par l'UIT [15] : Echelle continue de la qualité sur stimulus double (DSCQS), Echelle de dégradation sur stimulus double (DSIS) et Evaluation continue de la qualité sur stimulus unique (SSCQE).

**DSCQS : Double Stimulus Continuous Quality Scale** : Le but principal de la méthode DSCQS est de mesurer la qualité des systèmes par rapport à une référence. Les personnes qui sont montrées paires de séquences vidéo (la séquence de référence et la séquence altérée) dans un ordre aléatoire. Il est largement accepté comme une méthode de test précis avec peu de sensibilité aux effets de contexte, en tant que spectateurs sont présentés deux fois la séquence. Les téléspectateurs sont invités à évaluer la qualité de chaque séquence de la paire après la deuxième projection. Il est également utilisé pour mesurer la qualité du codage d'image stéréoscopique. [16]

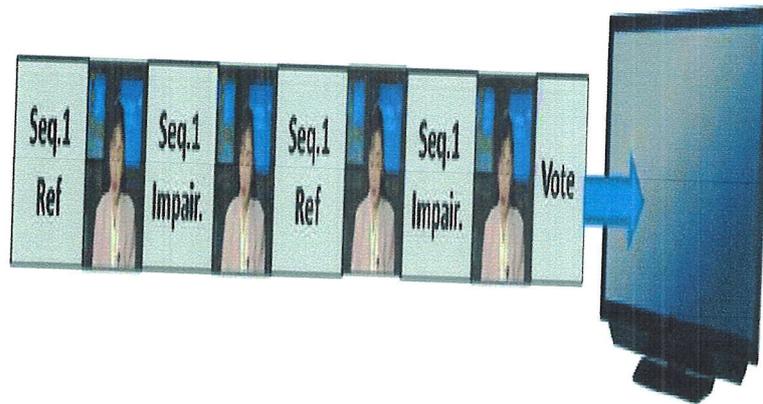


FIGURE 2.6 – Schéma d'un système de DSCQS

Le tableau 2.1 donne un exemple d'une échelle de mesure à 5 niveaux de notation utilisée pour évaluer la qualité d'une vidéo .

Echelle à cinq notes		
Qualité	Dégradation	
Excellente	5	Imperceptible
Bonne	4	Perceptible mais non gênante
Assez bonne	3	Légèrement gênante
Médiocre	2	Gênante
Mauvaise	1	Très gênante

TABLE 2.1 – Exemple d'échelle de notation

**DSIS : Double Stimulus Impairment Scale** : Comme dans la méthode de DSCQS, chaque essai se compose d'une paire de stimulus : la référence et l'essai. Cependant, dans la méthode de Double Stimulus (DSIS) [15], les deux stimulus sont toujours présentés dans le même ordre : la référence est toujours la première, suivi du test. Dans la méthode de DSIS, les observateurs comparent les deux stimuli dans un essai et évaluent la dégradation du stimulus d'essai en ce qui concerne la référence, en utilisant une échelle de dégradation de cinq niveaux. Ainsi, seulement une voix est faite pour chaque essai de DSIS.

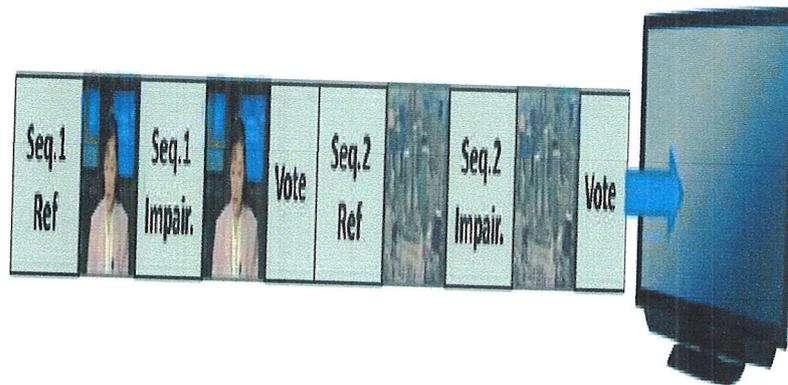


FIGURE 2.7 – Schéma d'un système de DSIS

**SSCQE : SINGLE-STIMULUS CONTINUOUS QUALITY EVALUATION :**  
Au lieu de voir des paires de courtes séquences séparées, les observateurs observent un programme de la durée 20-30 minutes en général qui a été traité par le système du test ; la référence n'est pas montrée (Single Stimulus Continuous Quality Evaluation, SSCQE). En utilisant un glisseur, les observateurs évaluent continuellement la qualité instantanément perçue sur l'échelle de DSCQS du mauvais à l'excellent (voir figure 2.8). [17]



FIGURE 2.8 – Schéma d'un système de SSCQE

Une échelle de notation utilisée pour évaluer la qualité d'une vidéo est donnée dans le tableau 2.2. Ces échelles peuvent varier de 1 à 5 jusqu'à 100.

Excellente
Bonne
Assez Bonne
Médiocre
Mauvaise

TABLE 2.2 – Echelle d'évaluation de qualité video

### 2.2.1.3 Analyse des résultats

La phase de préparation des données est déterminante pour la qualité de l'analyse. Des erreurs ou des incohérences peuvent être commises lors du processus d'évaluation des systèmes. Les données subjectives brutes peuvent être traitées pour obtenir une note moyenne d'opinion, l'écart type et l'intervalle de confiance, qui pourront aider à classer de manière plus précise les cas testés.

**MOS : Mean Opinion Score :** La première étape de l'analyse des résultats consiste à calculer la note moyenne ou le MOS (Mean Opinion Score) pour chacune des présentations.

$$MOS(i) = \frac{1}{N} \sum_{j=1}^N Note_i(j) \quad (2.1)$$

Où N est le nombre d'observateurs et  $Note_i(j)$  est la note de l'observateur j pour la dégradation i de l'image.

**Intervalle de confiance :** un intervalle de confiance est souvent associé à chaque note MOS. Il est généralement fixé à 95 % et est donné par la relation suivante :

$$Mos(i) - e_j, Mos(i) + e_j \quad (2.2)$$

Où :

$$e_j = 1.95 \frac{\sigma_j}{\sqrt{N}}$$

Et l'écart-type pour chaque présentation  $\sigma_j$  est donné par :

$$\sigma_j = \sqrt{\frac{\sum_{K=1}^N (Note_j(K) - MOS(K))^2}{N - 1}} \quad (2.3)$$

## 2.2.2 Evaluation subjective de l'audio

Les tests d'écoute subjectifs sont toujours considérés comme le moyen le plus fiable de mesure de la qualité des systèmes audio. Ils sont bien décrits et constituent des méthodes éprouvées d'évaluation de la qualité audio lorsque celle-ci est située au haut ou au bas de l'échelle.

### 2.2.2.1 La méthode MUSHRA : Multiple Stimuli with Hidden Reference and Anchor

La méthode MUSHRA est une méthodologie permettant de réaliser un test d'écoute de codec afin d'évaluer la qualité perçue de la sortie d'algorithmes de compression audio avec perte .Il est défini par la recommandation UIT BS.1534-3.[18] La méthodologie MUSHRA est recommandée pour évaluer la "qualité audio intermédiaire".

La méthode de test MUSHRA ne soit pas destinée à l'évaluation des faibles dégradations, il est cependant recommandé de faire appel à des auditeurs expérimentés afin de garantir la validité des données recueillies. Ces auditeurs doivent avoir l'habitude d'écouter les sons de façon critique. Des résultats plus fiables seront ainsi obtenus plus rapidement qu'avec des auditeurs non expérimentés. Il est également important de noter que la plupart des auditeurs non expérimentés acquièrent en général une plus grande sensibilité aux différents types d'artéfacts après les avoir fréquemment rencontrés. Un estimateur expérimenté est choisi pour son aptitude à procéder à un test d'écoute. Cette aptitude doit être qualifiée et quantifiée en termes de ses compétences, pour ce qui est de la fiabilité et du discernement dont il fait preuve lors de la répétition des évaluations. Ceux-ci sont définis comme suit [18]

- Le discernement : L'aptitude à percevoir des différences entre les éléments testés.
- La fiabilité : Le fait pour les évaluations répétées du même élément testé d'être proches les unes des autres.

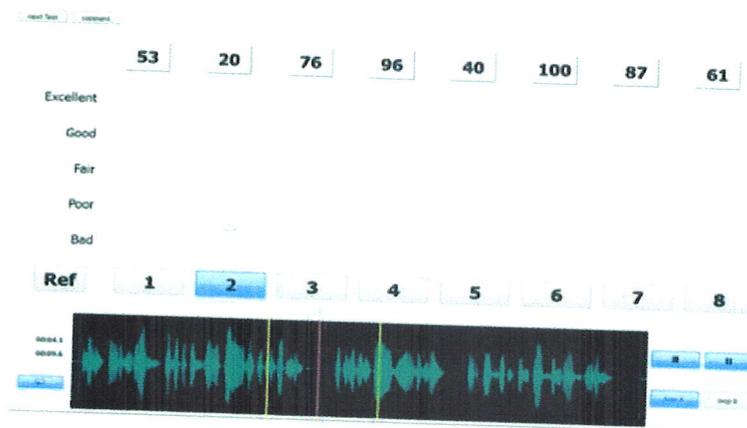


FIGURE 2.9 – Exemple d'un affichage sur ordinateur utilisé pour un test MUSHRA

### 2.2.3 Evaluation subjective de l'audiovisuel

La norme UIT-T P.911 [19] propose des méthodes d'évaluation subjective de la qualité audiovisuelle (AV) pour applications multimédias non interactives (contexte passif d'écoute et de visualisation : TV, multimédias, etc.). Le jugement de qualité s'effectue sur une seule et unique échelle à l'issue de la visualisation et de l'écoute de chaque séquence audiovisuelle de test. Les Trois méthodes proposées dans le cadre de cette norme sont :

#### 2.2.3.1 Méthode ACR : ABSOLUTE CATEGORY RATING

La méthode ACR ou méthode d'évaluation par catégories absolues consiste à attribuer une note de qualité après chaque séquence AV visualisée/entendue. La note de jugement attribuée doit refléter l'opinion du participant quant à la qualité audiovisuelle globale perçue, c'est-à-dire la qualité audio et vidéo combinée. Cette évaluation est réalisée sur une échelle catégorielle de cinq ou neuf points (intervalles) explicitée par cinq items (Excellent-Bon-Satisfaisant-Médiocre-Mauvais). Il est recommandé d'utiliser l'échelle en neuf points lorsqu'une plus grande puissance de discrimination est nécessaire, typiquement, lorsque l'on souhaite évaluer des codages à bas débit. [20]

9	Excellent
8	
7	Bon
6	
5	Satisfaisant
4	
3	Médiocre
2	
1	Mauvais

5	Excellent
4	Bon
3	Satisfaisant
2	Médiocre
1	Mauvais

FIGURE 2.10 – Echelle d'évaluation de qualité à 9 et 5 niveaux

La norme recommande des séquences d'une durée comprise entre huit et dix secondes, l'intervalle de temps conseillé pour le vote est égal ou inférieur à dix secondes. Le chronogramme recommandé par la norme UIT-T P.911. [19]

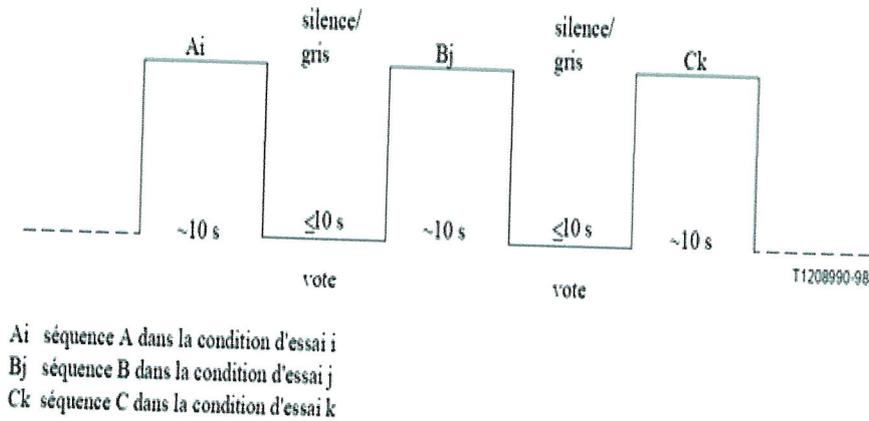


FIGURE 2.11 – Chronogramme de la méthode ACR

### 2.2.3.2 MÉTHODE DCR : DEGRADATION CATEGORY RATING

La méthode DCR ou méthode par évaluation de catégories de dégradations propose une présentation des séquences AV de test par paires. Les séquences constituant la paire sont identiques à la différence que la première est toujours présentée sans dégradations (référence) tandis que la seconde est traitée par le système à évaluer (donc susceptible de comporter des dégradations). La séquence traitée est toujours présentée après la référence. [19]

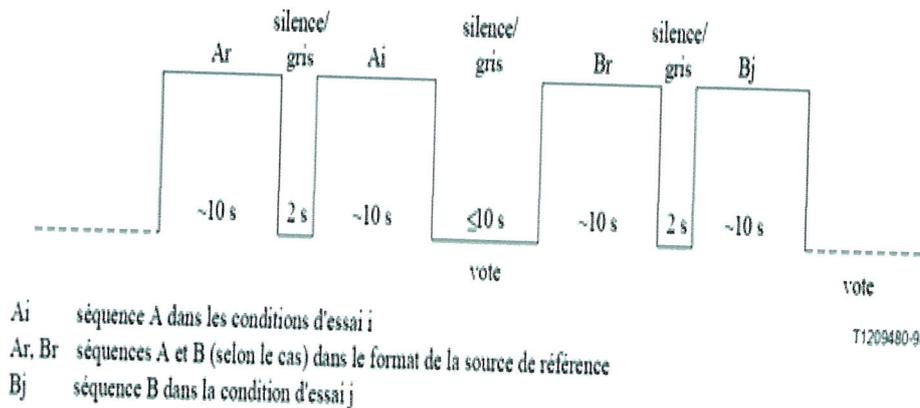


FIGURE 2.12 – Chronogramme de la méthode DCR

Seule la séquence traitée est évaluée par les participants en comparaison avec la condition de référence. L'échelle d'évaluation correspond ici à une échelle de perceptibilité de la dégradation comme présenté par la Figure 2.13 .



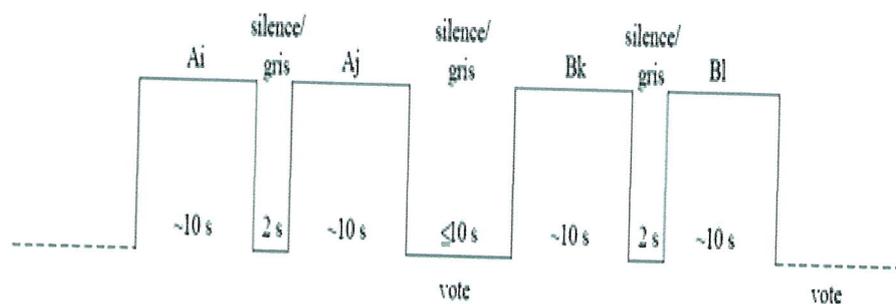
5	Imperceptible
4	Perceptible mais non gênant
3	Légèrement gênant
2	Gênante
1	Très gênante

FIGURE 2.13 – Echelle de dégradation à cinq niveaux

### 2.2.3.3 METHODE PC : PAIR COMPARISON

La méthode des comparaisons par paires implique que les séquences d'essai soient présentées en paires. Chaque paire est formée de la même séquence, présentée d'abord au moyen d'un système l'essai puis au moyen d'un autre système. La séquence de référence (sans dégradation) peut être incluse et sera traitée comme un système à l'essai additionnel.

Toutes les combinaisons de paires de séquences A, B, C, etc... devront être évaluées associées selon toutes les  $n(n-1)$  combinaisons possibles (AB, BA, CA, etc.) et présentées dans les deux ordres possibles (AB, BA, etc.). Le jugement de qualité AV globale est ici exprimé à travers un jugement de préférence pour l'une ou l'autre séquence de la paire qui doit réaliser après la présentation de chaque paire. Cette méthode est notamment préconisée pour la comparaison de systèmes quasi-équivalents et/ou de haute qualité. La durée recommandée pour les séquences de test est d'environ dix secondes, celle du temps de vote doit être inférieure ou égale à dix secondes.



Ai, Aj séquence A dans les  $i^e$  et  $j^e$  conditions d'essai, respectivement  
 Bk, Bl séquence B (selon le cas) dans les  $k^e$  et  $l^e$  conditions d'essai, respectivement

T1209490-98

FIGURE 2.14 – Chronogramme de la méthode PC

### 2.2.3.4 Comparaison entre les différentes méthodes de l'évaluation de la qualité audiovisuelle

Lors du choix d'une méthode d'essai, un critère important est la différence fondamentale entre méthodes faisant appel à des références explicites (par exemple DCR) et méthodes ne faisant pas appel à des références explicites (par exemple ACR, PC ). Cette deuxième classe de méthodes ne contrôle ni la transparence ni la fidélité.

Il convient d'utiliser la méthode DCR lorsque l'on contrôle la fidélité de transmission par rapport au signal de source. Ce facteur présente souvent de l'importance pour l'évaluation de systèmes de haute qualité. D'autres méthodes peuvent être utilisées pour évaluer les systèmes de haute qualité. Les observations spécifiques de l'échelle DCR (dégradation "imperceptible/perceptible") sont précieuses lorsque la détection d'une dégradation par l'observateur est un facteur important.

Lorsqu'il importe de vérifier la fidélité par rapport au signal de source, il convient donc d'utiliser la méthode DCR. La méthode DCR sera également appliquée pour l'évaluation de systèmes de haute qualité, dans le contexte des communications multimédias, cela grâce à la discrimination entre dégradation imperceptible/perceptible sur l'échelle DCR ainsi que grâce à la comparaison avec la qualité de référence.

La méthode ACR est facile et d'application rapide. Sa présentation des stimuli est semblable à celle de l'usage courant des systèmes. La méthode ACR convient donc bien pour des essais de qualification. Le principal mérite de la méthode PC est son haut pouvoir discriminatoire, qui est particulièrement précieux lorsque plusieurs objets d'essai sont de qualité presque égale.

Lorsqu'il faut évaluer un grand nombre d'objets au cours du même essai, la procédure fondée sur la méthode PC tend à être longue. Dans ce cas, un essai ACR ou DCR peut d'abord être effectué avec un nombre limité d'observateurs, suivi d'un essai PC effectué seulement sur les objets qui ont reçu à peu près la même note d'évaluation. [19]

## 2.3 Evaluation objective

### 2.3.1 Evaluation objective de la vidéo

#### 2.3.1.1 Métriques à référence complète (FR)

Avec cette approche, l'ensemble de la vidéo originale est disponible à titre de référence. En conséquence, les méthodes FR sont basées sur la comparaison d'une vidéo déformée avec

la vidéo d'origine.[21]

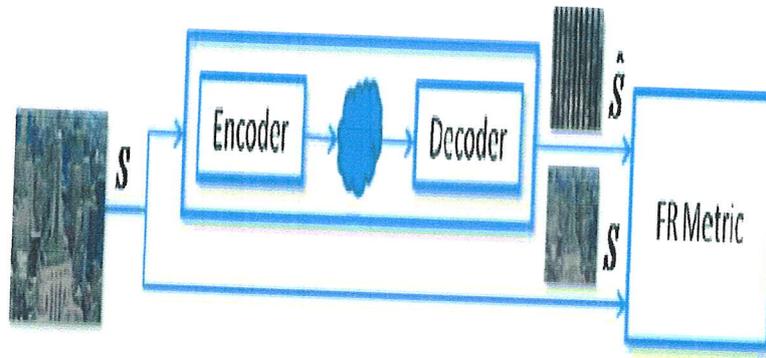


FIGURE 2.15 – Schéma de métrique à référence complète

**Approches basées EQM : Erreur Quadratique Moyenne :** L'erreur quadratique moyenne est l'une des métriques les plus populaires et les plus utilisées à ce jour. Cela est dû essentiellement à sa simplicité, au temps de calcul et surtout au fait que l'on puisse l'intégrer facilement dans un processus d'optimisation tel que la minimisation d'une fonction de coût ou de distorsion mathématique. Cette mesure n'est rien d'autre que la moyenne quadratique du signal erreur ou distorsion. Elle est donnée par :

$$EQM = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - I_d(i, j))^2 \quad (2.4)$$

Avec  $I(i, j)$  et  $I_d(i, j)$  le pixel de coordonnées  $(i, j)$  de l'image originale et de sa version dégradée, respectivement.  $N$  et  $M$  représentent respectivement le nombre de lignes et de colonnes de frame. A partir de cette définition, différentes variantes ont été proposées tel que le rapport signal à bruit crête (PSNR : Peak Signal Noise Ratio) restent les plus utilisées.

$$PSNR = 10 \log_{10} \frac{\text{Valeur crête du signal original}}{EQM} \quad (2.5)$$

Où la valeur crête du signal originale est la valeur maximal 255 par pixels pour les images .

**Approches Structurelles** Pour pallier la faiblesse des mesures type EQM, basées sur le pixel en général, certains auteurs ont orienté leur recherche vers les métriques de qualité basées sur une analyse des structures locales du signal image.

Parmi les métriques les plus populaires et les plus utilisées, on trouve celles proposées par Wang et al, appelées SSIM.

**SSIM :Structural SIMilarity** L'indice de similarité structurelle est une mesure basée sur l'hypothèse que le système visuel humain est adapté pour extraire des informations structurelles dans le champs de vision. Par conséquent, le changement de informations structurelles entre l'image déformée et l'image originale pourrait être une bonne approximation de la distorsion de l'image perçue.

La version de base de SSIM est décrite, où l'information structurelle est recueillie par une comparaison de la luminance, le contraste et la structure. Laisser  $x$  et  $y$  Deux signaux d'image non négatifs de longueur  $N$ . Tout d'abord, la luminance est comparée par une fonction d'intensités moyennes  $\mu_x$  et  $\mu_y$  sous la forme :

$$\mu_x = \frac{1}{N} \sum_1^N x_i \quad (2.6)$$

Où  $N$  : le nombre de pixels de chaque fenêtre.  
 $x_i$  : l'intensité d'un pixel.

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2.7)$$

Où  $C_1$  est la constante inclus pour éviter une instabilité lorsque  $\mu_x^2 + \mu_y^2$  est presque nul .  
 $C_1 = (K_1L)^2$  , où  $L$  est une gamme dynamique c'est 255 pour les images et  $K_1 \ll 1$  est un petit nombre. Le même représente les constantes dans les fonctions de contraste et de la structure, dont la description suit.

La comparaison du contraste est une fonction d'écart types  $\sigma_x$  et  $\sigma_y$  qui ressemble à :

$$\sigma_x = \left( \frac{1}{N-1} \sum_1^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (2.8)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2.9)$$

Où  $C_2 = (K_2L)^2$  et  $K_2 \ll 1$ . Notez également que, avec la même quantité de changement de contraste  $\Delta\sigma = \sigma_y - \sigma_x$ , cette fonction est moins sensible dans le cas de fort contraste de base que dans le cas de faible contraste de base, qui correspond au comportement de masquage de contraste de système visuel humain. Enfin, la comparaison de la structure est fonction de la corrélation entre les deux signaux sous la forme :

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (2.10)$$

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (2.11)$$

Où  $C_3$  est à nouveau une petite constante. Notez que coefficient de corrélation entre les signaux d'original est le même que celui entre les signaux normalisés et pourrait donc bien représenter la structure.

Ces fonctions sont combinées dans la mesure résultante comme :

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (2.12)$$

Où  $\alpha > 0, \beta > 0, \gamma > 0$  sont des paramètres utilisés pour modifier l'importance relative des composants individuels. Wang et al ont mis  $\alpha = \beta = \gamma = 1$  et  $C_3 = C_2/2$ , qui simplifié la mesure de :

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.13)$$

Le SSIM fonctionne mieux s'il est utilisée localement. Cela signifie pour calculer les statistiques locales  $\mu_x, \sigma_x$  et  $\sigma_{xy}$  dans une petite fenêtre qui est pixel par pixel déplacé sur la totalité de l'image et les résultats sont alors en moyenne. Les raisons d'une telle approche sont que différentes parties de l'image peuvent différer beaucoup et aussi humaine peuvent se concentrer sur une seule zone limitée à l'époque. Une telle approche peut également être utilisé pour créer la carte de qualité variable dans l'espace de l'image pour obtenir plus d'informations sur la distorsion de l'image.

SSIM même si elle joue beaucoup mieux que EQM, a des limites. Par exemple, la variante de base ne fonctionne pas bien dans les cas de transformation des images redimensionnées ou mis en rotation, même si la qualité de ces images est la même que celle de leurs images de référence. Ceci est partiellement résolu par le complexe Wavelet SSIM (CW-SSIM).

SSIM en substance, compare également les signaux à l'approche pixel à pixel il est encore tout à fait semblable à EQM.

**Approches inspirées du Système Visuel Humain (SVH) :** Cet approche est considérée beaucoup plus hybride. Son objectif principal est de donner des résultats proches de ceux d'une évaluation subjective tout en gardant la simplicité de calcul des méthodes objectives. a consiste à simuler le système visuel humain et le représenter par plusieurs canaux perceptuels ou visuels. Ces canaux correspondent à des décompositions spatiofréquentielles des informations de l'image. Cette décomposition permet la prise en compte des caractéristiques du cortex visuel qui possède plusieurs variations de sensibilité. Dans un modèle de simulation, un canal visuel est représenté par un sous modèle qui correspond à une variation de sensibilité. Le modèle le plus connu est le prédicteur de différence visuelle VDP (Visual Differences Predictor) [22] . Il modélise les trois principales variations de sensibilité du SVH : l'intensité lumineuse, le contenu fréquentiel et le contenu structurel de l'image. Ce modèle permet de mesurer la qualité d'une image dégradée à l'aide d'une carte indiquant la probabilité de détection de différence visuelle entre cette image et sa version originale. Le calcul de ce paramètre consiste à transformer les images, modifiées et originales, par une fonction d'adaptation à la luminance, appliquer une fonction de sensibilité de contraste puis calculer l'EQM entre les deux images. La deuxième étape permet de définir plusieurs filtres caractérisant les sous bande fréquentielles (canaux perceptuels). [23]

### 2.3.1.2 Métriques sans référence (NR) :

Cette classe de méthodes de qualité objective ne nécessite pas d'accéder à la vidéo d'origine mais recherche des artefacts par rapport au domaine de pixel d'une vidéo, utilise les informations incorporées dans le train de bits du format vidéo associé ou effectue une évaluation de la qualité sous forme hybride. D'approches basées sur les pixels et sur le flux binaire. [21]

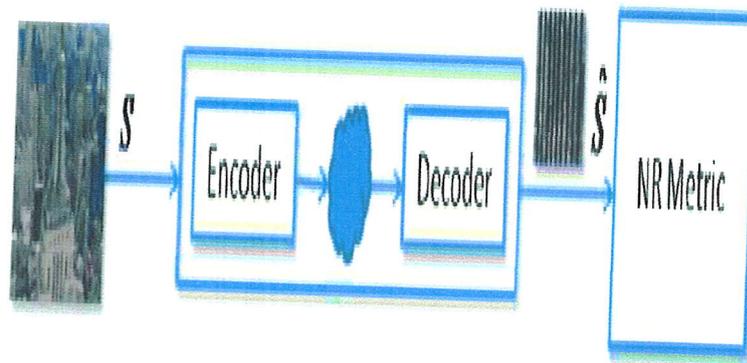


FIGURE 2.16 – Schéma de métrique sans référence

**Approches basées sur les statistiques de scènes naturelles (NSS : natural scene statistics) :** L'idée principale de ces approches est que les images de scènes naturelles

présentent des régularités statistiques affectés par la présence de distorsions, et ces changements statistiques dans la distorsion les images peuvent être mesurées pour évaluer la qualité de l'image. Parmi ces méthodes :

**BIQI (Blind Image Quality Index)** : Cette approche a été présentée par Moorthy et Bovik [24], il estime la qualité à partir d'une transformée en ondelettes utilisant la base d'ondelettes de Daubechies [25]. La transformation est effectuée sur trois échelles et trois orientations. Le coefficient de sous-bande de la transformation est paramétré à l'aide d'une distribution généralisée gaussienne GGD. Il y a trois paramètres de la distribution : moyenne, variance et paramètre de forme. Mais deux paramètres seulement sont utilisés qui sont la variance et le paramètre de forme parce que la moyenne dans la base des ondelettes est égale à zéro. Par conséquent, trois échelles, trois orientations et deux paramètres génèrent 18 caractéristiques pour l'image ( $3 \times 3 \times 2$ ). Après avoir extrait les caractéristiques, l'image testée est classée par le classificateur SVM multiclass selon l'une des cinq classes représentant cinq distorsion différentes (jpeg, jpeg2000, bruit blanc, erreur de lecture rapide et flou). Ce classificateur ne donne pas seulement une classification absolue, mais aussi en donnant le montant de chaque distorsion présente dans l'image.

**DIVIINE (Distortion Identification-based Image Verity and INtegrity Evaluation)** : Dans [26], Moorthy et Bovik ont présenté l'algorithme DIIVINE. Les caractéristiques ici sont extraites en décomposant l'image déformée d'abord en utilisant une décomposition en ondelettes utilisant une pyramide orientable sur deux échelles et six orientations (0,30,60,90,90,120,150). Ils ont constaté qu'un degré d'orientation accru améliore la performance. Les coefficients de sous-bande à travers ces orientations et échelles sont ensuite utilisés pour extraire un ensemble de caractéristiques statistiques. Au total, 88 fonctionnalités sont utilisées dans cette méthode. Ensuite, la qualité est estimée en utilisant le même cadre en deux étapes dans [27].

**Paramètres d'analyse comparative** : Les différents algorithmes d'évaluation de qualité sans référence fournissent différents scores de qualité. Donc, pour comparer la performance de ces algorithmes il faut avoir un système de comparaison, les paramètres utilisés pour faire la comparaison sont cités ci-dessous :

**Pearson coefficient de corrélation (PCC)** : Il est utilisé pour mesurer la dépendance entre les variables. Sa valeur est entre (-1,+1) où la valeur proche de '+1' indique que les deux variables ont une corrélation positive et la valeur proche de '-1' indique que les deux variables ont une corrélation négative. Une très faible ou un zéro indique les deux variables ne sont pas corrélées. L'équation de PCC entre X, Y est décrite comme suit :

$$\rho = \frac{\text{Covariance}(X, Y)}{\sigma_x \sigma_y} \quad (2.14)$$

Les deux variables X, Y sont le résultat de d'algorithmes et la qualité réelle obtenue par la base de données.

**Spearman Corrélation Coefficient (SCC) :** Fournit la relation entre deux variables classées. Sa valeur est entre (-1, +1) avec la même interprétation que celle de PCC. Le SCC est calculé comme suit :

$$\rho = 1 - \frac{6 * \sum d^2}{n(n^2 - 1)} \quad (2.15)$$

Où (d) est la différence dans les rangs de deux variables X et Y,  $d_i = x_i - y_i$ . Et n est le nombre de paires de valeurs.

### 2.3.1.3 Métriques à référence réduite (RR) :

Dans ce cas, il n'est pas nécessaire de donner accès à la vidéo d'origine mais uniquement de fournir des fonctionnalités représentatives des caractéristiques de la vidéo d'origine. La comparaison des informations réduites de la vidéo d'origine avec les informations correspondantes de la vidéo déformée fournit l'entrée pour les méthodes RR .[21]

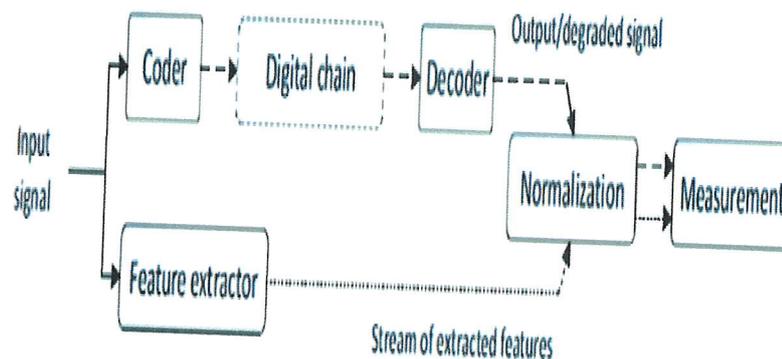


FIGURE 2.17 – Schéma de métrique à référence réduite

**Les approches avec référence réduite :** Parmi les méthodes d'évaluation de qualité d'image avec référence réduite citons :

Dans l'article[28], le score de qualité est estimé en mesurant les changements d'entropies pondérées des images de référence et déformées dans le domaine des ondelettes. Les

coefficients d'ondelettes de l'image naturelle et déformés sont modélisés sous forme de distributions GSM. La métrique de qualité est obtenue en tant que distance entre l'approximation de référence et l'image naturelle d'une image déformée. L'algorithme proposé fonctionne beaucoup mieux que l'algorithme FR MSE.[29]

Dans l'article[30], un nouvel algorithme RR IQA basé sur des statistiques d'image est proposé. Les statistiques sur les images sont modélisées dans le domaine des pixels, basé sur la distribution en dégradé. Les caractéristiques de qualité RR sont extraites sur la base d'un modèle statistique. La qualité est mesurée en déterminant la similarité entre l'image de référence et l'image déformée. La méthode proposée est aussi performante que le PSNR FR largement connu[29].

Dans cet article[31], une tentative d'approximation de la métrique RR SSIM à partir de FR SSIM a été effectuée par transformation de normalisation divisible DNT de l'image naturelle. Outre l'évaluation de la qualité, le concept de réparation d'image a été introduit en comparant les propriétés statistiques de sous-bande de l'image déformée à celles de l'image de référence.

Les caractéristiques sont extraites de l'image de référence en appliquant une transformation linéaire à plusieurs échelles (transformée en ondelettes), puis la représentation de la DNT est calculée en divisant chaque coefficient d'ondelette par une mesure d'énergie locale basée sur les coefficients voisins. Pour un résumé efficace des propriétés statistiques de l'image de référence, une moyenne nulle est ajustée sur les coefficients DNT. Pour l'estimation de RR SSIM, l'effet des distorsions sur les propriétés statistiques doit être cohérent avec le FR SSIM correspondant. La méthode proposée est comparée à FR (PSNR et SSIM) et à RR (ondelettes marginales et DNT marginales) et la mesure proposée est très compétitive dans la plupart des cas.[29]

## 2.3.2 Evaluation objective de l'audio :

### 2.3.2.1 Méthode intrusive :

Les modèles intrusifs comparent un signal original avec un signal dégradé testé. Les techniques intrusives peuvent être divisées à des méthodes tel que : PESQ , PEAQ , SSIM et PSNR.

**L'évaluation perceptive de la qualité vocale :** PESQ (Perceptual Evaluation of Speech Quality )normalisé par l'UIT-T [32] est une méthode de prédiction de la qualité subjective pour la téléphonie avec bande passante réduite. PESQ transforme les signaux de parole originale et dégradé en une représentation psychologique qui se rapproche de la perception

humaine, calcule la distance de perception et l'inscrit en un score objectif conforme au MOS.

**L'évaluation perceptuelle de la qualité audio :** Actuellement, le seul standard en vigueur est la méthode PEAQ (Perceptual Evaluation of Audio Quality) , connu aussi sous la référence Rec.ITU-R BS.1387 [33]. Cette recommandation constitue la synthèse d'un ensemble de techniques qui existaient avant et dont le principe est illustrée sur la figure 2.18. Il s'agit d'une méthode dite avec référence où on compare un signal de référence à sa version potentiellement dégradée récupérée à la sortie du système évalué.

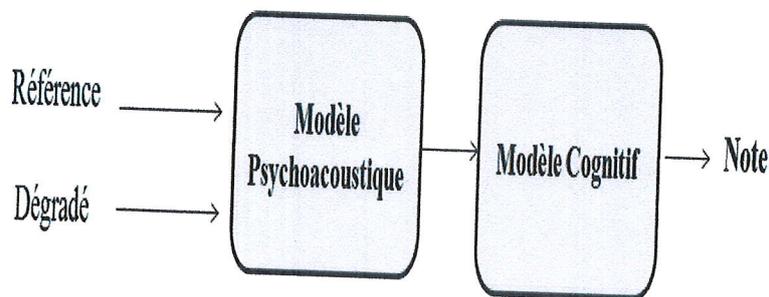


FIGURE 2.18 – Principe de la PEAQ

PEAQ est un algorithme qui modélise les principes psycho-acoustiques du système auditif humain et ces mêmes principes psycho-acoustiques sont utilisés dans de nombreux codec audio pour réduire le débit binaire tout en conservant un niveau acceptable de qualité audio. PEAQ peut être décrit comme étant constitué de deux parties : le modèle psycho-acoustique et le modèle cognitif comme le montre dans la figure 2.18.

#### **Le modèle psycho-acoustique :**

Le modèle psycho-acoustique transforme les signaux d'entrée dans le domaine temporel en une représentation de la membrane basilaire (à savoir un modèle de la membrane basilaire dans le système auditif humain) et après cette transformation, les signaux sont traités dans le domaine des fréquences avec l'utilisation d'une Transformée de Fourier rapide (FFT).

#### **Le modèle cognitif :**

Le modèle cognitif condense les informations à partir d'une séquence de trames produite par le modèle psycho-acoustique. Les sources d'informations les plus importantes pour des mesures de qualité sont les différences entre le Signal de référence et le Signal testé dans le domaine fréquentiel et celui des hauteurs sonores. Dans le domaine fréquentiel, on mesure les largeurs de bandes spectrales des deux signaux, ainsi que la structure harmonique de l'erreur.

Dans le domaine des hauteurs sonores, les mesures de l'erreur sont dérivées de la modulation de l'enveloppe de l'excitation, ainsi que de l'amplitude de l'excitation .

**la mesure de similarité structurelle moyenne :** L'indice de similarité structurelle est basée sur l'idée selon laquelle une mesure de changement dans l'information structurelle est une bonne approximation au changement de qualité perçue. Elle considère trois différences mesurées entre les signaux d'origine et les signaux dégradé : la luminosité, le contraste et la structure.

La luminosité est une comparaison des valeurs moyennes des signaux. Si X et y sont des segments correspondants de l'audio avec N chacun des échantillons, la comparaison de la luminosité est donnée par :

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2.16)$$

Où : La gamme dynamique des éléments de La gamme dynamique des éléments de X et Y est désigné par la variable L .

La comparaison du contraste ou de la variance est définie similaire à la comparaison de luminance donnée par l'équation 2.13, mais en ce qui concerne les écarts type relatifs des deux segments : à savoir :

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2.17)$$

revenant à la structure mesure en termes de l'original X et y vecteurs est donnée par :

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (2.18)$$

Enfin, les trois composants sont combinés pour donner une mesure globale de similitude comme :

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (2.19)$$

Où  $\alpha > 0$  ,  $\beta > 0$  ,  $\gamma > 0$  sont utilisés pour ajuster l'importance relative des trois composantes.

**le rapport signal à bruit crête : PSNR** PSNR comparer le signal audio d'origine avec le signal dégradé. il est mesuré en décibels (db). Il a comme équation :

$$PSNR = 10 \log_{10} \frac{\text{Valeur crête du signal original}}{RMSE} \quad (2.20)$$

Où la valeur crête du signal originale est la valeur maximal 1 pour les audio .

Et RMSE est l'erreur quadratique moyenne entre l'original et les signaux dégradés. Il a comme équation :

$$RMSE = \frac{1}{M.N} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - I_d(i, j))^2 \quad (2.21)$$

Où M.N : la taille de la liste audio .

### 2.3.2.2 Méthode non intrusive :

Bien que les méthodes intrusives soient plus précises, elles ne conviennent pas aux applications en temps réel, mais nécessitent une synchronisation difficile entre les signaux de référence et traités. Les méthodes d'évaluation objective de la qualité audio qui évaluent la qualité audio en utilisant uniquement le signal de test (ou dégradé) sont appelées méthodes non intrusives. Les techniques non intrusives peuvent être divisées en deux classes : approches à priori et sources [8].

- Une approche basée sur les PRIORIS :

Les approches à priori apprennent d'abord un ensemble de distorsions bien caractérisées, puis établissent une relation statistique entre cet ensemble et les opinions subjectives. Par exemple, la technique de qualité de la parole basée sur la sortie mesurée pour les systèmes de communication sans fil consiste à analyser les caractéristiques visuelles du spectrogramme du signal audio. La méthode calcule la variance et la plage dynamique de manière bloc par bloc, puis calcule la moyenne de tous les blocs pour obtenir le score de qualité final. Gray et al ont proposé une nouvelle utilisation de la technique de modélisation du tractus vocal pouvant être utilisée pour l'évaluation de la qualité non intrusive des flux de parole sur des réseaux. Les résultats rapportés ont montré l'efficacité de la technique, mais également la sensibilité au sexe du locuteur [8].

- Une approche basée sur la SOURCE :

Les approches basées sur la source peuvent être considérées comme des méthodes plus universelles, puisqu'elles émettent des hypothèses a priori propriétés de signal propres attendues plutôt que les distorsions cela peut se produire. De cette façon, ils peuvent gérer une large gamme de types de distorsion. L'une des premières tentatives de développement d'algorithme

d'évaluation de la qualité audio basés sur la source est le suivant : le modèle a comparé la variété de signaux audio neutres à distordus au moyen de la prédiction perceptuelle-linéaire (PLP). Cependant, la méthode est coûteuse en termes de calcul car elle est basée sur la technique de quantification vectorielle (VQ), et sa capacité généralisée est inférieure. Pour remédier à certains de ces inconvénients, Falk et Chan ont remplacé les QV par des modèles de mélange gaussien (MGM) et ont proposé une mesure de cohérence pour en estimer la qualité. Des résultats améliorés ont été obtenus plus tard une fois que des MGM propres et dégradés ont été utilisés. Une évaluation de la qualité de la parole basée sur la quantification vectorielle et une auto organisation .Falk et al[33]. ont mis au point une métrique normalisée SRMR (speech-to-reverberation-modulation energy ratio) basée sur une analyse par banc de filtres de modulation d'inspiration auditive des enveloppes temporelles des signaux de parole et de hauteur. En outre, peu de modèles ont été développés pour prédire les évaluations de la qualité audio par des auditeurs malentendants .Par exemple, HASQI (Hearing-Aid Speech Quality Index) prend en compte l'effet du bruit, de la distorsion non linéaire et du filtrage linéaire sur la qualité de la parole perçue ; Cependant, il est très sensible à la distorsion du motif loudness. À son tour, Beerends et al. ont présenté le PREQUEL (évaluation de la qualité de la reproduction perceptuelle pour les haut-parleurs) qui simule les enregistrements binauraux des signaux de référence à l'aide d'un simulateur de tête et du torse pour quantifier la qualité sonore perçue des haut-parleurs en évaluant leur sortie acoustique. Ces dernières années, le développement de méthodes hybrides gagne également du terrain [8].

### 2.3.2.3 Méthodes Paramétriques

Les modèles paramétriques estiment la qualité à l'aide de spécifications du processus de conception du réseau et / ou des paramètres, tels que l'écho, perte d'insertion pondérée en fréquence (dénommé «indice de sonie») et la perte de paquets. La plupart de ces spécifications peuvent être modélisées avec précision par un petit nombre de mesures statistiques. Un exemple bien connu d'approche paramétrique est la recommandation UIT-T P.563, qui utilise des dispositifs de mesure non intrusifs en service (INMD). Un INMD évalue les paramètres objectifs des canaux téléphoniques sur le trafic d'appels en direct sans entraver l'appel, et avec une connaissance du réseau et le système auditif humain produit des valeurs de qualité.[8]

### 2.3.3 Evaluation objective de l'audiovisuel :

Généralement, nous regardons des vidéos avec une bande-son d'accompagnement. Par conséquent, des métriques complètes de qualité audiovisuelle sont nécessaires pour analyser les deux modalités de contenu multimédia ensemble. La qualité audiovisuelle comprend un facteur : de l'interaction entre la qualité audio et la qualité vidéo.[34][35]

La tolérance vis-à-vis de l'audio est toujours plus élevée que l'inverse, probablement dans la mesure où il s'agit également d'un phénomène plus naturel dans le monde réel, où la lumière voyage plus rapidement que le son. Des résultats similaires ont été rapportés dans des expériences avec des clips non-parole montrant un batteur [36]. L'interaction entre les signaux audio et vidéo est un autre facteur influant sur l'évaluation de la qualité globale du contenu multimédia, comme le montrent des études en neuroscience .

Étant donné que la plupart des mesures de qualité audiovisuelle existantes sont proposées sur la base d'une combinaison d'évaluation de la qualité audio et vidéo, l'étude [35] analyse l'influence réciproque de la qualité audio, de la qualité vidéo et de la qualité audiovisuelle. Sur la base de l'analyse expérimentale, l'étude tire plusieurs conclusions générales comme suit :

-Premièrement, la qualité audio et la qualité vidéo contribuent à la qualité audiovisuelle globale, et leur multiplication obtient la corrélation la plus élevée avec la qualité globale.

-Deuxièmement, la qualité globale est dominée par la qualité vidéo en général, tandis que la qualité audio est plus importante que la qualité vidéo dans les cas où les débits binaires audio et vidéo codés sont faibles ou la qualité vidéo est supérieure à un certain seuil. Lorsque la qualité audio diminue, l'influence de la qualité audio augmente sur la qualité globale. De plus, avec les applications dans lesquelles l'audio est évidemment plus important que le contenu vidéo (comme la téléconférence, les actualités, les vidéos musicales, etc.), la qualité audio domine la qualité globale.

-Enfin, la qualité audiovisuelle est également influencée par d'autres facteurs, notamment les informations de mouvement et la complexité du contenu vidéo . [35]

Les études sur les métriques de qualité audio / vidéo ont principalement porté sur les applications à faible débit, telles que les communications mobiles, où le flux audio peut utiliser une partie importante du débit total [37][38]. Certaines études portent sur l'évaluation de la qualité audiovisuelle pour les applications de vidéoconférence [39][40][41]. Les données de test utilisées dans ces études sont très différentes en termes de plage de contenu et de distorsion, et ces modèles permettent d'obtenir de bonnes performances de prévision .

Bien que certaines études aient examiné les mesures de la qualité audiovisuelle, l'évaluation conjointe de la qualité audiovisuelle a progressé lentement. L'interaction entre la perception audio et vidéo est compliquée et la perception du contenu audiovisuel manque toujours d'une enquête approfondie. Actuellement, de nombreuses métriques de qualité sont proposées sur la base de la fusion linéaire de la qualité audio et vidéo, mais la plupart des études choisissent des paramètres de fusion de manière empirique sans support théorique et avec peu ou pas d'intégration dans le calcul métrique. Cependant, l'évaluation de la qualité audiovisuelle mérite d'être examinée plus avant en raison de son application étendue au codage du signal, à la transmission du signal , etc...

## Conclusion :

Dans ce chapitre nous avons présenté les méthodes d'évaluation de la qualité vidéo qui se divisent en deux grandes familles : les méthodes subjectives basées sur l'observateur humain, en plus l'implication de ces observateurs non experts dans la procédure d'évaluation pour affecter le score donné à une vidéo. Les deuxièmes techniques sont les méthodes objectives où l'introduction de l'observateur humain n'est pas nécessaire. Les résultats obtenus par ces méthodes doivent être conformes à l'évaluation subjective. Quelques méthodes d'évaluation objectives sont présentées dans ce chapitre telles que PSNR , SSIM , etc...

Nous avons présenté aussi les méthodes d'évaluation de la qualité audio qui se divisent en deux grands familles : les méthodes subjectives qui contient la méthode MUSHRA et les méthodes objectives basée sur les méthodes intrusives , les méthodes non intrusives et finalement les méthodes paramétriques .

Et nous avons présenté également les méthode d'évaluation de la qualité audiovisuel qui se divise au deux familles importantes : les méthodes subjectives qui contient les méthodes ACR , DCR ,PC et les méthodes objectives basée sur plusieurs études.

Dans le chapitre suivant nous allons voir le détail de l'évaluation objective de la qualité audiovisuel .

# Evaluation objective de la qualité audiovisuel

## Introduction

La base perceptuelle et cognitive de l'évaluation de la qualité audiovisuelle, c'est-à-dire à quel stade de la chaîne de traitement de la perception les modalités auditifs et visuels sont réellement fusionnées, n'est pas encore totalement déterminée.

dans ce chapitre nous allons voir L'évaluation objective de la qualité audiovisuel et ces différentes approches .

### 3.1 Les méthodes d'évaluation objectives de la qualité perçue audiovisuel :

Il existe de nombreuses théories sur l'évaluation de la qualité des modalités individuelles, y compris les processus psychophysiques impliqués dans leur perception de la qualité. Cependant, l'évaluation de la qualité audiovisuelle qui est un processus d'information multimodale, est un domaine relativement peu exploré.

#### 3.1.1 L'approche de la fusion des deux modalités :

Bien que divers détails du traitement neurophysiologique des données audiovisuelles restent inconnus, des études empiriques ont montré que les domaines auditif et visuel ont une influence réciproque sur la qualité audiovisuelle globale perçue.

Cependant, la majorité des chercheurs ont adopté la théorie de la fusion tardive, dans laquelle les canaux auditifs et visuels sont traités en interne pour produire des valeurs de qualité respectives qui sont intégrées à un stade avancé pour former une seule qualité globale perçue. [42]

La qualité audiovisuelle est donc généralement décrite comme une fusion de deux dimensions (qualités audio et vidéo), comme illustré à la Figure 3.1.

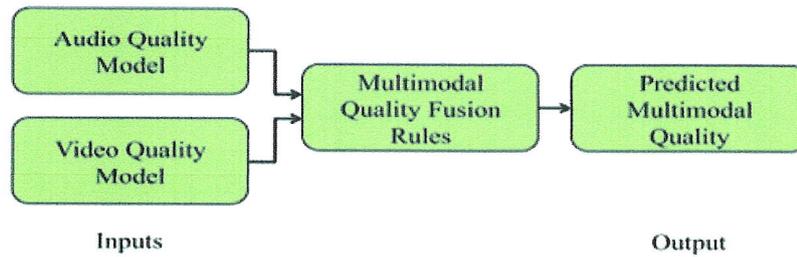


FIGURE 3.1 – Modèle d'estimation de la qualité multimédia de base

Le modèle de fusion le plus couramment utilisé et adopté dans plusieurs études [42], [43], [44] est celui présenté dans l'équation suivante :

$$QAV = a_0 + a_1QA + a_2QV + a_3QAQV \quad (3.1)$$

Où  $QAV$ ,  $QA$ ,  $QV$  et  $a_0$ ,  $a_1$ ,  $a_2$ ,  $a_3$  sont respectivement la qualité audiovisuelle, la qualité audio, la qualité vidéo et les poids prédits.

Toutefois, il convient de noter qu'il n'existe pas de valeurs ni de dérivations communément admises pour les quatre paramètres de fusion ( $a_0$ ,  $a_3$ ) dans l'équation 3.1, les valeurs rapportées dans la littérature vont de  $a_0 = [-3.34, 4.26]$ ,  $a_1 = [-0.19, 0.85]$ ,  $a_2 = [0, 0.89]$ , à  $a_3 = [-0.01, 0.26]$ .

Peu d'études sur la compréhension cognitive humaine suggèrent que les canaux audio et vidéo pourraient être intégrés dans une phase précoce de la formation de la perception humaine [45]. Sur cette base, plusieurs chercheurs [42], [46] ont proposé des modèles de qualité audiovisuelle comme une multiplication de qualité audio et vidéo d'égale importance, comme le montre l'équation suivante :

$$QAV = a_0 + a_1QAQV \quad (3.2)$$

De même, Martineza et al. [47] ont proposé trois mesures de la qualité perçue audiovisuelle. Le premier modèle est un modèle linéaire simple donné par l'équation suivante :

$$QAV = a_0 + a_1QA + a_2QV \quad (3.3)$$

La deuxième mesure est basée sur le modèle pondéré de Minkowski comme suit :

$$QAV = (a_1QA^P + a_2QV^P)^{\frac{1}{P}} \quad (3.4)$$

Où les valeurs de la puissance de Minkowski ( $P$ ) sont toutes comprises entre 1 et 1,2. Sur la base de ces résultats, nous avons varié la valeur de  $P$  dans la plage de 0,9 à 1,3 et répété la procédure d'ajustement pour chacune de ces valeurs.

Comme certaines études [46], [48] suggèrent que la modalité visuelle peut être plus dominante que l'audio dans la formation de la qualité audiovisuelle perçue, en particulier

pour les vidéos avec des données de mouvement élevées, ainsi les auteurs dans [42] présentent ainsi l'équation suivante :

$$QAV = a_0 + a_1QV + a_2QAQV \quad (3.5)$$

Bien que les modèles dans les équations prétendants atteint assez précise la qualité audiovisuelle prévue dans certaines études lorsque des durées de qualité audio et vidéo sont les mêmes, il ne reflète pas les différences de l'influence de seulement audio et vidéo uniquement des stimuli sur la qualité globale. En outre, ils ne peuvent pas non plus pleinement saisir certains autres facteurs influents, tels que l'objectif de l'évaluation, l'environnement de test, l'impact des dégradations.

### 3.1.2 L'approche de prévision de la qualité audiovisuelle :

L'UIT-T a proposé certains modèles normalisés de prévision de la qualité audiovisuelle, par exemple, les Recommandations UIT-T P.1201 [49], UIT-T G.1070 [50] et UIT-T G.1071[51].

**Le modèle UIT-T P.1201** est destiné à l'estimation de la qualité audiovisuelle des services de diffusion. C'est un modèle non intrusif d'information d'entête de paquet visant la surveillance de service et le benchmarking du streaming UDP. Le modèle prend en charge aussi bien les applications à basse résolution telles que la télévision mobile ainsi que les applications à plus haute résolution telles que l'IPTV. Le modèle utilise les informations extraites de l'entête du paquet ainsi que les informations fournies hors bande. Il fournit des prédictions distinctes de la qualité audio, vidéo et audiovisuelle sous forme de résultat en terme du MOS à 5 points. Le modèle a été validé pour la compression, la perte de paquets ainsi que le buffering des altérations de l'audio et de la vidéo avec des débits différents. Le contenu vidéo des différentes complexités spatiotemporelles avec différentes images clés (keyframes), cadences d'images (frame rates) et résolutions vidéo a été sélectionné. Le modèle ITU-T Rec. P.1201 a testé plus de 1166 échantillons à des résolutions inférieures et a testé plus de 3190 échantillons à des résolutions plus élevées. Les valeurs de corrélation RMSE et de Pearson pour la modélisation audiovisuelle ont été évaluées respectivement à 0,470 et 0,852 pour les applications à résolution inférieure et à 0,435 et 0,911 pour les applications à plus haute résolution.

**Le modèle UIT-T G.1070** représente un modèle de planification recommandé pour la téléphonie vidéo. Dans ce modèle, la qualité multimédia globale est calculée en employant des paramètres de réseau et d'application ainsi que des paramètres sur le terminal. Il propose un algorithme qui estime la qualité du vidéophone pour la qualité de l'expérience et la qualité des planificateurs de services. Le modèle fournit aussi des estimations sur la qualité multimédia qui tiennent compte de l'interactivité pour permettre aux planificateurs

d'éviter de sous-dimensionner le service. Le modèle contient trois fonctions principales d'évaluation de la qualité de la parole, de la qualité vidéo et de la qualité multimédia globale. La fonction d'estimation de la qualité de la parole est similaire à l'outil E-model et prend comme paramètres d'entrée le type de codec vocal, le taux de perte de paquets, le débit binaire et le niveau sonore d'écho de la parole. La fonction vidéo est générée pour le contenu «head-and-shoulders» et prend comme paramètres d'entrée le format vidéo, la taille d'affichage, le type du codec, le taux de perte de paquets, le débit binaire, l'intervalle d'images clés ainsi que le taux d'images. La fonction multimédia intègre séparément la qualité audio et la qualité vidéo en incluant l'asynchronisme audiovisuel (audiovisual asynchrony) et le délai de bout en bout. Sur des ensembles de données précis, la précision du modèle d'évaluation de la qualité des communications multimédias en terme de corrélation de Pearson est de 0,83 pour QVGA et de 0,91 pour la résolution QQVGA . L'application du modèle est limitée à la planification de la QoE et de la QoS. Des applications telles que le benchmarking et le suivi de la qualité ne sont pas couvertes par la recommandation.

Bien que le **modèle UIT-T G.1071** est recommandé pour la planification réseau des services de diffusion audio et vidéo. Cette recommandation concerne les domaines d'application à plus haute résolution (HR) tels que l'IPTV et les domaines d'application de résolution inférieure (LR) comme la TV mobile. L'application des modèles est limitée à la planification de la qualité d'expérience (QoE)/qualité de service (QoS). Le benchmarking et le suivi de la qualité ne font pas partie du cadre de cette recommandation. Le modèle prend en entrée les hypothèses de planification de réseau telles que la résolution vidéo, les types et profils de codecs audio et vidéo, les débits audio et vidéo et le taux de perte de paquets. Il fournit en sortie des prédictions distinctes de la qualité audio, vidéo et audiovisuelle définies sur l'échelle MOS à 5 points. Notons que des cas d'utilisation tels que la dégradation audio et vidéo au cours du buffering, les situations de transcodage, les effets du bruit et du délai sur l'audio, la diffusion audiovisuelle avec une adaptation significative du débit ne font pas partie du modèle. À noter aussi que des tests ont montré que, pour les applications à basse résolution utilisant les bases de données d'apprentissage et le test ITU-T P.1201.1, les coefficients d'estimation de la qualité audiovisuelle RMSE et de la corrélation de Pearson atteignent respectivement 0.5 et 0.83. Pour les applications à haute résolution et pour des bases de données d'apprentissage et de validation ITU-T P.1201.2, les coefficients d'estimation de la qualité audiovisuelle RMSE et Pearson atteignent 0,51 et 0,87 respectivement.

### 3.1.3 L'approche d'apprentissage automatique :

Le monde contemporain de l'apprentissage automatique consiste en un nombre incalculable d'algorithmes ainsi que de leurs implémentations dans diverses bibliothèques. Certaines de ces méthodes sont destinées uniquement aux problèmes de classification. Cependant, plusieurs algorithmes sont adaptés à des problèmes de classification aussi bien que de

régression. Nous décrivons brièvement les méthodes d'apprentissage automatique que nous avons étudiées en détail et avons adaptées afin d'obtenir les meilleures performances.

**Méthodes d'ensemble basées sur l'arbre de décision :** Les arbres de décision (Decision Trees, DT) sont des structures de données hiérarchiques qui peuvent être utilisées pour des problèmes de classification et de régression en utilisant efficacement la stratégie de «Diviser-et-conquérir» (divide-and-conquer). Un arbre de décision est composé de noeuds de décision internes où un test est appliqué à une entrée donnée. Il se compose aussi de branches conduisant à des valeurs d'attribut. Une valeur de classification ou de régression est attribuée par les noeuds feuilles. Le processus d'estimation provient du noeud racine, traverse les noeuds de décision jusqu'à ce qu'un noeud soit atteint.

La structure arborescente permet une découverte rapide des noeuds qui couvrent une entrée. Dans un arbre binaire, la traversée de chaque noeud de décision exclut la moitié des cas. En raison de la convergence rapide et de la facilité d'interprétation, les arbres binaires sont parfois préférés à des méthodes plus précises .

L'estimation peut être calculée que le modèle soit paramétrique ou non paramétrique. Pour l'estimation paramétrique, le modèle est construit sur l'ensemble de l'espace d'entrée à partir de données d'apprentissage ; une structure arborescente statique est formée. Le même modèle est ensuite utilisé pour faire des estimations dès que les données de test sont disponibles. Pour l'approche non paramétrique, la structure arborescente n'est pas statique et, au cours du processus d'apprentissage, elle pousse au fur et à mesure que les branches et les feuilles sont ajoutées .

Les arbres de décision présentent un biais (bias) faible et une variance très élevée, ce qui entraîne des problèmes d'ajustement lorsqu'ils se développent très profondément. Pour réduire la variance, les méthodes d'ensemble basées sur l'arbre de décision ont été construites. Les forêts d'arbres décisionnels sont en effet un ensemble de méthodes d'apprentissage pour la classification et la régression utilisant différents modèles d'arbres de décision pour obtenir une meilleure performance de prédiction. Durant la phase d'apprentissage, un tableau d'arbres de décision est formé et un sous-ensemble de données d'apprentissage, choisi au hasard, est utilisé pour former chaque arbre. Dans un problème de classification, les entrées sont soumises à chaque arbre dans la forêt d'arbres décisionnels et ceci afin d'obtenir un vote pour une classe. Un modèle de forêt d'arbres décisionnels recueille tous les votes, puis choisit la classe avec le plus grand nombre de votes. Ce comportement réduit les problèmes de variance élevée dont nous avons parlé plus haut. Cependant, comme il existe un compromis entre le biais et la variance, la classification RF introduit une légère augmentation du biais tout en réduisant la variance. Dans l'ensemble, ce modèle fournit des améliorations significatives en terme de précision de classification .

Les forêts d'arbres décisionnels n'ont que deux paramètres à régler pendant la phase d'apprentissage. Ces paramètres sont le nombre d'arbres dans la forêt et le nombre de variables dans le sous-ensemble aléatoire à chaque noeud .

Au lieu de rechercher un seul modèle supérieur, les chercheurs ont remarqué que le fait de combiner de nombreuses variations produit de meilleurs résultats avec un peu d'effort supplémentaire. Comme nous pouvons le remarquer, pour les forêts d'arbres décisionnels, les modèles d'apprentissage génèrent de nombreux classificateurs et combinent leurs résultats. Cette approche a récemment suscité un grand intérêt. Les deux méthodes d'apprentissage les plus connues sont le boosting et les techniques de bootstrap. Pour ces deux méthodes, l'algorithme d'apprentissage combine les prédictions de plusieurs modèles de base.

Dans la construction de modèles basés sur les arbres de décision avec des méthodes utilisant les techniques de bootstrap, chaque arbre est construit avec une variation aléatoire de l'ensemble de données d'apprentissage. La prévision est réalisée par un simple vote majoritaire. Le but est d'améliorer la stabilité et la précision. Cette approche réduit considérablement la variance et contribue à éviter les problèmes d'ajustement, mais elle augmente légèrement le biais. Bien qu'elle soit généralement appliquée aux arbres de décision, cette approche peut être aussi bien utilisée avec n'importe quel type de méthode .

Pour les méthodes de boosting, la prédiction dépend aussi des arbres antérieurs. Dans cette approche, les points mal prédits par les arbres précédents reçoivent un poids supplémentaire par les arbres successifs. Les méthodes de boosting visent principalement à réduire le biais et éventuellement la variance tout en créant un seul apprenant fort (strong learner) parmi un ensemble d'apprenants qui sont faibles .

**Régression symbolique et programmation génétique :** La technique de régression symbolique vise à identifier une expression mathématique sous-jacente qui correspond le mieux à un ensemble de données. Elle consiste à trouver simultanément la forme des équations ainsi que les paramètres. La régression symbolique commence par former une expression initiale en combinant aléatoirement des blocs de construction mathématiques et puis continuer à former de nouvelles équations en recombinaison les équations précédentes en utilisant la programmation génétique (GP).

La programmation génétique est une technique de calcul qui nous permet de trouver une solution à un problème sans connaître préalablement la forme de la solution. Elle est basée sur l'évolution d'une population de programmes informatiques où les populations sont stochastiquement transformées en de nouvelles populations génération par génération .

La programmation génétique découvre la performance d'un programme en l'exécutant, en mesurant son résultat et en comparant ce résultat à un objectif défini. Cette comparaison est appelée «fitness». Dans le domaine de l'apprentissage automatique, ceci est équivalent à trouver 'le score', 'l'erreur' ou 'la perte'. Dans chaque génération, les programmes qui réussissent sont marqués pour la reproduction et sont ensuite utilisés pour produire de nouveaux programmes pour la génération suivante. Le croisement (crossover) et la mutation sont les principales opérations génétiques aidant à créer de nouveaux programmes à partir d'un ensemble de programmes existants. Pour l'opération de croisement, un programme enfant est généré en joignant des parties choisies au hasard à partir de deux programmes sélectionnés

de la génération précédente. Cependant, pour la mutation, un programme enfant est créé à partir d'un seul parent de la génération précédente en modifiant aléatoirement un segment arbitraire .

La programmation génétique utilise généralement les arbres dans le but de manipuler des programmes. Les appels de fonction dans l'arbre sont représentés par les noeuds et les valeurs associées aux fonctions sont représentées par les feuilles . Les programmes de la programmation génétique combinent plusieurs composantes dans des formes plus avancées. Dans ce cas, chaque composant est représenté par un arbre qui se regroupe avec d'autres arbres sous le noeud racine.

Tout comme l'ensemble de méthodes que nous avons vu dans la section précédente, les populations initiales de la programmation génétique sont généralement générées de façon aléatoire. Ces populations initiales sont catégorisées comme pleine (full), croissance (grow) et combinée (ramped half and half) selon leur profondeur .

Les deux méthodes full et grow limitent la profondeur maximale des individus initiaux générés. Elles diffèrent les unes des autres quant à la taille et à la forme des arbres générés. Pour la méthode

full, les arbres sont générés où toutes les feuilles sont à la même profondeur. Les arbres de la méthode grow sont générés dans différentes tailles et formes. La méthode Ramped half-and-half propose de combiner les deux méthodes full et grow. Dans cette approche, la méthode full est utilisée pour construire la moitié de la population initiale et la méthode grow est utilisée pour construire l'autre moitié .

La programmation génétique choisit les individus de façon probabiliste en fonction de leur fitness, puis leur applique des opérations génétiques. Ce processus entraîne de meilleurs individus ayant probablement plus de programmes enfants que les individus inférieurs. Deux méthodes communes de sélection individuelle en programmation génétique sont la sélection du tournoi (tournament selection) et la sélection proportionnelle de fitness (fitness proportionate selection).

**Apprentissage profond :** L'apprentissage profond remonte aux années 1940 et a été renommé à plusieurs reprises, reflétant l'influence de différents chercheurs et de différentes perspectives. Cette appellation spécifique est très récente .

Un exemple typique d'un modèle d'apprentissage profond est «feedforward Deep Network» ou le perceptron multicouche (Multi-Layer perceptron, MLP) (Bengio et al., 2015). MLP ne fait aucune hypothèse sur les relations entre les variables. En général, ces modèles utilisent trois couches principales : une couche d'entrée de neurones représentant le vecteur d'entrée, une ou plusieurs couches intermédiaires «cachées» et des neurones de sortie qui représentent le vecteur de sortie. Les noeuds de chaque couche sont liés à tous les noeuds des couches adjacentes. Ces liens sont utilisés pour transmettre des signaux d'un neurone à l'autre .

Les non-linéarités sont représentées dans le réseau par les fonctions d'activation et de

transfert dans chaque noeud. Chaque noeud gère un calcul de base alors que leurs liens permettent un calcul global. Le comportement global d'un réseau neuronal est influencé par le nombre de couches, le nombre de neurones dans chaque couche, la façon dont les neurones sont liés et les poids associés à chaque lien. Le poids associé à chaque lien définit comment un premier neurone influence le deuxième neurone. Les poids sont révisés durant la période d'apprentissage. Avec cette approche, les couches masquées captent les complexités dans les données tandis que les poids sont ajustés dans chaque itération afin d'obtenir la plus faible erreur dans la sortie. L'algorithme d'apprentissage utilisé est celui de la rétropropagation du gradient (gradient descent back propagation) .

Dans l'approche de la rétropropagation du gradient et pendant la phase directe, le signal d'entrée est propagé par le réseau couche par couche. Dans le noeud de sortie, le signal d'erreur est calculé et est renvoyé au réseau dans ce qu'on appelle la phase de régression (backward phase). Durant cette phase, les paramètres du réseau sont modifiés afin de minimiser l'erreur de signal. Les méthodes d'apprentissage profond peuvent être utilisées dans les problèmes de régression ainsi que dans les applications de regroupement et de classification (clustering and classification).

## **Conclusion**

Dans ce chapitre nous avons parlé de l'évaluation objective de la qualité audiovisuel, cette dernier contient trois approches : l'approche de la fusion de deux modalités , l'approche de prévision de la qualité audiovisuelle ainsi que l'approche d'apprentissage automatique , où on a choisie la première puisque malgré que les autres approches atteignent une précision de prédiction élevée, leurs applications sont intrinsèquement limitées. Dans le prochaine chapitre on va donné plus de détail sur notre choix.

# La méthode choisie , les résultats , les discussions

## Introduction

Dans ce chapitre nous allons l'entrée de notre méthode quand nous a sélectionné d'après les approches qu'on a déjà vu dans le chapitre précédent . Dans ce cas on choisit de trouver des valeurs plus précises de la qualité , le choix de ces dernières faite la différence de qualité donc ces valeurs sont très importantes pour une évaluation plus fiable . Ensuite en va étudier et analyser les résultats obtenus par notre méthode et en les comparant avec les études qui ont déjà été réalisées .

## 4.1 Environnement de travail

### 4.1.1 Langage

Nous avons utilisés Python 2.7 comme langage de programmation, parmi les raisons de cette utilisation :

- ✓ Python fonctionne sur différentes plateformes (Windows, Mac, Linux).
- ✓ Il a une syntaxe simple similaire à la langue anglaise .
- ✓ Ce langage peut être traité de manière procédurale , de manière orientée objet ou de manière fonctionnelle.

### 4.1.2 Caractéristique de la plateforme

On utilisé une machine avec les caractéristiques suivantes :

- ✓ **PROCESSEUR** : 2.60 GHz Intel5(R) Core(TM) i5-4300M .
- ✓ **RAM** : 4.00 GO .
- ✓ **CARTE GRAPHIQUE** : Intel HD 4600 .
- ✓ **SYSTEME D'EXPLOITATION** : Windows 8 64 bits.

## 4.2 La base de donnée utilisé :

Pour valider notre méthode objective nous comparons nos résultats d'évaluation avec les résultats des études qui sont déjà été réalisé [52].

L'ensemble de données de qualité audiovisuelle de database INRS [53] composé de 160 vidéo audiovisuel, comprenant divers paramètres de compression du média et de distorsion du réseau, tels que la fréquence d'images, la quantification et les paramètres de réduction du bruit, ainsi que le taux de perte de paquets. Les valeurs de plage de paramètres de compression et de distorsion réseau sont sélectionnées pour correspondre aux cas d'utilisation de communications en temps réel. Le codec vidéo H.264 avec une résolution de 720p et le codec audio AMR-WB sont utilisés pour coder les flux vidéo et audio. Trente observateurs ont évalué la qualité audiovisuelle globale sur l'échelle de qualité subjective ACR (Absolute Category Rating) sur 5 niveaux dans un environnement contrôlé. L'ensemble de données comprend les valeurs MOS, les taux de perte de paquets mesurés au niveau du flux de bits pour les flux audio et vidéo, les paramètres de compression et diverses informations d'en-tête de paquet. Nous avons utilisé un logiciel open source pour produire des séquences audiovisuelles source, une diffusion en continu de bout en bout et un lecteur vidéo personnalisé. Ces outils et le jeu de données sont librement accessibles au public à des fins de recherche et de développement.

### 4.2.1 L'ensemble de données sur la qualité audiovisuelle de l'INRS :

Les détails sont donnés dans la publication [54] [55] [56]

Liste de fichiers :

✓ Dossier MOS :

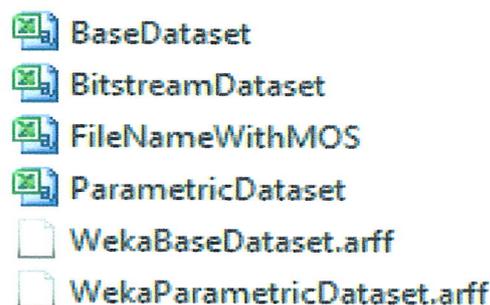


FIGURE 4.1 – Dossier MOS

Ce dossier contient divers fichiers de score MOS. Ces fichiers sont :

FileNameWithMOS.csv : contient le nom des fichiers audiovisuels et des scores MOS calculés.

GeneralFileName	MOS
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr0.1_amrwb_m08	4.10344827586
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr0.5_amrwb_m08	2.81481481481
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr0_amrwb_m08	4.10344827586
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr1_amrwb_m08	1.73214285714
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr5_amrwb_m08	1.15
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr0.1_amrwb_m08	3.94827586207
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr0.5_amrwb_m08	3.20833333333
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr0_amrwb_m08	3.8

FIGURE 4.2 – Fichier FileNameWithMos

Base Dataset.csv : contient 5 paramètres indépendants : fréquence d'images vidéo, paramètre de quantification, réduction du bruit, taux de perte de paquets pour les flux audio et vidéo et scores MOS calculés [54] [55].

VideoFrameRate	NR	QP	VideoPacketLossRate	AudioPacketLossRate	MOS
10	0	23	0.04	0.09	4.10344827586
10	0	23	0.46	0.42	2.81481481481
10	0	23	0.0	0.0	4.10344827586
10	0	23	1.23	0.98	1.73214285714
10	0	23	4.2	5.46	1.15
10	999	23	0.0	0.19	3.94827586207
10	999	23	0.55	0.51	3.20833333333

FIGURE 4.3 – Fichier Base Dataset

- ✓ **Output Video Files Folder** : ce dossier contient les fichiers vidéo de sortie sauvegardés.
- ✓ **Dossier Source Video Files** : Il comprend les fichiers vidéo source et les CAPS audio et vidéo utilisés dans GStreaming.
- ✓ **Dossier Subject Details** : il contient les scores ACR collectés auprès de tous les observateurs. Ce sont des notes brutes non traitées. Aucun critère de consolidation ou de rejet n'est appliqué.

GeneralFileName	ACRScore	Time
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q35_mr999_plr0.5_amrwb_m08	3	12792
ntia_HeadShoulders-Female15_h264_720p_baseline_fps25_q27_mr0_plr0.5_amrwb_m08	4	12151
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q35_mr999_plr1_amrwb_m08	3	12393
ntia_HeadShoulders-Female15_h264_720p_baseline_fps20_q35_mr0_plr1_amrwb_m08	3	12934
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_mr0_plr0.1_amrwb_m08	5	15534
ntia_HeadShoulders-Female15_h264_720p_baseline_fps25_q35_mr0_plr1_amrwb_m08	2	11872
ntia_HeadShoulders-Female15_h264_720p_baseline_fps20_q23_mr0_plr0.5_amrwb_m08	2	12354
ntia_HeadShoulders-Female15_h264_720p_baseline_fps25_q35_mr0_plr0_amrwb_m08	3	18874

FIGURE 4.4 – Dossier Subject Details

- ✓ **Subject Merged Folder** : Scores ACR consolidés avec informations de synthèse. Les scores ici sont après application des critères de rejet. Reportez-vous à [54] pour plus de détails.

### 4.3 Description de notre méthode :

On passe au détail de notre méthode étape par étape :

**La première étape** : Dans cette étape il est nécessaire de faire un détachement de l'audiovisuel de database INRS [53] et on obtient pour chaque fichier audiovisuel deux fichiers , l'un pour les vidéos et l'autre pour les audio .

**La deuxième étape** : nous calculons la qualité des vidéo et des audio de la database par deux méthodes (soit la méthode PSNR , soit la méthode SSIM), on les a choisie parce que ces derniers sont parmi les méthodes les plus utilisé dans l'évaluation de la qualité vidéo ainsi que la qualité audio , voir le tableau suivant :

Média	Métrique
vidéo	PSNR SSIM
audio	PSNR SSIM

TABLE 4.1 – Métrique de référence complète[57]

Comme c'est déjà mentionner dans **chapitre 2** les deux fonctions sont définie par :

**Equation PSNR** :

$$PSNR = 10 \log \frac{\text{Valeur crête du signal original}}{EQM}$$

Avec EQM est définie par

$$EQM = \frac{1}{M.N} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - I_d(i, j))^2$$

Pour que PSNR soit parfait il faut qu'il soit supérieur ou égale à 36 .

**Equation SSIM :**

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma$$

Avec  $l(x, y)$  est définie par

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

Et  $c(x, y)$  est définie par

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

ainsi que  $s(x, y)$  est définie par

$$s(x, y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

Intervalle de SSIM est entre  $[0, 1]$  et pour qu'il soit parfait il faut qu'il se rapproche de 1 .

**La troisième étape :** pour ce faire , on fusionne les deux qualité avec l'équation 3.1 , cette dernière est la plus utiliser selon l'étude [52] où ils ont cité que notre équation fonctionne bien dans les expérience qui sont déjà été effectuées ainsi que les résultats obtenus qui sont aussi très similaires aux résultats des études subjective .

Dans cette étape on vas s'approfondir dans notre méthode en détaillant son fonctionnement s'effectuera à partir de l'équation :

$$QAV = a_0 + a_1QA + a_2QV + a_3QAQV$$

où  $a_0 = -3.34$  ,  $a_1 = 0.85$  ,  $a_2 = 0.76$  ,  $a_3 = -0.01$  [58]

Tout d'abord on a fixé les valeurs  $a_1$  ,  $a_2$  ,  $a_3$  , tout en changeant à chaque fois la valeur de  $a_0$  sur tous le long de intervalle  $[- 3.34 , 4.26]$  pour obtenir enfin une liste de résultat de la qualité audiovisuel de la database INRS [55].

Après on a calculer le Coefficient de Corrélation linéaire (CC) entre la liste obtenue et les valeurs subjectives (Mos) avec cette équation :

$$CC = \frac{1}{NbJK} \frac{\sum_{j=1}^J \sum_{k=1}^{K(j)} (MOS_{jk} - \overline{MOS})(Mobj_{jk} - \overline{Mobj})}{\sqrt{\sigma^2 MOS . \sigma^2 Mobj}} \quad (4.1)$$

Avec :

$$NbJK = \sum_{j=1}^J K(j)$$

Le nombre des vidéos testées (on suppose que J versions originales ont subi chacune K(j) dégradations différentes).

Le MOS moyen comme équation :

$$\overline{MOS} = \frac{1}{NbJK} \sum_{j=1}^J \sum_{k=1}^K (j) MOS_{jk} \quad (4.2)$$

La mesure objective moyenne comme équation :

$$\overline{Mobj} = \frac{1}{NbJK} \sum_{j=1}^J \sum_{k=1}^K (j) Mobj_{jk} \quad (4.3)$$

Avec Mobj : c'est la qualité audiovisuel QAV obtenue de l'étape 3 .

La variance de MOS comme équation :

$$\sigma^2 MOS = \frac{1}{NbJK - 1} \sum_{j=1}^J \sum_{k=1}^K (j) (MOS_{jk} - \overline{MOS})^2 \quad (4.4)$$

La variance des mesures objectives comme équation :

$$\sigma^2 Mobj = \frac{1}{NbJK - 1} \sum_{j=1}^J \sum_{k=1}^K (j) (Mobj_{jk} - \overline{Mobj})^2 \quad (4.5)$$

La valeur du coefficient de corrélation CC est toujours comprise entre -1 et 1. Une valeur proche de 1 ou de -1 indique une bonne corrélation.

Le même travail a été exécuté avec la valeur de a1 sur intervalle [- 0.19 , 0.85] tout ont fixant a0 , a2 , a3 . Ainsi que a2 en respectant intervalle [ 0 , 0.89] et ça en réglant les a0 , a1 , a3 sur ces propre valeurs et encore on vas réalisé le même travaille sur le a3 sur tout son intervalle qui est [- 0.01 , 0.26] tout en fixent les valeurs de a0 , a1 , a2 .

Après on a sélectionné les valeurs de ai qui ont leurs corrélation proche de 1, ainsi qu'on va calculer la qualité audiovisuel avec ces meilleur de ai de chaque audiovisuel de database INRS , ensuite on doit faire un calcul de coefficient de corrélation CC entre les résultats obtenue et les résultats subjective (MOS) pour obtenir enfin une valeur standard de CC .

## 4.4 Analyse et interprétation des résultats

Puisque notre méthode propose une évaluation objective de la qualité audiovisuel, nous allons calculer la qualité audio et la qualité vidéo où on a choisie la méthode SSIM puisque elle est connue qu'elle donne des résultats plus satisfiable que PSNR . Les figures ci dessous montre quelques résultats obtenues :

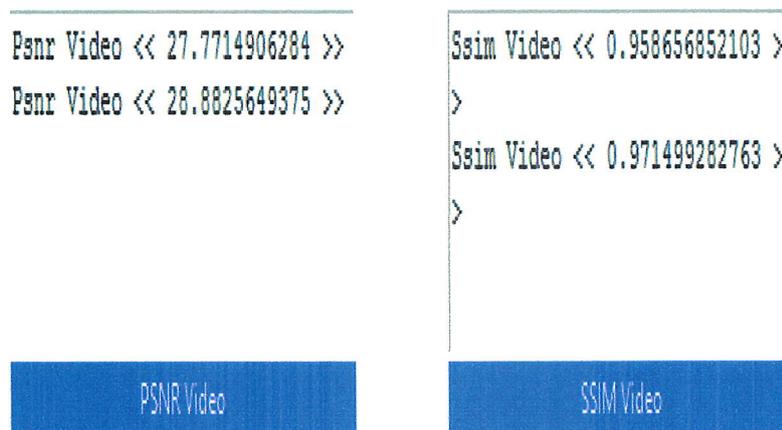


FIGURE 4.5 – Résultats obtenues de PSNR et SSIM pour un vidéo de la base INRS

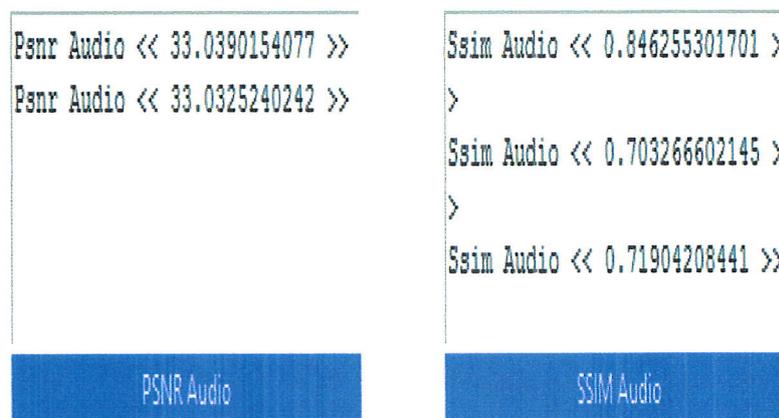


FIGURE 4.6 – Résultats obtenues de PSNR et SSIM pour un audio de la base INRS

Ainsi que nous allons utilisé une technique largement appliqué qui est la fusion de la qualité audio et la qualité vidéo avec un choix des valeur prédite qui vous serons démontré lors de ces figures :

**La valeur maximal de  $a_0$  :**

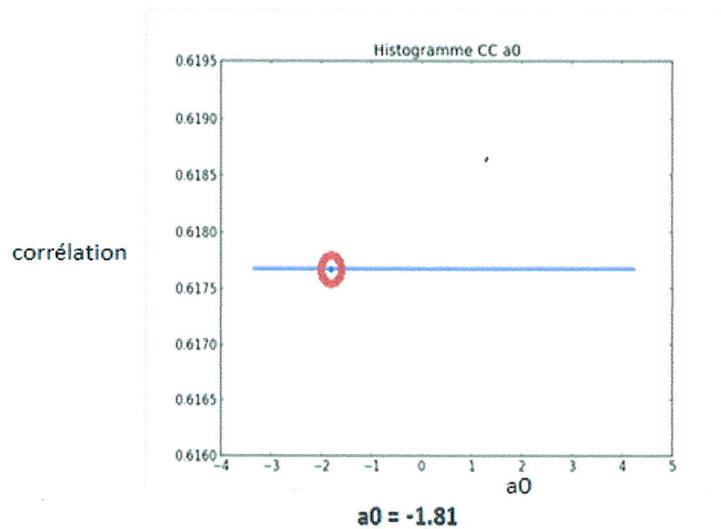


FIGURE 4.7 – Histogramme de a0

On remarque que le coefficient de corrélation CC de a0 est stable sur le long de son intervalle , alors on a choisi une valeur aléatoire de a0 de cette histogramme comme valeur maximal ( $a0 = -1.81$ ).

La valeur maximal de a1 :

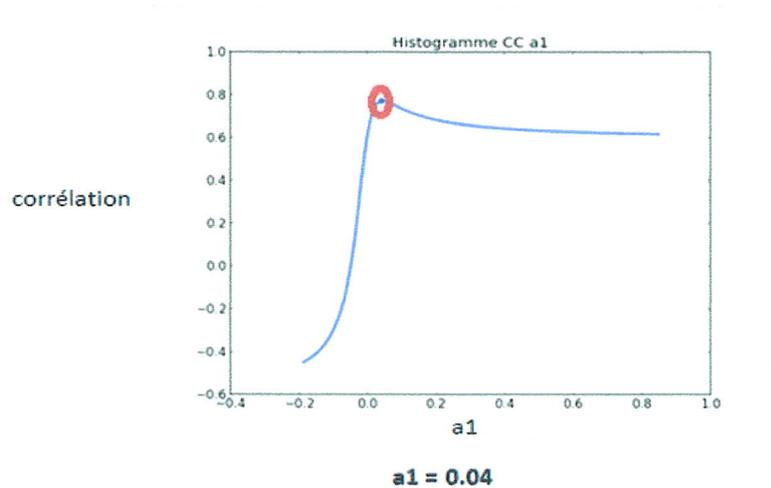


FIGURE 4.8 – Histogramme de a1

On constate que le coefficient de corrélation CC se change chaque fois qu'on change le a1 , donc on a sélectionné le CC la plus proche de 1 et on a pris le a1 qui lui correspond ( $a1 = 0.04$ ).

La valeur maximal de  $a_2$  :

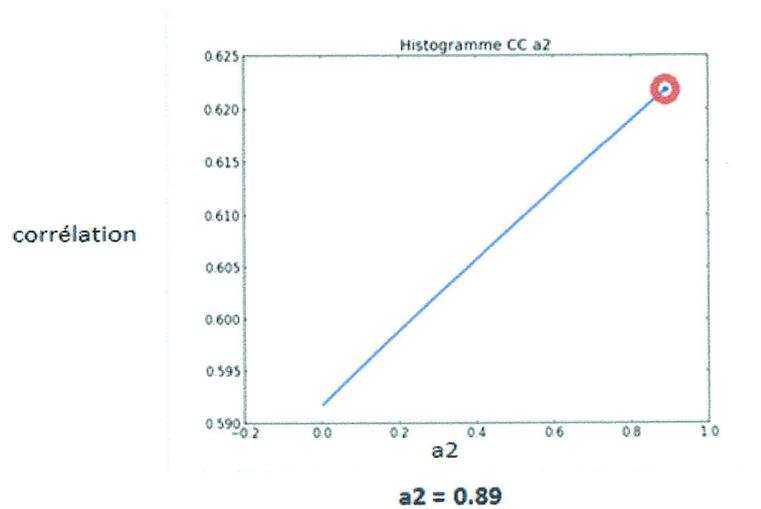


FIGURE 4.9 – Histogramme de  $a_2$

On observe que chaque fois qu'on change le  $a_2$  la corrélation s'augmente , alors on c'est prononcer sur la valeur maximal de coefficient de corrélation CC et on a gardé son  $a_2$  ( $a_2 = 0.89$ ).

La valeur maximal de  $a_3$  :

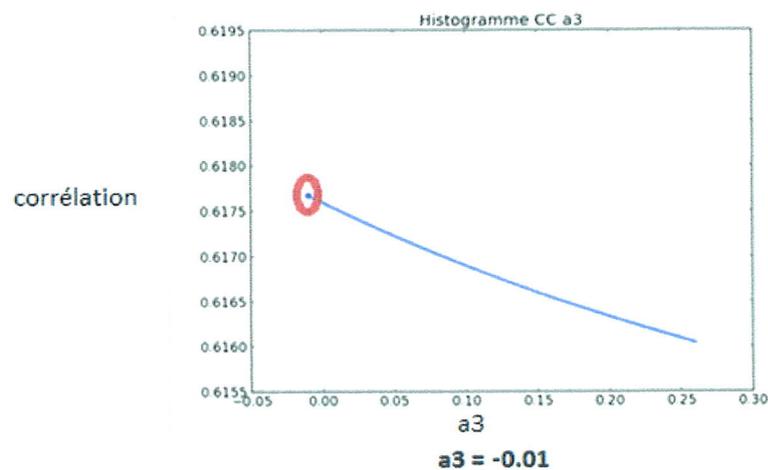


FIGURE 4.10 – Histogramme de  $a_3$

On remarque que chaque fois qu'on change les  $a_3$  la corrélation se diminue , alors on

est obligé de choisir le  $a_3$  correspond à la première valeur de CC qui est adéquat ( $a_3 = -0.01$ ).

Les résultats obtenue de notre méthode dans l'étape 3 est présenté dans la figure suivante :

Audiovisuel	MOS	SSIM Vidéo	SSIM Audio	AUDIOVISUEL
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr0.1_amrwb_m08	4.10344827586	0.999894847723	0.846005832008	-0.944712520972
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr0.5_amrwb_m08	2.81481481481	0.995831469564	0.686140213464	-0.953097183721
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr1_amrwb_m08	4.10344827586	0.996386011971	0.999993456712	-0.943180506001
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr1_amrwb_m08	1.73214285714	0.996781794095	0.61534224381	-0.954384132961
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr0_plr5_amrwb_m08	1.15	0.968293470451	0.64447295148	-0.978680282747
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr0.1_amrwb_m08	3.94827586207	1.0	0.703266602145	-0.948902001936
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr0.5_amrwb_m08	3.20833333333	0.996386011971	0.71904208441	-0.951619200719
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr1_amrwb_m08	3.8	0.999999223466	0.999999999104	-0.940000683377
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr1_amrwb_m08	2.0	0.989696173506	0.643361469925	-0.959803270632
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q23_nr999_plr5_amrwb_m08	1.25	0.971835689436	0.644705592229	-0.97554349175
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr0_plr0.1_amrwb_m08	3.18965517241	0.999464261278	0.999999999104	-0.940471450102
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr0_plr0.5_amrwb_m08	2.61111111111	0.997873429186	0.737140213464	-0.949762765812
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr0_plr1_amrwb_m08	3.48275862069	0.999999392751	0.999999999104	-0.940000534406
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr0_plr1_amrwb_m08	2.24137931034	0.998194272066	0.664992348508	-0.951645319453
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr0_plr5_amrwb_m08	1.25	0.976171208602	0.642456660221	-0.97178083488
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr999_plr0.1_amrwb_m08	2.39285714286	0.998284334458	0.719215291492	-0.949938144259
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr999_plr0.5_amrwb_m08	2.27777777778	0.998533214191	0.667752594177	-0.951263067044
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr999_plr1_amrwb_m08	3.5	1.0	0.999999999104	-0.940000000027
ntia_HeadShoulders-Female15_h264_720p_baseline_fps10_q27_nr999_plr1_amrwb_m08	2.39130434783	0.995067904297	0.639841183224	-0.955162772099

FIGURE 4.11 – Aperçus sur le fichier text des résultats obtenue

Le fichier raisonnable les résultats de la qualité objective pour les vidéos et les audio , ainsi que le résultat globale de la qualité audiovisuelle pour 160 vidéo audiovisuel de la base INRS . Le MOS est le résultat subjective de ces 160 vidéo qu'on obtenu avec la base (résultat de laboratoire INRS).

Nous avons les résultats subjectives (MOS) et les résultats objectives (notre méthode) où on a utilisé une mesure de performance (coefficient de corrélation) entre ces deux derniers pour voir la performance de notre méthode , dont on a obtenue un résultat standard ( $cc = 0.96$ ) pour faire une comparaison avec les résultats des méthodes qui sont déjà été exécuté par de nombreux laboratoire , on compare notre methode qui utilisé la fusion de 2 methodes avec les etudes empiriques qui sont fait sur différent labos KPN , Bellcore , ITS , NTT , ICRFE , BT , ICU sont utilisé la même methodes de base utilisé mais avec des parametre ai différentes voir le tableau ci dessous [52]

Laboratoire	a0	a1	a2	a3	Corrélation
KPN	1.12	0.007	0.24	0.088	0.98
	1.45	0	0	0.11	0.97
Bellcore	1.07	0	0	0.111	0.99
	1.295	0	0	0.107	0.99
ITS	-0.677	0.217	0.888	0	0.978
	1.514	0	0	0.121	0.927
	1.517	-0.0058	0.654	0.042	0.98
NTT	1.17	-0.144	0.186	0.154	0.96
	0.908	-0.192	0.258	0.193	0.96
ICRFE	-0.9222	0.5691	0.5064	0.1697	0.911
	-0.6313	0.2144	0.0124	0.1184	0.902
BT	1.15	0	0	0.17	0.85
	0.95	0	0.25	0.15	0.83
	4.26	0.59	0.49	0	0.97
	-3.34	0.85	0.76	-0.01	0.99
ICU	0	0.38	0.44	0.18	0.95
	0	0.43	0.32	0.26	0.95
	0	0.35	0.58	0.07	0.95
INRS	-1.81	0.04	0.89	-0.01	0.96

TABLE 4.2 – Comparaison entre les méthodes de la qualité audiovisuel et notre méthode

On a trouvé que notre méthode été classé comme un bon résultat comparer au d'autre qu'on a eu dans le tableau précédant des études empiriques des labos telle que la méthode KPN, sans oublier qu'il y a des mauvais et des moins mauvais résultats telle que ICRSE, ICU ....

## Conclusion

Dans ce chapitre, nous avons décrit, analysé et interprété les résultats de notre méthode. Une comparaison avec les résultats de nombreux laboratoires cités dans ce chapitre permet de classer notre méthode parmi les meilleures méthodes d'évaluation objectives de la qualité audiovisuel.

## Conclusion générale et perspectives

Dans le domaine informatique et plus exactement le domaine multimédia , l'audiovisuel a pris une grande place dans les temps moderne grâce à la fusion de la vidéo ainsi que l'audio.

Une poussée récente de services multimédias sur des réseaux filaires et sans fil a également déclenché une étude d'évaluation de la qualité perceptuelle. En particulier, il existe une forte demande pour des méthodes capables d'estimer et de quantifier la qualité du codage, de la transmission et du décodage (réception), des services, de l'expérience et de la satisfaction perçues par l'utilisateur final. Bien que l'évaluation de la qualité multimédia perceptuelle s'est avérée une tâche difficile, de nombreux efforts de recherche et de développement ont été consacrés à cette tâche et à ses applications, permettant ainsi des progrès significatifs dans ce domaine . Notre méthode d'évaluation de la qualité audiovisuelle est une méthode objective intrusive qui est la fusion de la qualité audio et la qualité vidéo avec des valeurs prédite pour obtenir une évaluation de qualité plus fiable et on clôture ce manuscrit par une comparaison entre cette méthode et les études empiriques des labos qui sont déjà été réalisé.

### **En perspectives de travail :**

- Dans le futur travail il est question de recherche sur les deux approches restant de l'évaluation objective de la qualité audiovisuel(l'approche de prévision de la qualité audiovisuelle et l'approche d'apprentissage automatique), on les compare avec notre méthode(l'approche de la fusion des deux modalités) pour choisir la meilleur pour l'évaluation de la qualité audiovisuel .

- Pour le moment notre résultat est acceptable parce que on les a appliqué sur une seule base de donnée INRS , et pour assurer la fiabilité de notre méthode , il faut encore s'élargir dans nos recherche . Parmi cet élargissement on va mettre l'application de notre méthode sur plusieurs base de donnée audiovisuel et là où on peut voir les bon et mauvais résultats et choisir la meilleur .

## Bibliographie

- [1] M.TAYACHI and Hamida MAKLOUF. *Techniques audiovisuelles*. Université Virtuelle de Tunis, (2006).
- [2] Revue des études humaines et sociales. -b/ lettres et langues.n° 19. p.37- 48. (Janvier 2018).
- [3] Dominique Brunet. *Métriques perceptuelles pour la compression d'images étude et comparaison des algorithmes JPEG et JPEG2000*, Mémoire présenté à la Faculté des études supérieures de l'Université Laval dans le cadre du programme de maîtrise en 114 mathématiques pour l'obtention du grade de Maître ès sciences (M.Sc.). faculté des sciences et de génie université Laval, Québec Septembre(2007).
- [4] T. N. Pappas and R. J. Safranek. "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Proc. (A. Bovik, ed.)*. Academic Press, (2000).
- [5] Le groupe Adobe Dynamic Media. Initiation à la video numérique. <http://www.adobe.com/fr/products/pdfs/dvprimer.pdf>, (juin 2000).
- [6] Mélanie Tremblay. Documents audiovisuels numériques. *Stage pratique Rabat (Maroc)*, (2012).
- [7] H.BENZENINE. Traitement du son. <http://dSPACE.univ-tlemcen.dz>.
- [8] T.H.Falk Z. Akhtar. Audio-visual multimedia quality assessment : A comprehensive survey. (2017).
- [9] D. Norman and S. Draper. User centered system design : New perspectives on human-computer interaction l. erlbaum associates. (1986).
- [10] S. R. Gulliver and G. Ghinea. The perceptual and attentive impact of delay and jitter in multimedia delivery. *IEEE Transactions on Broadcasting*, 53(2) :pp. 449-458, (juin 2007).
- [11] Requirements for an Objective Perceptual Multimedia Quality Model. *document ITU Rec.J.148*. ITU Telecommunication Standardization Sector, (2010).
- [12] S. Bech and N. Zacharov. *Perceptual Audio Evaluation Theory, Method and Application*. New York, NY, USA, (2006).

- [13] A. Chetouani. *Vers un système d'évaluation de la qualité d'image multicritères. Thèse de doctorat.* l'université paris 13, (2010).
- [14] H. Saadane C. Charrier, C. Larabi. Evaluation de la qualité des images. *Ecole d'hiver sur l'image numérique couleur*, (2005).
- [15] ITU-R Recommendation BT.500-10. *Methodology for the subjective assessment of the quality of television pictures.* March(2000).
- [16] J.Pedro López Velasco. Video quality assessment,video compression. *Université Politécnica de Madrid, Spain*(2012).
- [17] AHMED SEGHIR Zianou. Evaluation de la qualité d'image. *Thèse de doctorat en informatique*, 2012.
- [18] Recommandation UIT-R BS.1534-3. Méthode d'évaluation subjective du niveau de qualité intermédiaire des systèmes audio. *Service de radiodiffusion sonore*, (2015).
- [19] UIT T.Recommandation P.911. Méthodes d'évaluation subjective de la qualité audiovisuelle pour applications multimédias. <http://www.itu.int/pub/T-REC>, (1998).
- [20] UIT T.Recommandation P.911. Méthodes d'évaluation subjective de la qualité audiovisuelle pour applications multimédias : Corrigendum 1. <http://www.itu.int/pub/T-REC>, (1999).
- [21] Technologie Eyevinn. Évaluation de la qualité vidéo. <https://medium.com/@eyevinntechnology/video-quality-assessment-34abd35f96c0>, Avril (2018).
- [22] Margaret H Pinson Stephen D Voran Arthur A Webster, Coleen T Jones and Stephen Wolf. An objective video quality assessment system based on human perception. *In SPIE human vision, visual processing, and digital display IV*, 1913 :15–26, 1993.
- [23] Boudjadja Rima. Approche inter couches pour la transmission en temps réel de signaux multimédia sur les réseaux sans fil. 2018.
- [24] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices,. *IEEE Signal Processing Letters*, 17(5) :513–516, 2010.
- [25] I. Daubechies. Ten lectures on wavelets, society for industrial and applied 110 mathematics, 1992.20 i. daubechies, ten lectures on wavelets,. *Society for Industrial and Applied Mathematics*, 1992.
- [26] K. Ghazinour O. Alaql and C. C. Lu. Classification of image distortions based on features evaluation,. *IEEE International Symposium on Multimedia (ISM)*, 2015.
- [27] M. Meeker. Internet trends 2015-code conference. *Glokalde*, 2015.
- [28] Rajiv Soundararajan and Alan C. Bovik. Rred indices : Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Speech and signal Processing*, 21(2) :148–155, (2012).

- [29] Zainab Awan Noor Al Madeed and Somaya Al Madeed. Image quality assessment- a survey of recent approaches. pages 145–151, (2011).
- [30] Shibao Zheng Xiaolin Chen and Rui Zhang. Reduced reference image quality assessment based on image statistics in pixel domain. *In Advances on Digital Television and Wireless Multimedia Communications*, pages 148–155, (2012).
- [31] Abdul Rehman and Zhou Wang. Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing*, 21(8) :3378–3389, (2012).
- [32] ITU-T Recommendation P.862. "perceptual evaluation of speech quality (pesq), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and codecs". février (2001).
- [33] Method for Objective Measurements of Perceived Audio Quality. *document ITU-R Rec. BS.1387-1*, (2001).
- [34] S.Winkler et P. Mohandas. "l'évolution de la mesure de la qualité vidéo : du psnr aux métriques hybrides". *IEEE Transactions on Broadcasting*, 54(3) :660–668, (2008).
- [35] MM. Hannuksela M. Gabbouj et A.Perkis J. Vous, U. Reiter. "Évaluation de la qualité sur la base de la perception pour des services audiovisuels : une enquête". *Traitement du signal : Image Communication*, 25(7) :482–501, (2010).
- [36] D. Alais et D. Burr R. Arrighi. "synchronie perceptuelle des flux audiovisuels pour des séquences de mouvements naturels et artificiels". *Journal of Vision*, 6(3) :260–268, (2006).
- [37] T. Tebaldi O. Nemethova et M. Rupp M.Ries, R.Puglia. "estimation de la qualité audiovisuelle pour les services de streaming mobile". *Actes du colloque international sur les systèmes de communication sans fil, Sienne (Italie)*, (7 septembre 2005).
- [38] S.Jumisko-Pyykko. "j'aimerais voir les sous-titres et le visage ou tout au moins entendre la voix : Effets du rapport d'image et du rapport du débit audio-vidéo sur la perception de la qualité à la télévision mobile". *Outils et applications multimédias*, 36(1 2) :167–184, (2008).
- [39] JG.Beerends et FE.Caluwe. "l'influence de la qualité vidéo sur la qualité audio perçue et inversement". *Journal de la société d'ingénierie audio*, 47(5) :355–362, (1999).
- [40] C.Jones et DJ.Atkinson. "développement de modèles de qualité audiovisuelle fondés sur des opinions pour la vidéoconférence de bureau". *Actes de l'atelier international sur la qualité de service, Napa, Californie*, pages 196–203, (20 mai 1998).
- [41] L.Sun et E.Ifeachor M.Goudarzi. "estimation de la qualité audiovisuelle pour les appels vidéo dans les applications sans fil". *IEEE GLOBECOM*, (2010).

- [42] D.S. Hands. A basic multimedia quality model. *IEEE Trans. Multimedia*, vol. 6, no. 6, pp. 806-816, (Dec.2004).
- [43] B.Belmudez. Audiovisual quality assessment and prediction for videotelephony. *Berlin, Germany : Springer*, (2014).
- [44] M. M. Hannuksela M. Gabbouj J. You, U. Reiter and A. Perkis. "perceptual-based quality assessment for audio-visual services : A survey". *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 482-501, (2010).
- [45] H. E. Pashler. *The Psychology of Attention*. Cambridge, MA, USA :MIT Press, (1998).
- [46] J. Beerends and F. De Caluwe. "the influence of video quality on perceived audio quality and vice versa". *J.Audio Eng. Soc*, vol. 47, no. 5, pp. 355-362, (1999).
- [47] H. Martinez and M. Farias. "full-reference audio-visual video quality metric". *J. Electron. Imag.*, vol. 23, no. 6, p. 061108, (2014).
- [48] B. Lewcio A. Raake B. Benjamin, S. Moeller and A. Mehmood. "audio and video channel impact on perceived audio-visual quality in different interactive contexts ". *in Proc. IEEE Int. Workshop Multimedia Signal Process.* pp. 1-5, (Oct. 2009).
- [49] UIT-T RECOMMENDATION P.1201. " parametric non-intrusive assessment of audio-visual media streaming quality". (2012).
- [50] UIT-T RECOMMENDATION G.1070. " opinion model for video-telephony applications". (2012).
- [51] UIT-T RECOMMENDATION G.1071. "opinion model for network planning of video and audio streaming applications". (2015).
- [52] M. M. Hannuksela M. Gabbouj J. You, U. Reiter and A. Perkis. "perceptual-based quality assessment for audiovisual services : A survey". *Signal Process, Image Commun*, 25(7) :482-501, (2010).
- [53] Edip Demirbilek. Inrs audiovisual quality dataset. <http://github.com/dipdemirbilek/INRSAudiovisualQualityDataset>, (2016).
- [54] Jean-Charles Grégoire Edip Demirbilek. "the inrs audiovisual quality dataset". *ACM Multimedia Conference*, (2016).
- [55] Jean-Charles Grégoire Edip Demirbilek. "machine learning based parametric audiovisual quality prediction models for realtime communications". *ACM TOMM (Revised)*, (2016).
- [56] Jean-Charles Grégoire Edip Demirbilek. "machine learning based bitstream audiovisual quality prediction models for realtime communications". *IEEE International Conference on Multimedia and Expo*, (2016).

- 
- [57] Jesus Alcober Alberto J. Gonzalez. "context-aware multimedia service composition using quality assessment". *IEEE International Conference on Multimedia and Expo*, (2011).
- [58] D. S. Hands. "a basic multimedia quality model". *IEEE Trans. Multimedia*, 6(6) :806–816, (2004).

## RÉSUMÉ

Aujourd'hui, des systèmes de restitution audiovisuelle de haute qualité (format haute définition -HD- ou 3D -images et sons-) s'étendent à l'ensemble des terminaux (mobiles, ordinateur, télévision) et services (communications interpersonnelles, télévisuelles - TV-, etc.) disponibles. Face à cette constante évolution et dans un contexte fortement concurrentiel, un des principaux enjeux pour les acteurs de l'offre de services audiovisuels (AV) est de garantir une qualité d'expérience (QoE -Quality of Experience-) optimale à l'utilisateur. Cela est particulièrement vrai pour les nouvelles technologies à forte valeur ajoutée, telles que l'audio et la vidéo HD ou 3D. Ainsi, une attention toute particulière est portée à l'évaluation de la qualité d'expérience à travers le développement d'outils et la mise en place de méthodes d'évaluation.

La qualité audiovisuelle (QAV) ne résulte pas de la simple addition des qualités audio (QA) et vidéo (QV) mais de l'interaction des niveaux de qualité de chaque modalité. Nous avons étudié la qualité de ces données en précisant ses métriques d'évaluation existant et l'influence des QA et QV sur la qualité des audiovisuelles.

**Mots clés : Audiovisuel , Qualité , QAV , QV , QA , QoE .**

## ABSTRACT

Today , high-quality audiovisual reproduction systems (high-definition format -HD- or 3D -images and sounds-) extend to all terminals (mobile, computer, television) and services (interpersonal, televisual communications) - TV- etc.) available. Faced with this constant evolution and in a highly competitive context, one of the main challenges for audiovisual service providers (AV) is to guarantee an optimal quality of experience (QoE -Quality of Experience-) to the user . This is especially true for new high value-added technologies, such as audio and HD or 3D video. Thus, special attention is paid to evaluating the quality of experience through the development of tools and the implementation of evaluation methods.

Audiovisual quality (QAV) does not result from the simple addition of audio (QA) and video (QV) qualities, but from the interaction of the quality levels of each modality. We studied the quality of these data by specifying its existing evaluation metrics and the influence of QA and QV on the quality of audio-visuals.

**Key words : Audiovisual , Quality , QAV , QV , QA , QoE .**

