Republique Algerienne Democratiqueet Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientique
UNIVERSITE MOHAMMED SEDDIK BENYAHIA
JIJEL
FACULTE DE SCIENCES EXACTES ET D'INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

MEMOIRE DE MASTER

Présenté pour l'obtention du diplôme de :

MASTER

EN **INFORMATIQUE**

Option : INFORMATIQUE LÉGALE ET MULTIMEDIA

Thème

La méthode de combinaison pour évaluer la qualité audio-visuelle

Présenté par :
M. BARAMA
Fayssal

Encadrée par :
Dr. BOUDJRIDA
fatima

Promotion : 2021

REMERCIEMENTS

Nous tenons à remercier en premier Dieu qui nous a donné la force, la volonté et le courage pour réaliser ce modeste travail.

Nous remercions en particulier notre encadrant M^{me} BOUDJRIDA fatima pour la confiance qu'il nous a accordé en acceptant de diriger notre travail de mastère. C'est grâce à son aide, à ses précieux conseils, qu'elle n'a cessé de nous prodiguer, que ce mémoire a pu voir le jour . Nous vous serons toujours reconnaissants.

Nos remerciements s'adressent aussi aux membres du jury pour l'honneur qu'ils nous ont fait en acceptant de juger et d'examiner notre travail, ainsi que tous les enseignants et tous les enseignantes de l'université de Jijel.

Et en fin, un très grand merci à tous ceux qui, de près ou de loin ont contribué la realisation de ce mémoire .

RÉSUMÉ

La mesure de la qualité perçue des signaux audiovisuels par l'utilisateur final est devenue un paramètre important pour de nombreus réseaux et applications multimédias. Elle joue un rôle crucial dans l'élaboration du traitement, la compression, la transmission et les systèmes audiovisuels, ainsi que leur mise en œuvre, leur optimisation et leur test. Les fournisseurs de services mettent en œuvre différentes solutions de qualité de service pour offrir la meilleure qualité d'expérience aux utilisateurs. 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. Dans ce travaille,nous avons proposé une évaluation de la qualité audiovisuelle en combinant entre la qualité audio et la qualité vidéo à l'aide des trois fonction : (lineare, minkowski, power). En général, nous avons fait tester ces trois fonctions en ce on se basant sur des MOS(mean opnion score) subjectives et des MOS prédictives et les résultats ont été prometteus en terme de l'efficacité. Néanmoins, il reste divers problèmes à résoudre pour mieux comprendre les nombreuses complexités de la perception humaine des qualités individuelles et multimodales. Nous sommes encore loin d'une méthode d'évaluation multimodale fiable de la qualité, de l'expérience et de la perception, qui nécessitera des efforts de recherche interdisciplinaires dans différents domaines, tels que la vision humaine, la physiologie, etc. .

Mots clés : Evaluation de la qualité audiovisuel , qualité de service ,qualité d'experience , multimédias, MOS subjectives, MOS prédictives.

ABSTRACT

Measuring the perceived quality of audiovisual signals by the end user has become an important parameter for many networks and multimedia applications. It plays a crucial role in the development of audio-visual processing, compression, transmission and systems, as well as their implementation, optimization and testing. Service providers are implementing various Quality of Service (QoS) solutions to provide the best Quality of Experience (QoE) to users. Many research and development efforts have been devoted to this task and its applications, allowing significant progress in this area. In this work we proposed an evaluation of the audio-visual quality by combining audio and video quality using the three functions (lineare, minkowski, power). In general, we have tested these three functions based on subjective MOS (mean opinion score) and predictive MOS (mean opinion score) and the results have been promising. Nevertheless, various problems remain to be solved to better understand the many complexities of human perception of individual and multimodal qualities. We are still far from a reliable multimodal assessment method for quality, experience, and perception, which will require interdisciplinary research efforts in different fields, such as human vision, physiology, etc.

Keywords : audio-visual quality assessment, (QoE), (QoS), multimedia, subjec-

tive MOS, predictive MOS

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

LISTES DES ABRÉVIATIONS

QAV Qualité audiovisuelle

QA Qualité audio

QV Qualité vidéo

IPTV Internet Protocol Télévision

RV La réalité virtuelle

DSIS La méthode de double stimulus

DSCQS Double stimulus continuos méthode quality scale

SSCQE Single stimulus continuous Quality Evaluation

MOS Mean opinion score

DMOS Difference mean opinion score

AR Arbre de recherche

DL Deep learning

FPS Frame par seconde

FR Référence complète

RR Référence réduit

NR Sans référence

PESQ L'évaluation perceptive de Ia qualité vocale

PEAQ L'évaluation perceptuelle de la qualité audio

PSNR Le rapport signal / bruit crête

SSIM La mesure de similarité structurelle moyenne

JPEG Joint Photographic Experts Group

SVH System visuelle humaine

VQM Vidéo quality metrics

HASQI Hearing-Aid Speech Quality Index

SRMR speech-to-réverbération-modulation Energy ratio

MGM modèles de mélange gaussien

PLP prédiction perceptuelle-linéaire

VQ quantification vectorielle

PCC The Pearson liner correlation coefficient

OR The outlier ratio

RMSE Root mean square error

SROCC The Spearman rank order correlation coefficient

INTRODUCTION GÉNÉRALE

Les services multimédias connaissent une croissance considérable dans la popularité des services multimédias s'est accrue récemment en raison de l'évolution des systèmes de communication numériques. Deux modalités principales de médias, à savoir les signaux audio et vidéo, constituent le contenu de base dans la plupart des systèmes numériques. La qualité des signaux audio-visuels peut se dégrader pendant la compression avec pertes et la transmission par des réseaux de communication sujets aux erreurs. Par conséquent, mesurer avec précision la qualité des signaux audiovisuels déformés joue un rôle important dans les applications numériques. par exemple lors de l'évaluation des performances des codecs et des réseaux. Pour aider à améliorer les capacités de codage ou d'ajuster les paramètres du réseau en fonction d'une stratégie visant à maximiser la qualité perçue par l'utilisateur final.

L'évaluation subjective de la qualité audiovisuelle est considérée comme la méthode la plus précise pour refléter la qualité perçue par l'utilisateur final.[8]

Considérée comme la méthode la plus précise reflétant la perception humaine. Cependant, elle prend beaucoup de temps et ne peut être réalisée en temps réel. C'est pourquoi l'Union internationale des télécommunications (UIT) a publié des exigences pour un modèle perceptuel objectif de qualité multimédia.

Actuellement, la plupart des études concernant la compréhension de la perception

humaine de la qualité des systèmes multimédia se sont concentrés sur les modalités individuelles, c'est-à-dire l'audio et la vidéo séparément.

Dans ce mémoire nous nous intéressons à l'évaluation de la qualité des données audio-visuelle en combinant les sorties de mesures objectives individuelles de qualité audio et vidéo avec une fonction linéaire, Minkowski ou de puissance.

CHAPITRE 1

LA QUALITÉ AUDIOVISUELLE

1.1 Introduction

a

DANS un contexte extrêmement concurrentiel, la qualité d'expérience de l'utilisateur (QoE : Quality of Experience), est une des préoccupations principales des acteurs du domaine de l'offre de services audiovisuels (télévisuels, visioconférences, etc.). Actuellement l'évaluation de la QoE se réalise généralement à travers l'évaluation de la qualité, telle que perçue par les utilisateurs, des signaux audio et/ou vidéo restitués. Les méthodes d'évaluation utilisées sont recommandées par l'Union Internationale des Télécommunications(UIT). Ces approches reposent sur des mesures subjectives dont l'interprétation et la validité sont limitées par un certain nombre de biais. Elles ne permettent pas non plus de rendre compte fidèlement de l'influence de la qualité audiovisuelle sur la qualité d'expérience de l'utilisateur. Par exemple, ces méthodes n'apportent pas d'informations sur le coût pour l'utilisa-

teur, du point de vue de la fatigue ou de l'effort mental, induit par des dégradations du signal et pouvant à terme conduire à un rejet du système ou de la technologie de restitution. Le coût d'utilisateur peut être mesuré à partir d'indices de l'activité physiologique et oculaire de l'individu, ce type de mesures présente l'avantage de ne pas être soumis aux biais des mesures subjectives, capables de diminuer la fiabilité des réponses recueillies. Wilson et Sasse [1] ont montré que des fluctuations importantes de la qualité audio ou vidéo peuvent ne pas être consciemment perçues par les participants et pour autant être reflétées par l'activité physiologique.
Dans ce chapitre nous allons aborder la définition de l'audiovisuelle et quelques domaines d'application ensuite on passent à la qualité et ces facteurs d'influence et ces dégradation enfin, nous avons parlé de l'évaluation.

1.2 Définition

Le terme audiovisuel peut se référer à tout travail qui utilise, à la fois, du son et de l'image. Donc l'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 animée et/ou des images fixes.[2]

1.3 Domaines d'utilisation de l'audiovisuelle

Parmi les différents domaines d'utilisation de l'audiovisuelle, Nous mentionnons les suivants :

1.3.1 L'enseignement et l'apprentissage

L'emploi des supports audiovisuels à plusieurs avantages pour L'enseignementet l'apprentissage. Nous pouvons en citer les suivants :

- Le professeur possède la liberté et la responsabilité pour organiser le contenu audiovisuel.
- L'enseignant à la possibilité de présenter ce contenu avec des moyens didactiques et pédagogique appropriés.
- L'audiovisuel facilite la mémorisation.
- Permettre l'apprentissage à distance et cela a été très essentiel dans la période du Covid-19.

1.3.2 La réalité virtuelle

C'est une interface homme-machine avancée qui simule un environnement réaliste. Les participants peuvent se déplacer dans le monde virtuel. Ils peuvent le voir sous différents angles, l'atteindre, le saisir et le remodeler. Le cyberespace est considéré comme l'ultime environnement de réalité virtuelle, les derniers développements de la réalité virtuelle examines les applications dans les domaines de l'ingénierie et de la médecine.[3]

1.3.3 Les jeux vidéo (gaming)

En quelques années, le jeu vidéo est devenu l'un des loisirs les plus populaires au monde. Que ce soit en nombre de joueurs ou en chiffre d'affaires. Le secteur est l'un des plus dynamiques de l'économie. Ainsi, en 2019, le chiffre d'affaires du jeu vidéo s'est élevé à 4,81 milliards d'euros, et les gamers réclame toujours plus on

ce qui concerne le côté graphique (vidéo)où bien les performances de la qualité de son(audio).[4]

1.3.4 Vidéo conférence

En 1968 la vidéo conférence à été introduite pour la première fois et présenter comme une solution commercial à l'exposition universelle de New York. La technologie introduite s'appelait le Picture phone d'AT&T(American Telephone & Telegraph) les participant on peut s'asseoir et communiquer par vidéo ave la personne de l'autre côté pendant 10 minutes à la fois pour faire l'expérience du premier appareil de visiophone conçue pour les masses. Malheureusement cette machine particuliere été ridiculement chère maladroit et difficile à installer.[5], Dans nos derniers jours les applications sont gratuite et facile à utiliser comme Zoom et Google Meet. [5]

1.4 La Qualité

La notion de qualité est un concept abstrait et envisagé comme une construction de l'esprit, qui est facile à comprendre mais difficile à définir. Dans le domaine multimédia, la qualité est généralement utilisée avec un objectif d'ingénierie à l'esprit en raison du fait que la qualité est un critère clé pour évaluer les systèmes, les services ou les applications pendant les phases de conception et d'exploitation [6]. Bien que, selon le livre blanc QUALINET [7], « la qualité soit le résultat du processus de comparaison et de jugement d'une personne, qui comprend la perception, la réflexion sur la perception et description du résultat ». Contrairement aux définitions/concepts dans lequel la qualité est considérée comme « qualitas » (c'est-à-dire un ensemble de caractéristiques), QUALINET considère la qualité en termes de l'excellence , du de-

gré d'accomplissement des besoins et d'un « événement de qualité », où l'événement est un événement observable et déterminé dans l'espace (c.-à-d. là où il se produit), le temps (c.-à-d. lorsqu'il se produit) et le caractère (c.-à-d. ce qui peut être observé)[7] Récemment, la recherche et l'industrie se sont englobant l'utilisateur final comme le facteur le plus important dans l'évaluation de la qualité multimédia pour atteindre une bonne qualité tels que la qualité de l'expérience (QoE : Quality of Experience) ou la qualité de perception (QoP : Quality of Perception) plutôt que seulement la qualité du service (QoS : Ouality of Service).[8]

1.4.1 La qualité de service (QoS)

La QoS est souvent utiliser pour exprimer le niveau de performance des applications multimédias et des réseaux.cette dernière est généralement utilisée dans la littérature en fonction des facteurs de performance physiques et mesurables des réseaux, y compris les plates-formes de livraison, est « une collection de technologies de réseautage et d'outils de mesure qui permettent au réseau de garantir la réalisation de résultats prévisibles [6] ». Le terme QoS est généralement utilisé avec deux significations différentes :

— Premièrement, il fait référence aux concepts et aux mesures des performances du réseau par exemple, nervosité, retard.

— Deuxièmement, il s'agit de mécanismes tels que les services intégrés.

Plusieurs caractéristiques, telles que la performance, la réactivité, la disponibilité, l'adaptabilité, la fiabilité, la sécurité et les aspects d'application sont impliquées pour former le QoS. En raison de l'hétérogénéité des applications, QoS a été expliqué de façon diverse dans des publications indépendantes. Compte tenu de l'architecture multimédia de bout en bout, QoS peut être diviser en trois couches : utilisateur,

application et ressource.[8]

1.4.2 La qualité de l'expérience (QoE)

La satisfaction et la perception des utilisateurs sont façonnées par divers autres aspects qui peuvent/ne peuvent pas nécessairement être réglementés par la performance des composants de service spécifiques. Donc récemment, le terme Qualité d'expérience (QoE) a été introduit pour décrire comment un utilisateur perçoit la facilité d'utilisation, l'acceptabilité et la satisfaction du service. QoE s'en va au-delà des paramètres conventionnels d'intégrité QoS de bout en bout pour pouvrir une multitude d'aspects différents pour améliorer la qualité expérimentée par l'utilisateur. Nommément QoE est le QoS perceptuel du point de vue des utilisateurs.[8]

Voici la (FIGURE 1.1) qui schématise de maniére simplifiée la relation entre QoS et QoE :

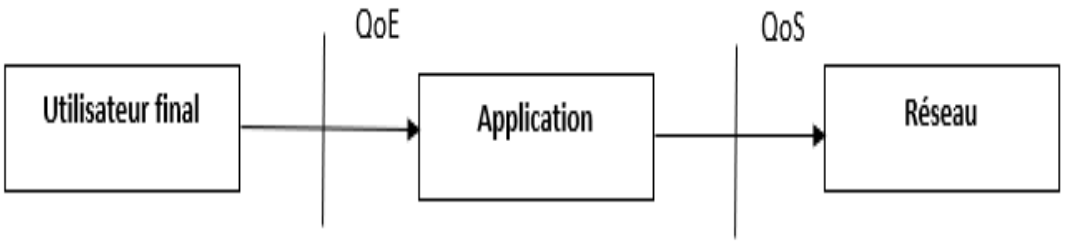

FIGURE 1.1 – La relation entre QoS et QoE

Il est primordial d'obtenir un QoE quantifie en traduisant performance du système ainsi que la perception des utilisateurs dans la forme de valeurs statistiques et interprétables. Le quantifie QoE peut être obtenu en utilisant soit des « mesures directes QoE » (c.-à-d. l'évaluation effectué par des sujets réels ; également appelée subjectives QoE) ou « mesures indirectes du QoE » (c.-à-d.l'enregistrement du comportement de l'utilisateur et de le relier avec QoE perçu(Également appelé QoE objectif). Dans cette dernière catégorie, l'utilisation des mesures physiologique a été récemment enquêté dans plusieurs études[8].

1.4.3 La qualité de perception (QoP)

QoS décrit la qualité technique du système, mais néglige la délié et l'aspect utilitaire des utilisateurs. Ainsi, pour remédier à cette limitation, Ghinea et Thomas [9] ont introduit la notion de qualité de perception (QoP) et l'ont définie comme « QoP est un terme qui englobe non seulement la satisfaction d'un utilisateur à l'égard de la qualité des présentations multimédias, mais aussi sa capacité d'analyser, de synthétiser et d'assimiler le contenu informationnel des écrans multimédias ». L'évaluation de la qualité multimédia en utilisant uniquement des facteurs subjectifs ou objectifs est insuffisant en raison de la nature multidimensionnelle du multimédia par conséquent QoP combine à la fois l'évaluation subjective basée sur la partie la plus large de la définition, c'est-à-dire la satisfaction de l'utilisateur à l'égard de la qualité des présentations multimédias (dénotées QoP-S), et l'objectif basé sur la deuxième partie de la définition, c'est-à-dire la capacité de l'utilisateur à analyser, synthétiser et assimiler le contenu informationnel du multimédia (indiqué par QoPIA : Quality of Perception Informational Analysis). QoP-S est composé de deux composants, c'est-à-dire QoPLOE (le niveau de plaisir de l'utilisateur tout en découvrant le contenu multi-

média) et QoPLOQ (jugement de l'utilisateur concernant l'objectif niveau de qualité attribué au contenu multimédia expérimenté). D'un point de vue spectique, QoPIA s'exprime habituellement en pourcentage pour refléter le niveau d'information d'un utilisateur assimilé à partir de contenu multimédia expérimenté.[9]

1.5 Les facteurs de dégradation de la qualité audiovisuelle :

La qualité audiovisuel est souvent dégradée à cause de certains facteurs dans les applications de streaming par exemple, les signaux vidéo et audio passent généralement par un pipeline de traitement composé de plusieurs étapes représentatives, dont la génération du contenu, le traitement, le codage côté serveur, la diffusion en continu sur le réseau et enfin le décodage et la présentation aux consommateurs du côté de l'utilisateur final. Diverses dégradations peuvent être introduites l'un ou l'autre des signaux vidéo et audio, ou les deux, qui dégrade la qualité d'expérience de l'utilisateur final. Les consommateurs modernes sont de plus en plus avertis en matière de technologie audio et vidéo (A/V), et attendent une qualité d'expérience élevée lorsqu'ils regardent et l'écoute en utilisant des systèmes A/V de plus en plus haute résolution et haute-fidélité, que ce soit sur des appareils mobiles ou sur leurs ordinateurs, Il y a donc une forte impulsion pour développer et déployer des systèmes de qualité audio et vidéo efficaces et précis ainsi que de déployer des modèles d'évaluation de la qualité audio et vidéo (QA/QV) efficaces et précis qui peuvent être utilisés pour surveiller et contrôler la qualité d'expérience de l'utilisateur final.[8]

1.5.1 Les facteurs d'influences

Cette section décrit les facteurs qui peuvent influencer la qualité d'échantillons audio et/ou visuels. En outre, les caractéristiques audio et visuelles qui sont couramment utilisées dans l'évaluation objective de la qualité sont étudiées, pour améliorer les algorithmes d'évaluation, il est apprécié de comprendre les facteurs complexes et fortement interdépendants qui ont un impact sur les comportements d'interaction des utilisateurs ainsi que la qualité perçue. Quelques facteurs sont inévitables, tandis que certains sont dus aux limites inhérentes au signal multimédia lui-même. Ces facteurs peuvent être regroupées en trois catégories : humaines, technologiques et les facteurs contextuels d'influences.[8]

1. **Facteurs d'influences humains :** Englobent les caractéristiques variables ou invariantes de l'utilisateur humain susceptibles d'avoir une incidence sur le jugement de la qualité, notamment la constitution physique / mentale / l'état émotionnel, les antécédents démographiques et socio-économiques. Ces attributs sont soit statiques (genre, âge) ou dynamiques (états mentaux, motivation). Les facteurs utilisateurs peuvent participer aux processus de qualité sensorielle et / ou cognitive. Le processus de qualité sensorielle précoce (c.-à-d. De bas niveau) est affecté par les états physique, émotionnel et mental de l'utilisateur, par exemple son acuité auditive, son humeur et son attention. Le processus de qualité cognitif (c'est-à-dire de niveau supérieur / descendant) se rapporte à l'interprétation des stimuli en fonction des connaissances et des antécédents de l'utilisateur, y compris les besoins, la motivation, les préférences, etc.[8]
2. **Facteurs d'influence technologiques :** Englobent l'agent (un partenaire d'interaction) et les facteurs fonctionnels du système. Les exemples de facteurs

d'agent sont des attributs techniques (par exemple, la reconnaissance vocale). Les exemples de facteurs fonctionnels sont les capacités fonctionnelles (par exemple, le nombre de tâches) et les caractéristiques du domaine (par exemple, le système de divertissement). Les facteurs système peuvent également être divisés en quatre classes en fonction du réseau (associé à la transmission de données sur un réseau, par exemple, la bande passante), lié au périphérique (associé au système / périphérique de communication, par exemple, un smartphone haute résolution) , liée au support (c'est-à-dire associée à la configuration du support, par exemple, cadences de prise de vue) et au contenu (associée, par exemple, à la quantité d'informations sur le support, par exemple contenu vocal / parlé / musical).[8]

3. **Facteurs d'influence contextuels :** Englobe l'environnement physique (par exemple, le bureau) et les facteurs de service (par exemple, les attributs système non physiques, par exemple les restrictions d'accès au système). Les facteurs de contexte peuvent également être décomposés en contexte physique (par exemple, emplacement et caractéristiques de l'espace, par exemple lieu paisible / bruyant), contexte temporel (par exemple aspect temporel de l'expérience, par exemple mois juin ou printemps), contexte social (par exemple interrelations entre utilisateurs, par exemple, dépendances hiérarchiques (patron et employé), contexte économique (par exemple, perspective professionnelle, par exemple, coût par utilisation), contexte de tâche (expérience de l'utilisateur pour la qualité perçue, par exemple, effet du multitâche tout en évaluant la qualité) et le contexte technique et informationnel (à savoir la relation entre les systèmes et les périphériques concernés ou facultatifs, par exemple l'interconnexion des périphériques via Bluetooth).

1.5.2 Dégradations des signaux audiovisuels

Afin de mieux comprendre l'évaluation de la qualité audiovisuelle il peut être utile d'examiner de près les différents artefacts qui se manifestent couramment dans les signaux audio et vidéo. La dégradations audio/visuelles se manifestent par les propriétés du dispositif de capture du signal, du mécanisme de codage, de décodage, de compression ou de transmission, ou du dispositif final utilisé.
On a deux types de dégradation celle quelle sont auditif (la réverbération, le bruit auditive) et on à celle quelle sont visuelle (Flou, effets de bloc, bruit, ringing-effect).[8]
– **Le bruit auditif :** On peut définir le bruit auditif comme étant une dégradation dans La qualité audio, cela produit des sons agaçants. Il y a plusieurs solutions comme les casques circumaural fermés qui donnent à l'utilisateur l'impression que le son provient de l'intérieur de sa tête. Cet effet apparaît également avec les casques ouverts, qui présentent une réduction du bruit plus faible.[13]
– **La réverbération :** La réverbération est la persistance du son dans un lieu après l'interruption de la source sonore. La réverbération est le mélange d'une quantité de réflexions directes et indirectes donnant un son confus qui décroît progressivement.[14]
– **Le flou :** C'est un effet esthétique qui donne à voir un contour imprécis. Ce type d'erreur qui apparaît principalement dans la compression JPEG et JPEG2000 d'où la perte de netteté de l'image. Cela se caractérise par une image plus floue, dont les bords des objets sont plus diffus.[15]
– **Effets de bloc :** La principale source d'erreur lors de la compression JPEG est ce qu'on appelle l'effet de blocs. Visuellement, cette distorsion se manifeste généralement au niveau des frontières entre blocs et apparaît comme des contours verticaux et horizontaux dont la visibilité dépend fortement de la distribution spatiale du signal image. En effet, tous les blocs sont encodés indépendamment les uns des autres.

Il peut donc arriver qu'à la frontière entre deux blocs, il y ait une discontinuité facilement perceptible par l'œil humain.[15]

– **Effet d'oscillations parasites :** Cette dégradation est due en général à l'étape de quantification ou de décimation des coefficients hautes fréquences. Elle se manifeste sous forme d'oscillations au voisinage des régions à fort contraste et est souvent définie comme un bruit autour de ces régions. Ce sont les ondelettes dont le support croise le bord d'un objet qui créent ce type d'artefact.[15]

– **Bruit :** On peut définir le bruit comme étant une dégradation dans l'image, provoquée par une perturbation externe. Généralement, on peut savoir les types d'erreurs à attendre, et donc le type de bruit sur l'image, d'où nous pouvons choisir la méthode la plus adaptée pour réduire les effets. On l'appelle aussi le bruit impulsionnel, le bruit de grenaille, ou le bruit binaire. Cette dégradation peut être causée par de fortes perturbations soudaines dans le signal d'image. Son apparence est éparpillée au hasard en pixels blancs ou noirs (ou les deux) sur l'image, Par contre, ce bruit est obtenu en ajoutant n pixels blancs et n pixels noirs aléatoirement dans une image. On le caractérise souvent par le pourcentage de pixels remplacés.[15]

1.5.3 Facteur de temps

Une autre caractéristique importante et bien connue de la qualité audiovisuelle est la synchronisation entre la qualité audio et vidéo, cette synchronisation des canaux audio et vidéo affectent plus considérablement l'évaluation objective que subjectivement. La désynchronisation spatiale ou temporelle est la dégradation la plus importante du contenu multimédia audiovisuelle. Dans les systèmes de reproduction de contenus audiovisuels, la lecture de stimuli auditifs et visuels synchronisés est considérée comme obligatoire. Il est intéressant de constater que les seuils de détec-

tion de la non-synchronisation ne sont pas temporellement symétriques.
Hollier et Rimel and all. [16] ont réalisé un certain nombre d'expériences axées sur les systèmes de communication audiovisuels pour examiner cette asymétrie temporelle avec différents types de stimuli. Ils ont comparé une scène audiovisuelle de tête parlante avec une scène de stylo rebondissant et un stimulus audiovisuel. Dans lequel une hache frappe un objet une seule fois. Ils ont conclu que la tendance générale de l'asymétrie de la détection des erreurs est apparente pour tous les types de stimulus. En outre, le caractère distinct de la hache stimulus entraîne une plus grande probabilité de détection que pour le stimulus du stylo. Pour le stimulus de la tête parlante, le taux de détection d'erreurs est cohérent avec les autres stimuli lorsque l'audio est en retard sur la vidéo, mais il est plus élevé que celui de la hache ou du stylo lorsque l'audio est en avance sur la vidéo.
Apparemment, les sujets testés ont comparé les stimuli artificiels présentés dans le laboratoire avec les expériences de la vie réelle. Dans la vie réelle, en raison de la nature physique des différentes vitesses de propagation du son et de la lumière, l'audio ne peut jamais devancer le percept visuel.
Ces conclusions sur la détection des erreurs de synchronisation de détection des erreurs de synchronisation se reflètent également dans les seuils de synchronisation recommandés dans la norme UIT J.100 [17], qui sont de 20 ms pour l'avance audio et de 40 ms pour le retard audio. La recommandation suggère ces valeurs fixes pour tous les types de contenu télévisuel et vise à garantir que les erreurs de synchronisation restent imperceptibles pour toutes les variétés possibles de contenu. Ce seuil relativement bas signifie que le système perceptuel humain est généralement assez sensible aux erreurs de synchronisation.

1.6 Evaluation de la qualité audiovisuelle (QAV)

Il existe essentiellement deux catégories de méthodes d'évaluation de la qualité (QA), à savoir les méthodes subjectives faisant appel à des observateurs humains pour évaluer la qualité des contenus multimédia et les méthodes objectives qui calculent la qualité automatiquement à l'aide de modèles mathématiques.[8]

1.6.1 Evaluation de la qualité subjective

Afin de mesurer de manière fiable la qualité perçue par les systèmes humains auditifs et/ou visuels, des tests subjectifs sont effectués où des groupes d'observateurs humains formés ou naïfs fournissent la qualité[19].
Cette procédure d'évaluation est connue sous le nom de l'évaluation subjective de la qualité qui vise à quantifier la plage des avis que les utilisateurs expriment lorsqu'ils voient/entendent le contenu numérique.
L'évaluation subjective de la qualité est effectuée en général dans un environnement bien contrôlé à l'aide des recommandations normalisées par exemple, Union internationale des télécommunications.
Les recommandations communément utilisées pour les testes de la qualité audiovisuelle sont UIT-T P.913 , UIT-T P.920 et UIT-T P.1401 [20].
– **ITU-T P.913 :** Méthodes d'évaluation subjective de la qualité vidéo, audio et audiovisuelle de la vidéo sur Internet et la distribution de qualité télévision dans n'importe quel environnement.[21]
– **ITU-T P.920 :** Méthodes d'essai interactives pour communications audiovisuelles. Cette recommandation est destinée à définir des méthodes interactives d'évaluation permettant de quantifier l'influence des procédés de codage et des temps de trans-

mission sur des communications audiovisuelles point à point ou multipoint. Cette méthode est fondée sur des essais d'opinion en conversation.[**?**]

– **ITU-T P.1401 :** cette recommendation présente un cadre pour l'évaluation statistique des algorithmes de prédiction objective de la qualité indépendamment du type de média évalué.

D'une manière générale, les études subjectives sur la qualité d'expérience peuvent être étiquetés comme techniques qualitatives ou quantitatives.[23]

— **Techniques qualitatives :** Elles saisissent les perceptions, les sentiments et les opinions des êtres humains par le biais de comportements verbaux, par exemple les commentaires sur les blogs et les sites.

— **Techniques quantitatives :** Elles permettent de saisir les perceptions et les Sentiments humains et les intentions à travers les chiffres et les statistiques.

1.6.2 Protocoles d'évaluation subjective de la qualité audiovisuelle :

Il y a essentiellement trois grandes par familles communes d'évaluation subjective définies par l'UIT, échelle continue de la qualité sur double stimulus DSCQS(Double Stimulus Continuos Quality Scale), échelle de dégradation sur double stimulus DSIS(Double Stimulus Impairment Scale) et l'évaluation continue de la qualité sur stimulus unique SSCQE(Single-Stimulus Continuous Quality-Scale).[2]

1) Echelle de qualité de la méthode continue double stimulus(DSCQS)

Le but principal de la 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 (voir FIGURE 1.2). 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.[24]

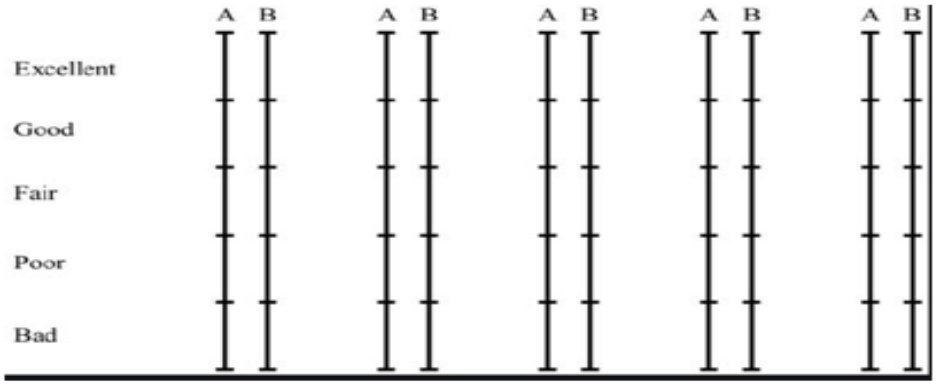

FIGURE 1.2 – Le protocole DSCQS pour évaluation subjective de la qualité perçu.

2) Echelle de dégradation sur double stimulus (DSIS)

Comme dans la méthode de DSCQS chaque essai se compose d'une paire de stimulus, la référence de l'essai. La figure suivante illustre le protocole DSIS pour évaluation subjective de la qualité perçu. Cependant, dans la méthode de double stimulus (DSIS), les deux stimulus sont toujours présentés dans le même ordre où la référence est toujours le premier 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

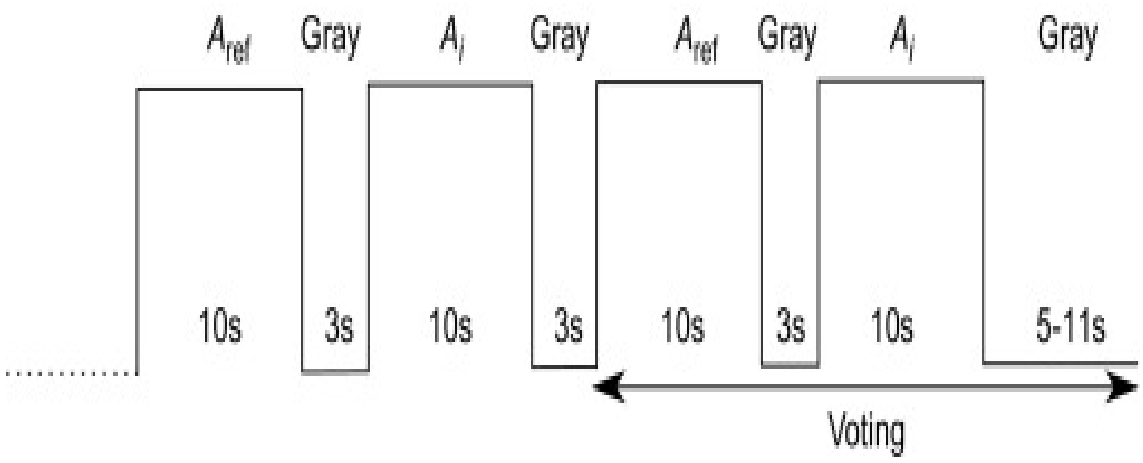

FIGURE 1.3 – Le protocole DSIS pour évaluation subjective de la qualité perçu.

de cinq niveaux. Ainsi, seulement une voix est faite pour chaque essai de DSIS.

En générale, on utilise la méthode suivant :[2]

– **Méthode Degradation Category Rating (DCR) :** 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 a 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. Seule la séquence traitée est évaluée par les participants en comparaison avec la condition de référence.[25]

3) L'évaluation continue de la qualité sur stimulus unique (SSCQE)

Au lieu de voir des paires de courtes séquences séparées, les observateurs observent un programme d'une durée 20-30 minutes en général qui a été traité par le système du test, la référence n'est pas montrée (SSCQE : Single Stimulus Continuous

Quality Evaluation.

En utilisant un glisseur, les observateurs évaluent continuellement la qualité instantanément perçue sur l'échelle de DSCQS du mauvais à l'excellent. La figure suivante présente le protocole SSCQE pour évaluation subjective de la qualité perçu.

FIGURE 1.4 – Le protocole SSCQE pour évaluation subjective de la qualité perçu.

En générale on utilisent la méthode suivante :[24]

– **Méthode Absolute Category Rating (ACR) :** la méthode ACR ou méthode d'évaluation par catégories absolues consiste a attribuer une note de qualité après chaque séquence AV visualisée/entendue. La note de jugement attribuée doit rejeter l'opinion du participant quant a la qualité audiovisuelle globale perçue, c'est-a 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 recommande 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.[25]

4. Le protocole d'évaluation à stimulus comparatif

Les méthodes comparatives permettent d'évaluer la qualité audiovisuelle en fonction d'une ou plusieurs autres audiovisuelle, venant toutes de la même audiovisuelle de référence [24] comme la méthode suivante :

– **Méthode Pair Comparaison (PC) :** 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 a l'essai additionnel. Toutes les combinaisons de paires de séquences A, B, C, etc. Devront être évaluées associes 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 exprime a 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. [2] Voici la figure ci-dessous qui schématise de manière simplifiée le chronogramme de la méthode Pair Comparaison (PC).

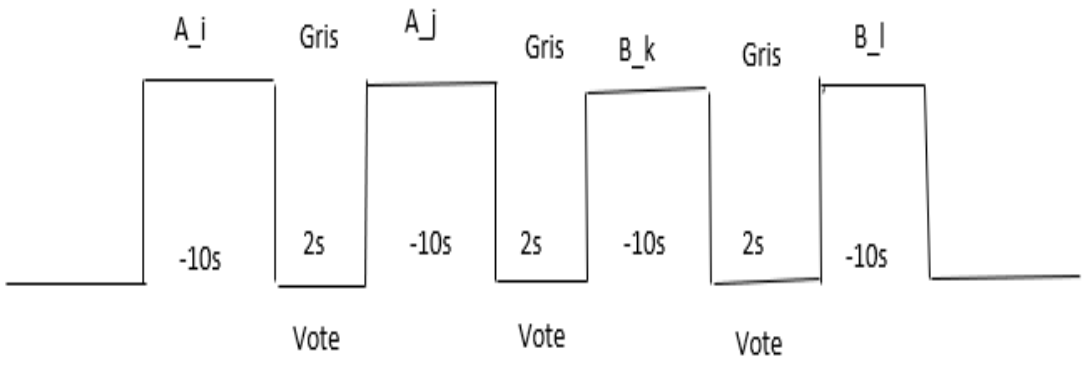

FIGURE 1.5 – Chronogramme de la méthode PC.

1.6.3 Analyse des résultats d'évaluation subjective de la qualité audiovisuelle (QAV) :

La première étape de l'analyse des résultats consiste à calculer la note moyenne ou le MoS (Mean opinion Score) pour chacune des présentations.
Le MoS est calculé à l'aide De l'équation suivante :[15]

$$MoS(i) = \frac{1}{N} \sum_{j=1}^{i} Note_i(j) \tag{1.1}$$

ou N est le nombre d'observateurs et Note est la note donnée par ces observateurs.

1.7 Conclusion

Dans ce chapitre nous avons présenté certaines définitions associées à l'audiovisuel notamment la vidéo et l'audio ensuite nous avons mentionné certaine domaine d'utilisation de l'audiovisuel tout en passons par la qualité (QoE, QoS, QoP), puis

nous avons introduit deux catégories de méthodes d'évaluation de la qualité audiovisuelle : les méthodes subjectives sont divisées en trois protocoles d'évaluation : protocole d'évaluation de stimulus simple incluant la méthode ACR, évaluation de double stimulus incluant la méthode DCR et le protocole d'évaluation de stimulus comparatif qui contient la méthode PC. Dans le chapitre suivant nous verrons l'évaluation objective est quelque approches de cette évaluation .

CHAPITRE 2

L'ÉVALUATION OBJECTIVE DE LA QUALITÉ AUDIOVISUELLE

2.1 Introduction

Les méthodes d'évaluation objective sont des méthodes calculatoires basées sur la mise en place d'un algorithme qui calcule la distance entre deux séquences audiovisuelles. Cet algorithme produit des valeurs numériques, exprimant la qualité audiovisuelle, appelées métriques. Les valeurs de ces métriques doivent refléter les notes subjectives données par les observateurs humains lors de test d'évaluation. L'intérêt majeur de ces métriques de qualité réside dans la possibilité de surveiller de manière automatique la qualité audiovisuelle en temps réel, La précision des mesures objectives de la qualité audiovisuelle ne sont pas encore assez bonne pour remplacer les tests subjectifs, mais ces tests subjectifs sont limités par ces facteurs :

— Ils sont longs et coûteux. Cela est dû au fait que les résultats subjectifs sont obtenus par des expériences avec de nombreux observateurs.

— Ils ne peuvent pas être incorporés dans des applications en temps réel telles que la compression et la transmission.

— Leurs résultats dépendent fortement des conditions physiques et de l'état émotionnel des observateurs. De plus, d'autres facteurs tels que le dispositif d'affichage et les conditions d'éclairage affectent les résultats de ces expériences.

Dans ce chapitre on va parler des domaines d'application des métriques d'évaluation objective après on vas les classifier selon leurs différences[17].

2.2 Domaine d'application des métriques d'évaluation objective de la qualité AV

Ces métriques on plusieurs domaine d'application par exemple :

— Ils peuvent être utilisés pour surveiller la qualité de vidéo dans les systèmes de contrôle de la qualité. Par exemple, les systèmes d'acquisition de vidéo peuvent utiliser une métrique objective pour surveiller et s'adapter automatiquement afin d'obtenir la meilleure qualité.

— Ils peuvent être utilisés pour évaluer les algorithmes de traitement de vidéo. Par Exemple, si un certain nombre d'algorithmes d'amélioration de vidéo sont disponibles, une métrique objective peut être utilisée pour choisir l'algorithme qui fournit des vidéos de meilleure qualité.

— Ils peuvent être utilisés pour optimiser les systèmes de transmission et de traitement de vidéo. Par exemple, dans un réseau de communication visuelle, une

métrique objective peut être utilisée pour optimiser les algorithmes d'allocation de débit dans les deux phases de codage et de transmission[30].

2.3 Classification des modèles d'évaluation objective de la qualité audiovisuelle

Les modèles de mesure de la qualité objective peuvent être classées en cinq groupes, conformément à la recommandation de UIT [31], en fonction du type de données entrée utilisées par les métriques [32] [33],Ou bien selon leur type d'informationssupplémentaires soit en réfèrence complet ou réfèrence reduite ou sans réfèrence,la classification et clarifiée dans la figure qui suit :

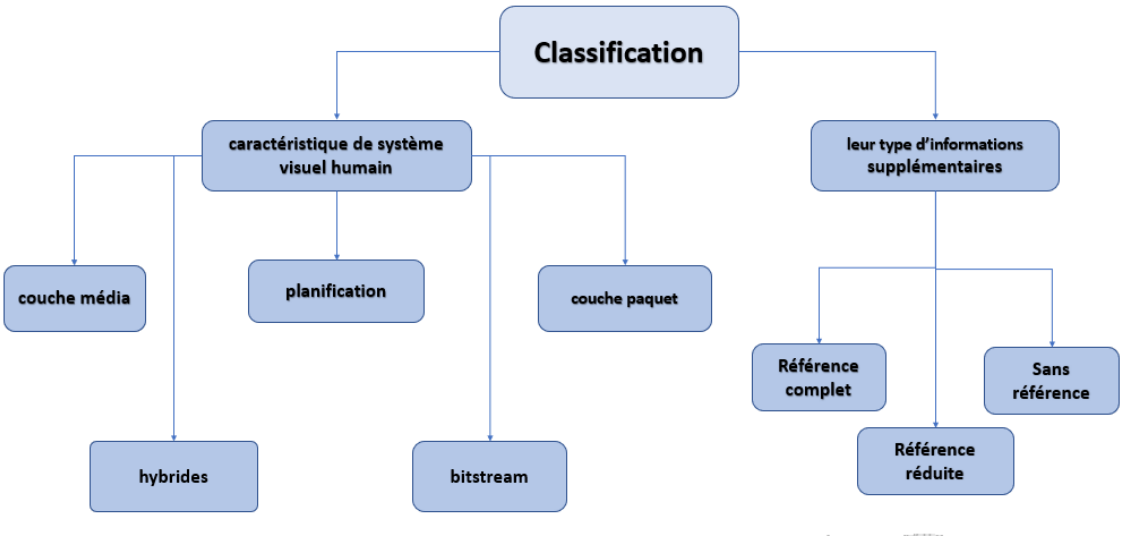

FIGURE 2.1 – la taxonomie métrique de la qualité [33],

2.3.1 Modèles de couche média

Les modèles de cette catégorie ne nécessitent aucune information sur le système en question. En particulier, ces modèles utilisent uniquement des échantillons audio ou vidéo pour estimer la qualité et peuvent être appliqués à des applications telles que l'optimisation de codec et la comparaison de codec. Les modèles fondés sur les médias ou les signaux comprennent des aspects de la perception humaine et évaluent les caractéristiques physiques du signal envoyé. Ils utilisent le signal décodé comme entrée pour calculer un score de qualité[32] [33].

2.3.2 Modèles paramétriques de couche paquet

Les solutions permettant de prévoir la qualité dans ce groupe sont légères et prédisent l'impact des configurations de l'encodage et des altérations du réseau sur la qualité multimédia. Ils utilisent généralement l'information extraite des entêtes des paquets et n'ont pas accès aux données du paquet. Ces méthodes conviennent aux cas où les données sont chiffrées [32] [33].

2.3.3 Modèles de planification

Ces modèles utilisent des paramètres de codage et de réseau pour prédire la qualité. Ils exigent donc une connaissance a priori du système en question. Les modèles de planification sont semblables aux modèles paramétriques ; la différence est d'où l'information d'entrée sera acquise. Ces modèles sont basés sur l'information de service disponible durant la phase de planification, alors que les modèles paramétriques prennent les informations d'entrée d'un service existant[32] [33].

2.3.4 Modèles de flux binaire (bitstream)

Ces modèles prédisent la qualité à l'aide d'informations de flux de bits et de couche de paquets codées utilisées dans les modèles paramétriques de couche de paquets. Ces modèles traitent en général les entêtes et le payload du flux binaire vidéo. Ils traitent l'entête du flux binaire pour extraire des informations de transport telles que le flux de transport (Transport stream, TS) et/ou les champs timestamps et les numéros de séquence du protocole Real-time Transport Protocol (RTP). Le but est de détecter la perte de paquets. Ces modèles traitent le payload du flux binaire vidéo afin d'extraire un certain nombre de caractéristiques telles que le type d'image, le nombre de tranches, le paramètre de quantification (Quantization Paramètre, QP), le vecteur de mouvement, le type de chaque macrobloc (MB) et ses partitions ainsi que les coefficients de transformation du résidu de prédiction[32] [33].

2.3.5 Modèles hybrides

Les modèles de cette classe intègrent en général deux ou plusieurs des modèles susmentionnés. Les modèles hybrides d'évaluation de la qualité exploitent les informations des entêtes de paquets, du flux élémentaire et des images reconstruites. L'information sur les images reconstruites est obtenue à partir de la séquence vidéo traitée, générée par un décodeur externe plutôt que par un décodeur interne du modèle[32] [33].

2.4 Classification selon leur type d'informations supplémentaires

Par ailleurs, Les modèles de qualité peuvent aussi être regroupés selon le type d'informations supplémentaires qu'ils traitent. Les modèles avec référence (FR) traitent généralement la séquence source originale, alors que les modèles avec référence réduite (RR) utilisent seulement une quantité limitée de l'information dérivée de la séquence source. Les modèles sans référence (NR) utilisent des séquences transmises sans utiliser aucune information du signal original.

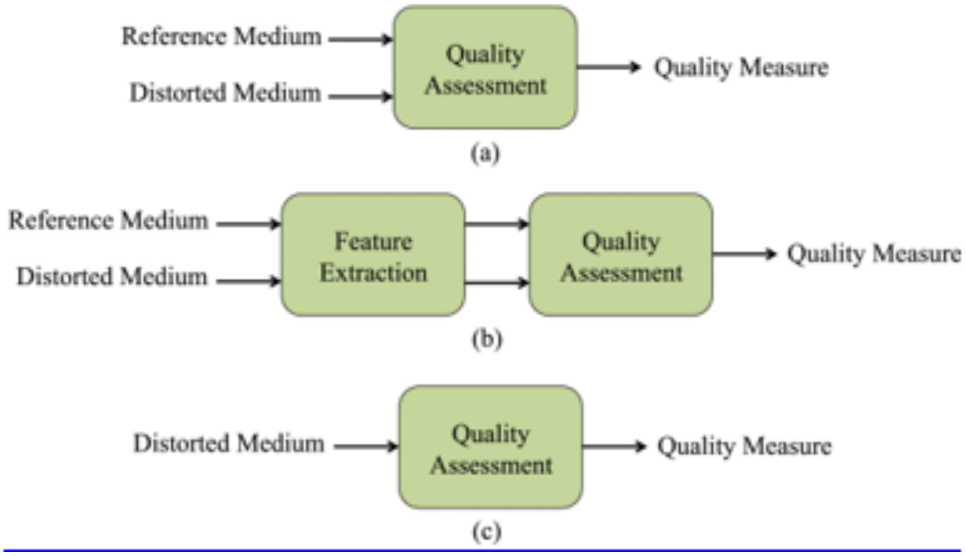

FIGURE 2.2 – Aperçu de la méthode de référence complète, la référence réduit, la méthode sans référence

Les méthodes FR mesurent la dégradation du signal de test par rapport à un signal de référence, ce qui nécessite la disponibilité du signal original complet. Bien qu'il fournisse une évaluation de la qualité objective très précise en raison de l'utili-

sation du signal original, cela est considéré comme coûteux et souvent non applicable à tous les services et applications, par exemple la surveillance IPTV[11].

Les méthodes RR évaluent la qualité en comparant une petite quantité de caractéristiques respectives extraites d'échantillons de référence et de test. Comme les méthodes RR utilisent des informations provenant du signal source, elles sont assez précises mais moins que les méthodes FR. FR et RR sont essentiels pour le contrôle de la qualité en temps non réel.

Les méthodes NR prédisent la qualité en utilisant uniquement le signal de test sans exiger de signal de référence explicite. Étant donné que ces méthodes ne nécessitent pas le signal de référence et émettent des hypothèses sur le contenu multimédia et les types de distorsions, elles sont moins précises.

En ce qui concerne les exigences de référence, FR et RR sont également désignés comme étant à deux extrémités, tandis que NR, en tant que métriques à une extrémité. En outre, en fonction de la convivialité, les méthodes objectives peuvent également être classées en méthodes hors service et en service. Dans le premier cas, aucune contrainte de temps n'est imposée et la séquence d'origine peut être disponible. Dans le deuxième cas, des contraintes de temps sont imposées et la qualité est évaluée lors de la diffusion en continu[11].

2.5 Classification selon l'approche d'évaluation utiliser

L'objectif des mesures objectives de la qualité vidéo est de donner des prédictions de qualité, qui sont en accord avec les résultats de l'évaluation subjective. Ainsi, une bonne métrique objective devrait tenir compte du processus psychophysique du

système de vision et de vision et de la perception humaine. Les principales caractéristiques de l'HVS comprennent la modélisation de la sensibilité au contraste et à l'orientation, la sélection de la fréquence, l'espace et le temps.

De la sensibilité au contraste et à l'orientation, de la sélection des fréquences ; le masquage des motifs spatiaux et temporels et la perception des couleurs.la sélection des fréquences, le masquage des motifs spatiaux et temporels et la perception des couleurs [34]. Parmi les approches qui prennent en compte le HVS on va site :

2.5.1 L'approche d'arbre de décision

Les arbres de décision sont des structures de données hiérarchiques qui peuvent être utilisées pour la classification et problèmes de régression en utilisant efficacement la stratégie diviser pour mieux Régner. Un arbre de décision est composé des nœuds de décision internes où un test est appliqué à une entrée donnée et des branches à une classification ou la valeur de régression par les nœuds des feuilles. Le processus d'estimation commence au nœud racine, traverse les nœuds de décision jusqu'à ce qu'un nœud de feuille soit atteint[11].

Les chercheurs soulignent que les apprenants arborescents ne sont pas très stables en raison de leur capacité d'anticipation limitée. Les méthodes d'ensemble tentent de surmonter les problèmes rencontrés chez les apprenants simples de l'arbre de base[11].

Parmi les travaux qui ont utilisé l'approche de l'arbre de décision on a le travail de Demirbilek, il a utilisé l'implémentation en Python des forêts d'arbres décisionnels et des techniques de bootstrap. Il à générer des deux modèles qui utilisent l'ensemble de données étendu avec 125 fonctionnalités. Rappelons que la version paramétrique de l'ensemble de données de qualité audiovisuelle de l'INRS comprend un total de 34 fonctionnalités.et il s'attende à des changements statistiquement significatifs dans la

performance du modèle avec cette augmentation radicale du nombre de fonctionnalités. Dans les deux modèles, il a mis la taille de l'arbre à 200 et $max_features$ à « tous » lors de la recherche de la meilleure répartition. Il n'a pas limité la profondeur de l'arbre. On ce qui concerne les résultats les arbres de décision ont surpassé largement les modèles d'apprentissage profond et ceux de la programmation génétique qu'il a implémenté lui mémé.

Pour le modèle basé sur les forêts d'arbres décisionnels, les valeurs des coefficients de corrélation RMSE et celui de Pearson valent respectivement 0,3082 et 0,9439. Ces valeurs étaient ces résultats pour le modèle basé sur les techniques de bootstrap. Les modèles des forêts d'arbres décisionnels et ceux basés sur les techniques de bootstrap ont performé de manière très similaire.

Toutefois, le modèle basé sur les forêts d'arbres décisionnels avait un très petit avantage par rapport au modèle basé sur les techniques de bootstrap. Les deux modèles ont également surpassé les modèles formés sur l'ensemble de données paramétriques. À partir de ces résultats, nous pouvons dire que les extensions bitstream nous ont aidés à construire des modèles plus performants[11].

2.5.2 L'approche de Deep-learning

L'apprentissage approfondi remonte aux années 1940 et a été rebaptisé à de nombreuses reprises, reflétant l'influence de différents chercheurs et de différentes perspectives. Ce n'est que récemment qu'elle a été appelée Deep-Learning. Un exemple typique de modèle d'apprentissage approfondi est le réseau approfondi ou multicouche perceptron (MLP).

Un perceptron multicouche ne fait aucune hypothèse sur les relations entre les variables. En général, ces modèles utilisent trois couches principales : un neurone

d'entrée est une couche qui représente le vecteur d'entrée, une ou plusieurs couches intermédiaires "cachées" et les neurones de sortie qui représentent le vecteur de sortie. Les nœuds de chaque couche sont reliés à tous les nœuds des couches adjacentes. Ces liens sont utilisés pour transmettre les 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 nœud. Chaque nœud effectue un calcul de base tandis que leurs liens permettent un calcul global. Le comportement global d'un réseau de neurones est influencé par le nombre de couches, le nombre des neurones de chaque couche, la façon dont les neurones sont liés et les poids associés à chaque lien. Le site de poids associé à chaque lien définit comment un premier neurone influence le deuxième neurone. Au cours de la période d'entraînement, les poids sont révisés. Avec cette approche, des couches cachées permettent de saisir les complexités dans les données, tandis que les pondérations sont ajustées à chaque itération afin d'obtenir l'erreur la plus faible dans la production. L'algorithme d'apprentissage utilisé est la rétropropagation par descente de gradient Dans l'approche de rétropropagation, pendant la phase avant, le signal d'entrée est propagé à travers le réseau, couche par couche. Dans le nœud de sortie, le signal d'erreur est calculé et ensuite ce signal d'erreur est envoyé au réseau en sens inverse, ce que l'on appelle la phase de retour. Au cours de cette phase de retour en arrière, les paramètres du réseau sont modifiés afin de minimiser l'erreur de signal. Les méthodes d'apprentissage approfondi peuvent être utilisées dans les problèmes de régression ainsi que dans le regroupement et la demandes de classement.

On ce qui concerne l'évaluation de la qualité audiovisuelle il faut examiner les ensembles de données accessibles au public ainsi que l'INRS de qualité audiovisuelle pour savoir quel type d'informations nous sommes en mesure de construire des modèles d'estimation de la qualité perçue[11].

Dans l'approche de rétropropagation, pendant la phase aller, le signal d'entrée est propagé à travers le réseau couche par couche. Dans le nœud de sortie, le signal d'erreur est calculé, puis ce signal d'erreur est envoyé au réseau dans le sens arrière, qui est appelé la phase arrière. Au cours de cette phase arrière, les paramètres du réseau sont modifiés afin de minimiser l'erreur de signal. Les méthodes de Deep Learning peuvent être utilisées dans les problèmes de régression ainsi que dans les applications de clustering et de classification [11].

Pour Maki en 2013, ils ont proposé un modèle à référence réduite paramétrique, c'est-à-dire qui utilise les caractéristiques du signal d'origine non pas pour se comparer aux mêmes caractéristiques que celles reçues, mais pour alimenter une fonction qui produira une estimation de qualité. La fonction est mise en œuvre au moyen d'un réseau de neurone, ils ont testé et comparé deux familles de réseau de neurone (Neural Networks NN) :

— les perceptrons multicouches (MultiLayer Perceptrons MLP).

— les réseaux neuronaux aléatoires (Random Neural Networks RNN),puisque la méthodologie PSQA utilise généralement des réseaux de neurones aléatoires (RNN) et a été utilisée avec succès pour l'estimation de la qualité vidéo et vocale sans références. Ils ont implémenté les modèles PSQA avec des MLP et des RNN. Pour les RNN, plusieurs topologies à trois couches ont été testées et pour les deux types de NN, les estimateurs résultants ont été évalués au moyen d'une validation croisée 10 fois. Mais le MLP a obtenu de meilleurs résultats que les RNN.

Le MLP a été configuré pour avoir une seule couche cachée, et le nombre de neurones d'entrée était égal au nombre de caractéristiques d'entrée. Le nombre de neurones cachés a été défini par la formule 2n + 1, où n est le nombre d'entrées. La

fonction tangente-hyperbolique a été choisie comme fonction d'activation des nœuds cachés, tandis que la fonction linéaire a été choisie pour le neurone de sortie. La formation du MLP a été effectuée par l'algorithme d'optimisation de la propagation inverse de Levenberg-Marquardt et algorithme d'optimisation de back-propagation[35].

2.5.3 L'approche de Combinaison

Trois expériences psychophysiques ont été menées afin de comprendre la contribution des composantes audio et vidéo à la qualité perceptive audiovisuelle globale. Il a été observé que les caractéristiques du contenu audiovisuelle sont importantes pour déterminant le MOS, prouvant qu'il existe une corrélation entre les activités spatiales et temporelles et les valeurs MOS recueillies lors des expériences. En faisant une analyse du contenu audio, Les chercheurs ont conclu que les séquences audios classées comme autres étaient plus sensibles à la dégradation par compression que les autres types de séquences audio. En observant séparément les résultats MOS audio et vidéo, il a été possible d'observer que la compression de la composante vidéo avait un impact plus important sur la qualité audiovisuelle globale que la compression de la composante audio. En utilisant une métrique vidéo et une métrique audio, ces chercheurs ont pu obtenir trois modèles objectifs de qualité audiovisuelle modèle linéaire, un modèle de Minkowski pondéré et un modèle de puissance.

Tous les modèles ont présenté de bons ajustements avec les données subjectives, avec des PCC supérieurs à 0,84. Ces modèles objectifs sont très simples et peuvent être utilisés pour prédire la qualité des signaux audio-visuels, à condition de disposer d'une métrique de qualité audio et d'une métrique de qualité vidéo.

D'autres études sont nécessaires afin de mieux comprendre comment les contenus vidéo et audio interagissent et affectent la qualité audiovisuelle. Plusieurs aspects

de la perception audiovisuelle nécessitent une attention particulière. Par exemple, la recherche sur la perception de la qualité audiovisuelle d'un point de vue neurophysiologique permettra de comprendre comment les canaux sensoriels visuels et auditifs sont combinés sur le plan perceptif. Un autre aspect est l'étude des interactions intermodales entre les composantes audio et vidéo et sa dépendance au contexte expérimental, surtout, du contenu audiovisuel. L'étude de l'impact des erreurs de synchronisation audiovisuelle (par exemple, la synchronisation des lèvres) sur la qualité audiovisuelle doit également faire l'objet d'un travail plus approfondi[36][37].

2.6 Analyse de performance des résultats objective :

Un aspect important de la modélisation de la qualité perçue est qu'un modèle objectif ne devrait pas prédire une opinion moyenne subjective de manière plus précise qu'un sujet de test moyen. L'incertitude des votes subjectifs est calculée par l'écart-type et l'intervalle de confiance correspondant. Ces paramètres statistiques visent à déterminer l'incertitude des sujets par fichier, ou par condition de test[20].
La performance d'un modèle est évaluée via trois métriques statistiques, utilisées pour informer de la précision du modèle, de sa consistance et de sa monotonie[20] [26]

– **La précision :** saisit la capacité du modèle à prédire les évaluations de qualité subjectives avec de faibles erreurs.

Lorsque les données sont tirées données de test avec une distribution proche de la normale, ces critères sont obtenus en calculant l'erreur de prédiction.
L'erreur de prédiction (c'est-à-dire l'exactitude) est obtenue à l'aide de l'erreur qua-

dratique moyenne (RMSE :Root Mean Square Error). La précision d'un modèle est habituellement déterminée par une interprétation statistique de la différence entre les valeurs MOS du test subjectif et sa prédiction sur une échelle généralisée. Un modèle précis a pour but de prédire la qualité avec l'erreur la plus faible en terme de RMSE lors des tests subjectifs [20].

$$RMSE = \sqrt{\frac{1}{N-1}\sum_{N} MoS(i) - MoS_p(i)} \tag{2.1}$$

où i est l'index de la séquence, et N est le nombre de séquences utilisées pour comparer les scores de qualité estimés aux scores subjectifs, tandis que la division à $(N-1)$ assure un estimateur sans biais pour rmse avec un intervalle de confiance à 95%.

– **La consistance :** reflète le degré auquel le modèle maintient l'exactitude des prévisions sur la plage des séquences de test. En calculant le rapport de valeurs aberrantes **(outlier ratio)**.

La consistance du modèle est obtenue en calculant soit le rapport des valeurs aberrantes (Outlier Ratio OR), soit la distribution des erreurs résiduelles. [20][26]

$$OR = \frac{TotalNoOutliers}{N} \tag{2.2}$$

Les valeurs OR sont définies comme les points pour lesquels l'erreur de prévision Perror dépasse l'intervalle de confiance de 95% de la valeur MOS moyenne, c.-à-d. si :

$$|Perror(i)| > \frac{z \times \sigma(MoS(i))}{\sqrt{Nsubj}} \tag{2.3}$$

$$\sigma(MoS(i)) = \sqrt{\frac{MoS(i) \times (1 - MoS(i))}{N}} \tag{2.4}$$

où $\sigma(MOS(i))$ représente l'écart-type des scores individuels associés à l'échantillon

de médias i, et Nsubj est le nombre d'électeurs par échantillon de médias i. La limite d'intervalle de confiance de 95% définie par la variable z est déterminée en fonction de Nsubj. Si Nsubj>30, alors la distribution gaussienne peut être utilisée, et donc z=1.96. Si Nsubj<30, la distribution t-Student est utilisée et la variable z = t et sa valeur dépend du Nsubj, respectivement le degré de liberté $df = Nsubj - 1$ [20][26].

– Enfin, **La monotonie (linéarité)** correspond au degré auquel les prédictions du modèle conviennent avec l'ampleur relative des évaluations subjectives de la qualité. En calculant le coefficient de corrélation de Pearson ,lorsqu'il n'est pas possible de vérifier que les données sont tirées d'une distribution proche de la normale, le coefficient de Spearman Rank est utilisé dans la littérature au lieu du coefficient de corrélation de Pearson comme mesure de la monotonie. [20][26]

Dans la littérature, deux métriques couramment utilisées pour le calcul de la linéarité d'un modèle existent : le coefficient de Spearman et le coefficient de corrélation de Pearson. Le coefficient de corrélation de Pearson est utilisé chaque fois que les données échantillonnées ont une distribution presque normale. Dans d'autres cas, le coefficient de Spearman est utilisé pour qualifier la linéarité entre les scores de qualité subjective prédits et réels. Le coefficient de corrélation de Pearson R , mesure la relation linéaire entre la performance d'un modèle et les données subjectives. [20][26]

$$R = \frac{\sum_{i=1}^{N}(X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \times \sqrt{\sum(Y_i - \bar{Y})^2}} \tag{2.5}$$

X_i indique le score subjectif MOS et Y_i le score objectif (MOS_p). N représente le nombre total d'échantillons pris en compte dans l'analyse Le coefficient de corrélation

de Spearman est défini comme suit :

$$R_s = \frac{\sum_{i=1}^{N}(RO(i) - \bar{RO}) \times (RO_e(i) - \bar{RO_e})}{\sqrt{\sum_{i=1}^{N}(RO(i) - \bar{RO})^2} \times \sqrt{\sum_{i=1}^{N}(RO_e(i) - \bar{RO_e})^2}} \tag{2.6}$$

Cette formule est similaire au coefficient de corrélation de Pearson, sauf le fait que l'ordre de classement des scores (rank order) de qualité subjectifs ($RO(i)$) et prédites ($RO_e(i)$) est pris au lieu des scores de qualité eux-mêmes. Cette métrique mesure donc si l'augmentation (diminution resp.) d'une variable est associée à l'augmentation (diminution resp.) de l'autre variable, indépendamment du surface de l'augmentation (diminution resp.). Cette mesure est une mesure non paramétrique de la monotonie. [20][26]

$$z = 0.5 \ln(\frac{1+R}{1-R}) \tag{2.7}$$

$$\sigma_z = \sqrt{\frac{1}{N-3}} \tag{2.8}$$

L'intervalle de confiance de 95% pour le coefficient de corrélation est déterminé à l'aide de la distribution gaussienne, qui caractérise la variable z et est donnée par l'équation[20][26]

$$z \pm 1.96 \times \sigma_z \tag{2.9}$$

2.7 Conclusion

Dans ce chapitre nous avons parlé de l'évaluation objective de la qualité audiovisuelle et ces domaines d'application, après nous l'avons classifié selon des différents modelés et leur type d'information supplémentaire ensuit on a fait une classification selon l'approche utilisé dans l'évaluation par mis les quels on à parler de l'approche

de combinaison quelle va être essentiel dans le chapitre suivant.

CHAPITRE 3

COMBINAISON DE MÉTRIQUES AUDIO-VISUELLE

3.1 Introduction

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[40].

Dans ce chapitre nous allons voir l'évaluation de la qualité audiovisuelle en utilise la méthode de combinaison entre les métriques audio et vidéo pour obtenire la qualité

audio-visuel perçue les chemins et montrer dans la figure suivante .

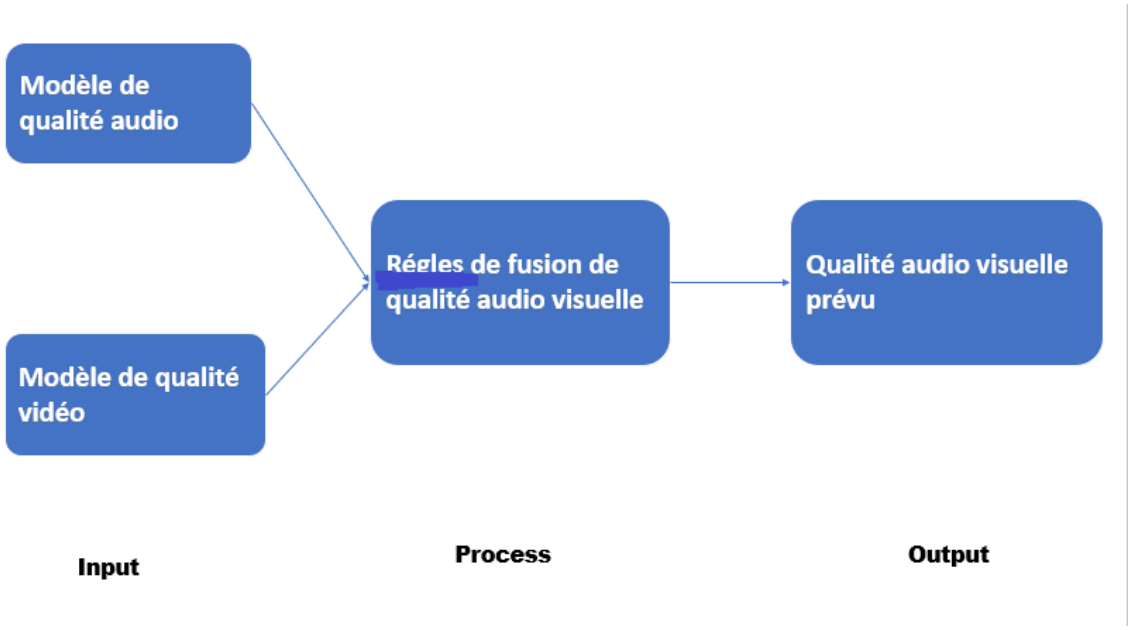

FIGURE 3.1 – Modèle de combinaison de la qualité objective de la qualité AV

3.2 L'approche de Combinaison pour les résultats subjectives :

Une série d'expériences subjectives ont été menées pour étudier l'influence mutuelle entre AQ, VQ, et AVQ depuis les années 1990. Comme mentionné précédemment, les qualités de cinq présentations différentes le tableau suivant :

Stimuli	Assessment	Quality abreviation
Audio only	Audio quality	AQ
Audio+video	Audio quality	AQ-V
Video only	video quality	VQ
Audio+video	video quality	VQ-A
Audio+video	Audio-visual quality	AVQ

TABLE 3.1 – L'évaluation de la qualité pour cinq présentations différentes

Ont pu être évaluées pour analyser l'influence mutuelle entre AQ, VQ et AVQ. Dans l'expérience menée par Beerendset De Caluwe. [39], il a été démontré que AQ_V et VQ_A n'améliorent pas immédiatement la prédiction de la qualité auditive et visuelle. Par conséquent, la plupart des expériences ont été menées pour évaluer AQ, VQ et AVQ, et explorer leur relation. Ces expériences tentent de dériver la QVA en fonction de l'AQ et de la VQ, mais l'influence d'une modalité sur l'autre modalité n'est pas étudiée. Par conséquent, la plupart des recherches se sont concentrées sur la dérivation d'un modèle permettant de déterminer la QVA à partir de la QA et de la QV. Un modèle de modèle de fusion couramment utilisé [44][40][39][38] est

$$AVQ = a_0 + a_1 AQ + a_2 VQ + a_3 AQVQ \tag{3.1}$$

Où les paramètres a_1, a_2, a_3 représentent les différents poids de la qualité audio et vidéo, ainsi que le facteur de multiplication de la qualité globale. Le paramètre a_0 n'est pas pertinent pour la corrélation entre la qualité prédictive et la qualité perçue, mais il améliore l'ajustement en termes de résidu entre les deux.

La qualité audiovisuelle globale est influencée par un couple de facteurs, parmi lesquels les QA et QV individuels sont les plus importants.

Le synchronisme, c'est-à-dire le décalage entre les stimuli audio et vidéo, est un autre élément clé.

Hayashi et al.[78]ont proposé de prendre en compte la synchronisation audio-vidéo dans le modèle de fusion.

La dégradation de la qualité (DQ) due au retard audiovisuel pour les services de visiophonie est dérivée, et la qualité multimédia perçue est calculée à partir de la AVQ et de la DQ en utilisant un modèle similaire à celle présentée dans l'équation précédente, Les résultats d'une évaluation de la qualité peuvent également être influencés par d'autres facteurs.

Cependant, il est difficile de modéliser ces facteurs externes dans une approche calculable. De plus, comme indiqué précédemment, nous avons supposé que les stimuli étaient parfaitement synchronisés dans cette étude.

Bien que la méthode de fusion de l'équation (3.1) ait été reconnue par de nombreux chercheurs, il n'existe pas de valeurs ou dérivations communément admises pour les quatre paramètres de fusion. Au contraire, les valeurs optimales des paramètres de fusion sont différentes selon les études pour différents éléments de contenu audiovisuel. Le tableau suivant résume les paramètres de fusion que nous avons pu trouver dans la littérature.

Donc les valeurs rapportées dans la littérature pour l'équation linéaire vont de : a_0 = [- 3.34 , 4.26], a_1 = [- 0.19 , 0.85], a_2=[0 , 0.89], a_3 = [- 0.01 , 0.26]. Peu d'études sur la compréhension cognitive humaine suggèrent que le canal audio et vidéo pourraient être intégrés dans une phase précoce de la perception humaine [45]. Sur cette base, plusieurs chercheurs [42],[41] ont proposé des modèles de qualité audiovisuelle comme

Laboratoire	a_0	a_1	a_2	a_3
KPN	1.12	0.007	0.24	0,088
	1.45	0	0	0.11
Bellcore	1.07	0	0	0.111
	1.295	0	0	0.107
ITS	-0.677	0.217	0.888	0
	1.514	0	0	0.121
	0.517	-0.0058	0.654	0.042
NIT	1.17	-0.144	0.186	0.154
	0.908	-0.195	0.258	0.193
ICRFE	-0.9222	0.5691	0.5064	0.1697
	-0.6313	0.2144	0.0124	0.1184
BT	1.15	0	0	0.17
	0.95	0	0.25	0.15
	4.26	0.59	0.49	0
	-3.34	0.85	0.76	-0.01
EPFL	1.98	0	0	0.103
	-1.51	0.456	0.77	0
ICU	0	0.38	0.44	0.18
	0	0.43	0.32	0.26
	0	0.35	0.58	0.07

TABLE 3.2 – les paramètres de fusion utilisé par des diffèrent laboratoire [79].

une multiplication de qualité audio et vidéo avec une importance égale, comme le montrent l'équation suivant :

$$Q_{AV} = a_0 + a_1 Q_A Q_V \tag{3.2}$$

De même, Martinez et al. [43] ont proposé trois mesures de la qualité perçue de l'audiovisuel. Le premier modèle est un modèle linéaire simple tel que donné par l'équation suivante :

$$Q_{AV} = a_0 + a_1 Q_A + a_2 Q_V \tag{3.3}$$

La seconde métrique est basée sur le modèle pondéré de Minkowski tel que donné par

$$Q_{AV} = (a_1 Q_A^p + a_2 Q_V^p)^{\frac{1}{P}} \tag{3.4}$$

Où l'exposant P est obtenu à partir de l'ajustement du modèle de Minkowski.

La troisième métrique est un modèle de puissance tel que donné par

$$Q_{AV} = (a_1 + a_2 Q_A^{p1} Q_V^{p2}) \tag{3.5}$$

3.3 L'approche de combinaison pour les résultats objectives

La même idée utilisée pour la méthode de combinaison subjective est utilisé dans cette approche, sauf que les métriques auditive et visuelle sont des métriques objectives. L'idéal est que la méthode objective de combinaison doit être capable d'imiter les prévisions de qualité d'un observateur humain, cette méthode objective de la modélisation de la qualité audiovisuelle consiste à développer des fonctions permettant de prédire la qualité audio et vidéo de manière indépendante, et ensuite de les combiner en utilisant une autre fonction de prédiction de la qualité audiovisuelle perçue globale[41]. Dans cette approche tout le protocole suit un chemin objectif d'où on fait une évaluation objective de la qualité vidéo, ensuite une évaluation objective de la qualité audio et on les combine grâce à des fonctions spécifiques.

Catégorie	Métrique
Visual	VMAF
	STRRED
	VQM
	SSIM
	MS-SSM
	VIFP
	FSIM
	GMSD
	SPEED
Auditif	PEAQ
	STOI
	VISQOL
	LLR
	SNR
	SEGSNR

TABLE 3.3 – Les différent exemples des métriques audio et vidéo.

3.3.1 Evaluation objective de la qualité vidéo

Une vidéo est une succession d'images à une certaine cadence. L'humain a comme caractéristique d'être capable de distinguer environ 20 images par seconde, En affichant plus de 20 images par seconde, il est possible de tromper L'œil et de lui faire croire à une image animée. On caractérise la fluidité d'une vidéo par le nombre d'images par secondes, elle est exprimée en FPS (Frames par second)[43].

Les techniques d'évaluation objective de la qualité peuvent également être classées

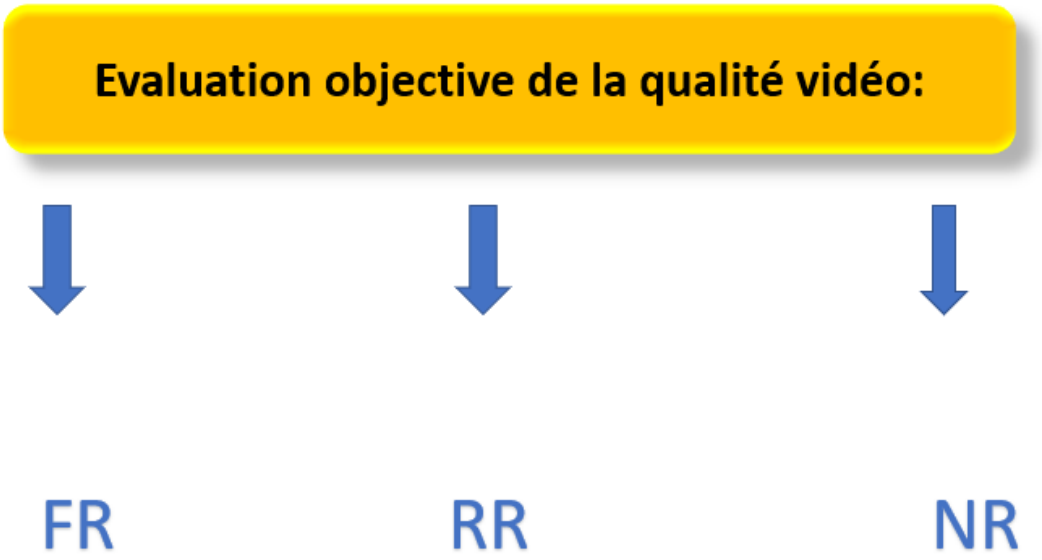

FIGURE 3.2 – Modèle d'évaluation objective de la qualité vidéo [8].

en trois catégories : référence complète (FR), référence réduite (RF) et non-référence (NR) en fonction de la disponibilité de la référence (original / idéal), informations partielles sur la référence, ou pas de référence pour évaluer la qualité, respectivement.

Dans le cas des mesures FR, la référence complète est nécessaire au point de mesure pour obtenir l'estimation de la qualité. Pour les mesures du RR, seule une partie de la référence est mise à disposition au point de mesure par un canal auxiliaire.

Enfin, pour les mesures de NR, l'estimation de la qualité est obtenue en aveugle, en utilisant seulement la vidéo de test.

Métriques à référence complète (FR)

On a deux types de métriques dans le FR :

— Métriques simples sans prendre en compte le SVH.

— Métriques avec prise en compte des caractéristiques SVH.

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

1. **Métriques basée sur l'erreur quadratique moyenne (MSE)**

   Pour déterminer le rapport ressemblance il faut faire une comparaison entre la vidéo dégradée et la vidéo parfaite par la mesure de l'erreur quadratique moyenne entre les pixels de ces deux vidéos [48] :

   Cette mesure n'est rien d'autre que la moyenne quadratique du signal erreur ou distorsion. Elle est donnée par :

$$MSE = \frac{1}{M \times N} \sum_{k=m-1}^{m} \sum_{k=n-1}^{n} (I_0(m,n) - I(m-n)) \tag{3.6}$$

   Avec (M × N) qui représente la taille de la vidéo, et I (m, n) et I' (m, n) sont respectivement les amplitudes de pixels sur les frames parfaite et déformée. Il est vraisemblable que l'œil tienne beaucoup plus compte des erreurs à grandes amplitudes, ce qui favorise la mesure quadratique [47][49].

2. **Métriques le rapport crête signal sur bruit (PSNR)**

   Cette mesure permet de quantifier la fidélité qui existe entre deux vidéos [48] (le rapport entre la puissance maximale possible d'un signal (amplitude de pixel) et la puissance du bruit), elle est une fonction de MSE :

$$PSNR = 10log_{10}(\frac{I_{max}^2}{MSE}) \tag{3.7}$$

   Pour les vidéos en couleur il faut calculer le PSNR sur chacun des trois plans colorimétriques puis faire la moyenne, dans une vidéo à niveau de gris, Imax désigne la luminance maximale possible. Une valeur de PSNR infinie correspond à une image non dégradée (MSE converge vers le zéro), et cette valeur décroît en fonction de la dégradation. Le PSNR lie donc le MSE à l'énergie maximale de la vidéo.

3. **Métriques Structural SIMilarity SSIM** 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 champ 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 perçu. La version de base de SSIM est décrite, ou l'information structurelle est recueillie par une comparaison de la luminance, le contraste et la structure.

   Equation SSIM :

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{3.8}$$

   Avec l(x,y) est défnie par :

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{3.9}$$

   Et c(x,y) est défnie par :

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{3.10}$$

   Et s(x,y) est défnie par :

$$s(x,y) = \frac{2\sigma_x\sigma_y + C_3}{\sigma_x + \sigma_y + C_3} \tag{3.11}$$

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

   Le SSIM fonctionne mieux s'il est utilisé localement. Cela signifie pour calculer

les statistiques locales μ_x, σ_x et σ_{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 utiliser 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[50].

Métriques à référence réduite (RR)

Les méthodes de qualité vidéo à RR permettent d'extraire les éléments les plus caractéristiques de la vidéo, et la qualité perçue est puis estimé en comparant ces caractéristiques dans la vidéo sous test, Les mesures vidéo RR peuvent être grossièrement classées en paquets visibilité des pertes, statistiques sur les scènes psychophysiques et naturelles techniques basées.

1. **Méthodes basées sur la visibilité de perte de paquets**

   Dans [51] et [52], une analyse de données structurée en arborescence basée sur CART (Classification et régression) et un modèle linéaire généralisé (GLM), respectivement, est effectuée pour classer si la perte de paquet est visible ou

invisible. Dans [53] et [54], les pertes de paquets multiples et H.264, en considérant les trames dans lesquelles la perte de paquets se produit, ont été étudiées dans l'amplitude et l'angle du mouvement, tandis que dans [55] et [52], la visibilité de la perte de paquets via SSIM, et la méthode d'induction de règles patient (PRIM) et le groupe d'images (GoP) sont adoptés pour la classification de perte de paquets. Aabed et AlRegib [56] exploité le flux optique pour évaluer les dégradations de la qualité du service de streaming vidéo dues aux erreurs de codage et de réseau.

2. **Méthodes psychophysiques**

   Les approches de ce groupe sont développées sur la modélisation HVS. Par exemple, [16] a utilisé plusieurs caractéristiques liées au VHS, telles que le flou et le bouchage, qui se distinguent par une analyse d'amplitude harmonique et des valeurs de force harmonique locale pour l'estimation de la qualité. De même, [15] a modélisé l'estimation de la qualité RR en utilisant la fonction de sensibilité au contraste de HVS par transformation de contour. La méthode de [57] combine la perception des couleurs, la décomposition psychophysique en sous-bandes et l'effet de masquage avec une similarité structurelle pour atteindre la métrique RR. Les chercheurs on développé une métrique RR pondérée pour la saillance permettant de simuler la perception de la qualité, appelée carte d'importance de la qualité perceptuelle (PQSM), à utiliser pour estimer la distorsion visuelle. Le PQSM est un tableau et ses éléments représentent les niveaux de signification relatifs de la qualité de la perception pour les régions correspondantes pour les images / vidéos. En particulier, la méthode décrite dans [58] utilise l'attention visuelle, la fixation / le mouvement des yeux et le trajet vision / rétine. Depuis, la caractéristique de sélectivité du HVS (Système visuel

humain) accorde plus d'attention à certaines zones / régions du signal visuel en raison de certaines combinaisons de caractéristiques saillantes de la vidéo, de signaux de connaissance de domaine et de l'association d'autres supports (audio, par exemple). Karacali et Krishnakumar [59] a conçu une métrique RR en temps réel appelée SPQR (Simplified Perceptual Quality Region) pour une application de visioconférence qui détecte le visage et ses différences entre les images. Une métrique de qualité RR pour les vidéos stéréo a été proposée dans [6] et [60], respectivement, en utilisant la visualisation avec des filigranes nuls sans disparité basés sur des vecteurs gradients et les caractéristiques temporelles de la perception vidéo et binoculaire dans HVS.

3. **Statistiques de scènes naturelles**

   Ces algorithmes supposent que les vidéos du monde réel sont constituées de scènes naturelles. Ainsi, leurs caractéristiques statistiques seraient perturbées par tout type de distorsion pouvant être utilisée pour quantifier la qualité perçue. Le modèle RR standard basé sur des statistiques de scènes naturelles (NSS) appelé métrique statistique d'images naturelles du domaine des ondelettes (WNISM) a été proposé dans [61]. La transformation de normalisation par division (DNT) a été utilisée pour surmonter les limites de la transformation en ondelettes dans [62]. Alors que, dans [63], la transformation de Tetrolet a été utilisée pour calculer les dépendances statistiques et la qualité. Ma et al. [64] argumenté et montré empiriquement que la densité gaussienne généralisée (GGD) peut dépeindre la distribution des coefficients dans le domaine de la DCT réorganisée (RDCT) pour une meilleure prédiction de la qualité vidéo RR. La divergence de Kullback-Leibler, la différence d'entropie pondérée dans les bandes DCT et la transformée en ondelettes discrète (DWT) de gradients

localement pondérés ont été utilisées avec succès pour estimer le haut niveau de qualité perçue dans [65] [66] [67], respectivement.

Métriques sans référence (NR)

Cette classe de méthodes de qualité objective ne nécessite pas d'accéder à la vidéo d'origine mes 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 bases sur les pixels et sur le flux binaire.

1. **Approches basées sur l'apprentissage**

   Dans ces approches les caractéristiques sont dérivées pendant le processus d'apprentissage lui-même. Ces approches extraient d'abord les caractéristiques des données d'entrée par le biais d'apprentissage, qui sont ensuite mises en commun pour produire une visibilité globale de la distorsion, puis converties en un score de qualité perceptuelle par ajustement à un modèle de régression. Les approches basées sur l'apprentissage des caractéristiques sont plus efficaces que les approches basées sur NSS en raison de leurs capacités d'apprendre automatiquement de meilleures fonctionnalités à partir de pixels d'image brutes.

2. **Approches basées sur l'hybrides**

   Les techniques qui combinent les statistiques de flux de bits codées et de supports décodés sont appelées méthodes d'estimation de la qualité hybride sans référence. Les méthodes hybrides peuvent être divisées en deux catégories : les caractéristiques ou les artefacts basés sur les pixels et le train binaire, et les statistiques des coefficients de transformation.

3. **BIQI**

Cette approche a été présentée par Moorthy et Bovik [68], ils estiment la qualité à partir d'une transformée en ondelettes utilisant la base d'ondelettes de Daubechies 9/7 [69]. 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 gaussienne généralisée (DGG)[69].

3.3.2 Evaluation objective de la qualité 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, on va parler de 3 méthodes pour évaluer la qualité objective de l'audio : Méthode intrusive, Méthode non intrusive, Méthode paramétrique.

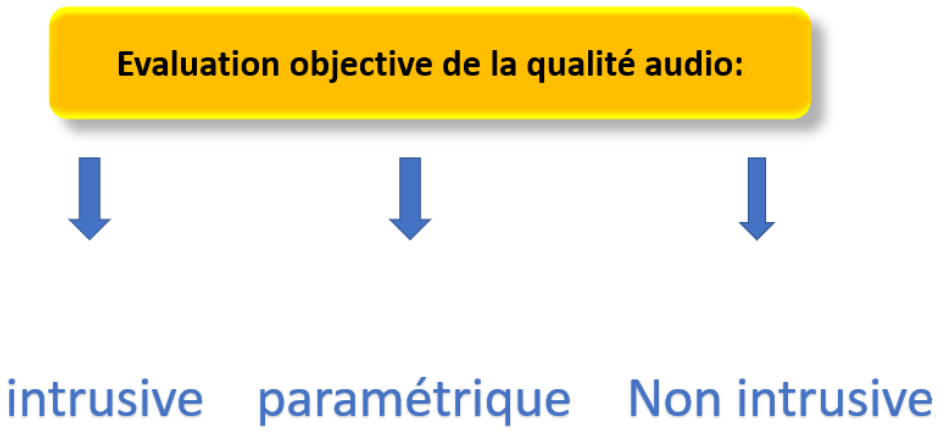

FIGURE 3.3 – Modèle d'évaluation objective de la qualité audio

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, PSNR et SSIM.

1. **L'évaluation perceptive de Ia qualité vocale (PESQ)**

   Normalisé par l'UIT-T [70] est une méthode de prédiction de la qualité subjective pour la téléphonie avec la 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.

2. **L'évaluation perceptuelle de la qualité audio (PEAQ)**

   Actuellement, le seul standard en vigueur est la méthode PEAQ, connu aussi sous la référence Rec.ITU-R BS.1387 [71]. Cette recommandation constitue la synthèse d'un ensemble de techniques qui existaient avant et dont le principe est illustré sur la figure 2.18. Il s'agit d'une méthode dite avec référence ou on compare un signal de référence à sa version potentiellement d´dégradé récupérée à la sortie du système évalué.

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 divisés en deux classes : approches a priori et sources[11].

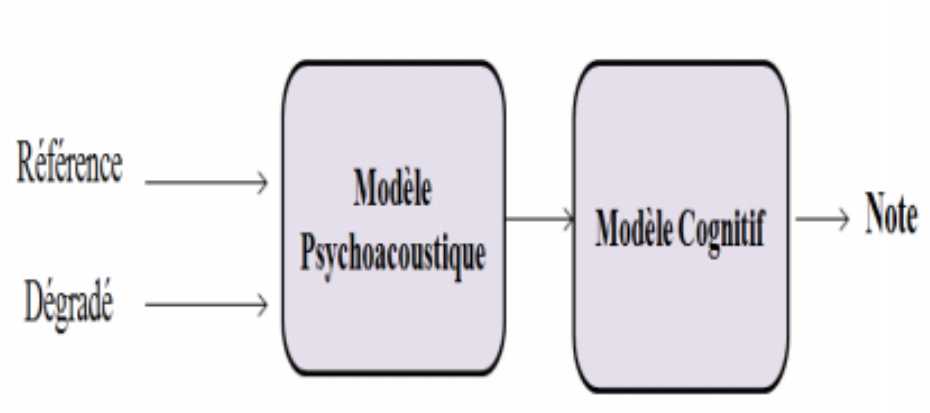

FIGURE 3.4 – principe de la PEAQ

1. **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[80] 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[11].

2. **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 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 [72]. Ont mis au point une métrique normalisée SRMR (speech-to-réverbération-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. A son tour, Beerends et al [81]. 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[73].

3) Méthode paramétrique

Les modèles paramétriques estiment la qualité à I '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 modalisées 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 canal téléphoniques sur le trafic d'appels en direct sans entraver L'Apple, et avec une connaissance du réseau et le système auditif humain produit des valeurs de qualité[11].

3.4 Conclusion

Dans ce chapitre nous avons parlé de l'approche de combinaison pour l'évaluation de la qualité audiovisuelle, on a introduit les trois fonctions (linéaire, Minkowski, power) qui sert a combinée entre les deux métriques auditive et visuelle subjective et objective, en passant par les métriques d'évaluation objective. Dans le chapitre suivant, nous appliquons cette approche d'évaluation avec les trois fonctions linéaire, Minkowski, et power.

CHAPITRE 4

LES TESTS ET RÉSULTATS EXPÉRIMENTAUX

4.1 Introduction

Dans ce chapitre, nous voulons étudier l'utilisation des modèles combinés pour intégrer des estimations de la qualité audio et vidéo uniques dans le but de prédire la qualité audiovisuelle globale. Pour obtenir les estimations de la qualité audio et vidéo, nous allons utiliser un ensemble de métriques de qualité audio et vidéo matures et suffisamment testées, après on va inclure un terme croisé multiplicatif (qualité audio $\times$ qualité vidéo), est utilisé pour prédire la qualité audiovisuelle. Dans ce chapitre, en plus de tester des modèles linéaire et puissance, nous avons également testé des modèles Minkowski.

4.2 Environnement de travail

4.2.1 Langage

Nous avons utilisé Matlab2015 comme langage de programmation, parmi les raisons de cette utilisation :

–Interfaces de langage et de bibliothèque externes, y compris Python, Java, C, C++, .NET et les services Web.

–À mesure que la taille et la complexité de vos projets augmentent, MATLAB fournit des fonctionnalités pour prendre en charge les pratiques de développement de logiciels collaboratifs. Par exemple, vous pouvez intégrer vos fichiers MATLAB aux systèmes de contrôle de source Git™ ou Subversion® ou tester la fonctionnalité et les performances de votre code. Pour partager du code avec d'autres, emballer des projets ou d'autres fichiers en tant que boîte à outils.

–Lorsque on travaille dans l'éditeur, MATLAB il identifie automatiquement les problèmes de codage potentiels. Les fonctionnalités de débogage aident à diagnostiquer des problèmes spécifiques. De plus, on peut générer des rapports qui nous aidons à mettre à jour notre code lorsque nous effectuez une mise à niveau vers une version plus récente de MATLAB.

4.2.2 Caractéristique de la plateforme

On utilise une machine avec les caractéristiques suivantes : –SYSTEME D'EXPLOITATION : Microsoft Windows 10 pro. –PROCESSEUR : 2.20 GHz Intel5(R) Core (TM) i5-5200M. –RAM : 8.00 GO. –CARTE GRAPHIQUE : NVIDIA GFORCE 8200.

4.3 La base de données utilisée

Les expériences font partie de la base de données de qualité audio-visuelle UnB (UnB-AVQ) [74]. Dans ces expériences, six séquences vidéo haute définition originales (avec des composants audio et vidéo) de The Consumer Digital Vidéo Library sont utilisées. Des images représentatives des séquences originales sont représentées sur la figure 4.1. Chaque séquence dure huit secondes et a une résolution spatiale et temporelle de 1280x720 (720p) avec un espace colorimétrique de 4 :2 :0 et 30 images par seconde (fps) respectivement. Pour l'expérience I, chacune des séquences de test

FIGURE 4.1 – Exemples d'images de vidéos originales utilisées dans les expériences subjectives : (a) « Boxer », (b) « Park Run », (c) « Crowd Run », (d) « Basketball », (e) « Music, » Et (f) « Reporter ».
[75]

vidéo originales (sans audio) a été compressée à l'aide du codec H.264. Quatre valeurs de débit binaire différentes ont été utilisées : 30, 2, 1 et 0,8 Mbps. Cette conception de test a abouti à 6(séquences originales) $\times$ 4(valeurs de débit) = 24 conditions de test.

Pour l'expérience II, seule la composante audio des vidéos a été prise en compte.

Le composant audio a été compressé à l'aide de la norme de codage MPEG-1 layer-3. Trois valeurs de débit binaire ont été utilisées : 128, 96 et 48 kbps. Cette conception de test a abouti à 6(séquences originales) × 3(valeurs de débit) = 18 conditions de test.

Pour l'expérience III, les composants audio et vidéo des séquences de test ont été compressés. Les composants vidéo ont été compressés avec H.264, en utilisant les mêmes valeurs de débit binaire utilisées dans l'expérience I (30, 2, 1 et 0,8 Mbps). Les composants audios ont été compressés avec la norme de codage MPEG-1 couche-3, en utilisant les mêmes valeurs de débit binaire utilisées dans l'expérience II (128, 96 et 48 kbps). Compte tenu des trois valeurs de débit binaire des composants audio et des quatre valeurs de débit binaire des composants vidéo (3 débits audio × 4 débits vidéo) pour les six originaux, cela a donné un total de $3 \times 4 \times 6 = 72$ conditions de test[75].

Les jugements donnés par les sujets à n'importe quelle séquence de test sont appelés scores subjectifs (mos-subj).

4.4 Détails de l'expérience

4.4.1 Evaluation de la qualité audiovisuelle basé sur les MOS-subjectives (MOSa, MOSv)

On va utiliser les MOS_a et les MOS_v donnée par la base de données UNB et les combine à l'aide des fonctions linéaire, Minkowski et power pour obtenir une qualité audiovisuelle globale, La fonction linéaire est donné par l'équation suivante :

$$AVQ = a_0 + a_1 AQ + a_2 VQ + a_3 AQVQ \tag{4.1}$$

Où les paramètres a_0, a_1, a_2, a_3 représentent les différents poids de la qualité audio et vidéo, ainsi que le facteur de multiplication de la qualité globale. Le paramètre a0 n'est pas pertinent pour la corrélation entre la qualité prédictive et la qualité perçue, mais il améliore l'ajustement en termes de résidu entre les deux.

Bien que la méthode de fusion de l'équation précédente ait été reconnue par de nombreux chercheurs, il n'existe pas de valeurs ou dérivations communément admises pour les quatre paramètres de fusion. Au contraire, les valeurs optimales des paramètres de fusion sont différentes selon les études pour différents éléments de contenu audiovisuel. Le tableau suivant résume les paramètres de fusion que nous avons pu trouver dans la littérature.

Laboratoire	a_0	a_1	a_2	a_3
KPN	1.12	0.007	0.24	0,088
	1.45	0	0	0.11
Bellcore	1.07	0	0	0.111
	1.295	0	0	0.107
ITS	-0.677	0.217	0.888	0
	1.514	0	0	0.121
	0.517	-0.0058	0.654	0.042
NIT	1.17	-0.144	0.186	0.154
	0.908	-0.195	0.258	0.193
ICRFE	-0.9222	0.5691	0.5064	0.1697
	-0.6313	0.2144	0.0124	0.1184
BT	1.15	0	0	0.17
	0.95	0	0.25	0.15
	4.26	0.59	0.49	0
	-3.34	0.85	0.76	-0.01
EPFL	1.98	0	0	0.103
	-1.51	0.456	0.77	0
ICU	0	0.38	0.44	0.18
	0	0.43	0.32	0.26
	0	0.35	0.58	0.07

TABLE 4.1 – les paramètres de combinaison linéaire utilisé par des diffèrent laboratoire

[76]

De même, On va aussi utiliser les deux fonctions Minkowski et power proposé par Martinez et al. [77]. Donné par l'équation suivante :

La fonction Minkowski :

$$Q_{AV} = (a_1 Q_A^p + a_2 Q_V^p)^{\frac{1}{P}} \tag{4.2}$$

Où p = 0.0001, a_1 =0.7024, a_2 =0.2976 [77].

La fonction power :

$$Q_{AV} = (a_1 + a_2 Q_A^{p1} Q_V^{p2}) \tag{4.3}$$

Où p_1 = 1.3213, p_2=0.6533, a_1 =-12.9734, a_2 =-0.0109 [77].

On va calculer la corrélation entre notre résultat (QAV) et le MOSav de la base de données en utilisent le coefficient de corrélation de Pearson (PCC) et La Racine de l'erreur quadratique moyenne (RMSE).

4.4.2 Evaluation de la qualité audiovisuelle basé sur les MOS-prédictives (Qa, Qv)

La première étape : Nous allons calculer la qualité audio ainsi que la qualité vidéo individuellement pour cela on a utilisé les métriques suivantes : SSIM, PSNR pour Audio et SSIM, PSNR, MS-SSIM3 pour vidéo

Ensuite on va calculer la corrélation entre nos résultats (Qa, Qv) et le (MOSa, MOSv) respectivement en utilisent le coefficient de corrélation de Pearson (PCC), Racine de l'erreur quadratique moyenne (RMSE),

Deuxième étape : ensuite on combine ces métriques à l'aide des trois fonctions (linéaire, Minkowski, power), le résultat de la combinaison est corrélé avec MOSav de la base en utilisent le coefficient de corrélation de Pearson (PCC)

4.5 Analyse des Résultats et Discussion

On ce qui concerne l'évaluation de la qualité audiovisuelle basé sur les MOS-subjective (MOSa, MOSv) on a obtenu des très bons résultats où PCC est supérieure à 0.9 dans les trois combinaisons, un petit avantage est observé pour le modèle Minkowski et power où RMSE est très grande dans la combinaison linéaire que celle des deux autres (Figure 4.3).

Donc on constate que nos 3 modèles de combinaison sont efficaces et on peut passer à notre 2eme partie de travaille

Fonction de combinaison	Audio	Vidéo	PCC	RMSE
Linéaire	MOSa	MOSv	0.9088	51.3134
Minkowski	MOSa	MOSv	0.9213	12.3642
Power	MOSa	MOSv	0.9266	6.5716

TABLE 4.2 – résultat des combinaisons basé sur les MOS-subjective

Dans cette partie on a suivi un chemin tout objectif en implémentant des métrique audio (psnr, ssim) et vidéo (psnr, ssim, mssim3) individuellement et les figures 4.4, 4.5, 4.6, 4.7 et figures 4.8 montre les résultats de chaque métrique implémentée respectivement

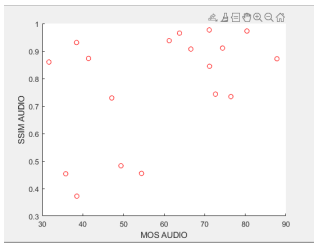

FIGURE 4.2 – MOS audio versus ssim

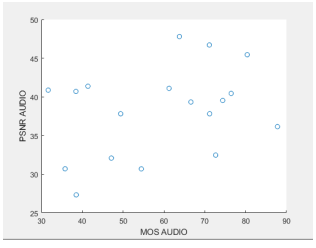

FIGURE 4.3 – MOS audio versus psnr

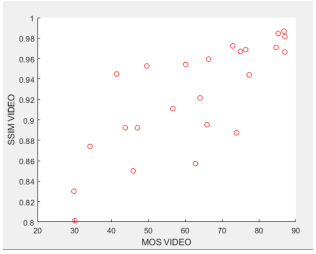

FIGURE 4.4 – MOS vidéo versus ssim

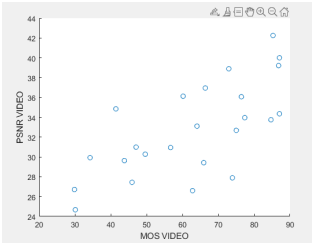

FIGURE 4.5 – MOS vidéo versus psnr

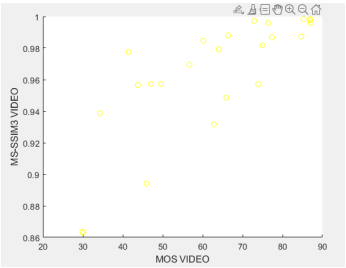

FIGURE 4.6 – MOS vidéo versus mssim3

On a remarqué que ces métriques nous ont donné des résultats visuels où le PCC est supérieur a 0.68 et par contre un des résultats auditifs inférieur à 0.5 mais qu'il permet de continue notre chemin de travail ces résultats sont montres dans (Figure 4.9).

	Métrique	PCC	RMSE
Audio	PSNR	0.334	5.7419
	SSIM	**0.4517**	**0.1822**
Vidéo	PSNR	0.6889	3.7122
	SSIM	0.7785	0.0364
	MS-SSIM3	**0.7816**	**0.0237**

TALBE 4.3 – résultat des métriques qui calcule le MOS-prédictives

Les prévisions individuelles des mesures de qualité audio et vidéo ont été intégrées à l'aide de trois modèles de combinaison : linéaire, Minkowski et puissance. Les mesures audio et vidéo ont été combinées, ce qui a donné lieu à 18 mesures de qualité audiovisuelle qu'ils nous ont donnée des résultats de corrélation où le PCC est inferieure a 0.7 et supérieur à 0.35 les résultats sont montre dans le (Figure 4.10).

	Audio	Vidéo	PCC
Linear	PSNRa	PSNRv	0.5702
	SSIMa	SSIMv	0.6708
	PSNRa	MS-SSIM3	0.4426
	SSIMa	PSNRv	0.6054
	PSNRa	SSIMv	0.5050
	SSIMa	**MS-SSIM3**	**0.6967**
Minkowski	PSNRa	PSNRv	0.5813
	SSIMa	SSIMv	0.4306
	PSNRa	MS-SSIM3	0.5209
	SSIMa	PSNRv	0.5407
	PSNRa	**SSIMv**	**0.6097**
	SSIMa	MS-SSIM3	0.3890
Power	**PSNRa**	**PSNRv**	**0.5960**
	SSIMa	SSIMv	0.4109
	PSNRa	MS-SSIM3	0.4882
	SSIMa	PSNRv	0.5256
	PSNRa	SSIMv	0.5472
	SSIMa	MS-SSIM3	0.3756

TABLE 4.4 – résultat des métriques qui calcule le MOS-prédictives

En observant que les performances des modèles de combinaison sont importantes,

de noter que le meilleur résultat de la combinaison linéaire et la combinaison entre SSIM audio et le MS-SSIM3 vidéo et le meilleur résultat de la combinaison Minkowski est entre PSNR audio et SSIM vidéo aussi que le meilleur résultat de la combinaison power et entre PSNR audio et PSNR vidéo donc les trois sont différents ce qui nous permet de dire que la combinaison de qualité audiovisuelle globale est plus influencée de la multiplication pondéré entre la qualité audio et la qualité vidéo que par le poids donner pour l'audio ou la vidéo.

4.6 Conclusion

Dans ce chapitre nous avons fait l'évaluation de la qualité audiovisuelle basée sur les MOS-subjectives et les MOS-prédictives et les résultats obtenus ont montré l'efficacité de cette technique et cela nous permet de classer notre méthode parmi les bonnes méthodes d'évaluation objectives de la qualité audiovisuelle.

CONCLUSION 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,le multimédia immersif ansi que les jeux de réalité virtuelle, vidéoconférence et présentations multimédias éducatives, pour n'en nommer que quelques-unes. Ces applications multimédias font désormais partie intégrante de la vie quotidienne et devraient connaître une croissance exponentielle supplémentaire, donc l'audio et la vidéo sont deux modalités de base dans la plupart des applications multimédias et l'évaluation de la qualité de ces deux signaux numériques est l'un des problèmes fondamentaux et complexes du traitement multimédia, donc on a évalué la qualité audiovisuelle en utilisant la méthode de combinaison pour regler cette problématique, En observant les performances des métriques individuelles de notre travaille, nous avons remarqué que les trois modèles de combinaison ont une bonne capacité d'intégration, Il convient de noter que, l'un des objectifs de ce travail était de tester différents modèles de combinaison et d'étudier leur capacité d'intégration, en termes de performances

de précision. Bien que les résultats soient prometteurs, nous pensons qu'une amélioration dans la métrique individuelle permet d'obtenir des meilleures performances

En perspectives de travail :

Dans le futur travail il est question d'utiliser des métriques plus complexes et les combinaient à l'aide de ces fonctions après effectuer des tests en utilisant des bases de données audiovisuelles plus diverses, contenant plusieurs types de dégradations audio et vidéo et enfin faire des comparaisons avec d'autres modèles de combinaison comme fuzzy logique et l'apprentissage automatique.

BIBLIOGRAPHIE

[1] G. M. Wilson and M. A. Sasse, "Do users always know what's good for them ? utilising physiological responses to assess media quality," in *People and computers XIV—Usability or else !*, pp. 327–339, Springer, 2000.

[2] Z. Akhtar and T. H. Falk, "Audio-visual multimedia quality assessment : A comprehensive survey," *IEEE access*, vol. 5, pp. 21090–21117, 2017.

[3] J. Zheng, K. Chan, and I. Gibson, "Virtual reality," *Ieee Potentials*, vol. 17, no. 2, pp. 20–23, 1998.

[4] M. El Mansouri, *Le jeu vidéo didactique ou serious game : processus de conception, ingénierie didactique et game design.* PhD thesis, Université Côte d'Azur, 2019.

[5] C. Licoppe, M. Verdier, and L. Dumoulin, "Courtroom interaction as a multimedia event," *The Electronic Journal of Communication/La revue electronique de communication*, vol. 23, no. 1-2, p. sp, 2013.

[6] S. Möller and A. Raake, *Quality of experience : advanced concepts, applications and methods*. Springer, 2014.

[7] K. B. et al, *Qualinet white paper on definitions of quality of experience*. in Proc. 5th Qualinet Meet, 2013.

[8] Z. Akhtar, "Audio-visual multimedia quality assessment : A comprehensive survey," *IEEE Transactions on Broadcasting, 53(2) :pp. 449-458*, (2017).

[9] G. Ghinea and J. P. Thomas, "Quality of perception : user quality of service in multimedia presentations," *IEEE Transactions on Multimedia*, vol. 7, no. 4, pp. 786–789, 2005.

[10] B. Cadon, "Détruire ou altérer le fonctionnement des machines numériques, la résistance du 21eme siécle," *Revue Possibles*, vol. 45, no. 1, pp. 39–49, 2021.

[11] E. Demirbilek, *Modéles d'apprentissage automatique d'estimation de Qualité percue dans les commenications en temps réel.* PhD thesis, (2017).

[12] A. M. Rohaly, P. J. Corriveau, J. M. Libert, A. A. Webster, V. Baroncini, J. Beerends, J.-L. Blin, L. Contin, T. Hamada, D. Harrison, *et al.*, "Video quality experts group : Current results and future directions," in *Visual Communications and Image Processing 2000*, vol. 4067, pp. 742–753, International Society for Optics and Photonics, 2000.

[13] M.-N. Garcia, *Parametric Packet-based Audiovisual Quality Model for IPTV Services*. Springer, 2014.

[14] X. Min, G. Zhai, J. Zhou, M. C. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Transactions on Image Processing*, vol. 29, pp. 6054–6068, 2020.

[15] M. L. Fentazi, O. Bouzit, *et al.*, "Évaluation aveugle de qualité des images fi xes," 2020.

[16] M. P. Hollier, A. N. Rimell, D. S. Hands, and R. M. Voelcker, "Multi-modal perception," *BT Technology Journal*, vol. 17, no. 1, pp. 35–46, 1999.

[17] Z. Mahrouk, "série j :transmission des signaux radiophoniques, télévisuels et autres signaux multimédias," 1999.

[18] S. Bech and N. Zacharov, *Perceptual audio evaluation-Theory, method and application*. John Wiley & Sons, 2007.

[19] S. Bech and N. Zacharov, *Perceptual Audio Evaluation Theory, Method and Application*. New York, NY, USA : Wiley, 2006.

[20] P. ITU-T, "1401 : Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," *ITU-T Recommendation*, p. 1401, 2012.

[21] P. ITU-T RECOMMENDATION, "Subjective audiovisual quality assessment methods for multimedia applications," 1998.

[22] R. P. ITU-T, "920, interactive test methods for audiovisual communications," *International Telecommunications Union Radiocommunication Assembly*, 1996.

[23] P. ITU-T, "1401 : Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," *ITU-T Recommendation*, p. 1401, 2012.

[24] G. T. Tchendjou, *Contrôle des performances et conciliation d'erreurs dans les décodeurs d'image*. PhD thesis, Université Grenoble Alpes, 2018.

[25] D. B. EDDINE., *Évaluation de la qualité perceptuelle des signaux multimédias : évaluation multicritére basée sur la fusion des métriques*. PhD thesis, (21/12/20).

[26] M.-N. Garcia, *Parametric packet-based audiovisual quality model for IPTV services*. PhD thesis, (2014).

[27] A. Mohan, K. Gauen, Y.-H. Lu, W. W. Li, and X. Chen, "Internet of video things in 2030 : A world with many cameras," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–4, 2017.

[28] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Göring, and B. Feiten, "A bitstream-based, scalable video-quality model for http adaptive streaming : Itu-t p. 1203.1," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, 2017.

[29] M.-N. Garcia, P. List, S. Argyropoulos, D. Lindegren, M. Pettersson, B. Feiten, J. Gustafsson, and A. Raake, "Parametric model for audiovisual quality assessment in iptv : Itu-t rec. p. 1201.2," in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 482–487, IEEE, 2013.

[30] A. Mohan, K. Gauen, Y.-H. Lu, W. W. Li, and X. Chen, "Internet of video things in 2030 : A world with many cameras," in *2017 IEEE international symposium on circuits and systems (ISCAS)*, pp. 1–4, IEEE, 2017.

[31] S. Zadtootaghaj, S. Schmidt, S. S. Sabet, S. Möller, and C. Griwodz, "Quality estimation models for gaming video streaming services using perceptual video quality dimensions," in *Proceedings of the 11th ACM Multimedia Systems Conference*, pp. 213–224, 2020.

[32] A. Raake, M.-N. Garcia, W. Robitza, P. List, S. Göring, and B. Feiten, "A bitstream-based, scalable video-quality model for http adaptive streaming : Itu-t p. 1203.1," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, IEEE, 2017.

[33] M.-N. Garcia, P. List, S. Argyropoulos, D. Lindegren, M. Pettersson, B. Feiten, J. Gustafsson, and A. Raake, "Parametric model for audiovisual quality assess-

ment in iptv : Itu-t rec. p. 1201.2," in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 482–487, IEEE, 2013.

[34] S. Winkler, *Digital video quality : vision models and metrics*. John Wiley & Sons, 2005.

[35] T. Mäki, D. Kukolj, D. Đordević, and M. Varela, "A reduced-reference parametric model for audiovisual quality of iptv services," in *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 6–11, IEEE, 2013.

[36] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics : A survey," *Journal of visual communication and image representation*, vol. 22, no. 4, pp. 297–312, 2011.

[37] U. Engelke and H.-J. Zepernick, "Perceptual-based quality metrics for image and video services : A survey," in *2007 Next Generation Internet Networks*, pp. 190–197, IEEE, 2007.

[38] M. H. Pinson, W. Ingram, and A. Webster, "Audiovisual quality components," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 60–67, 2011.

[39] M.-N. Garcia, R. Schleicher, and A. Raake, "Impairment-factor-based audiovisual quality model for iptv : influence of video resolution, degradation type, and content type," *EURASIP Journal on Image and Video Processing*, vol. 2011, pp. 1–14, 2011.

[40] S. Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," *IEEE transactions on multimedia*, vol. 8, no. 5, pp. 973–980, 2006.

[41] J. G. Beerends and F. E. De Caluwe, "The influence of video quality on perceived audio quality and vice versa," *Journal of the Audio Engineering Society*, vol. 47, no. 5, pp. 355–362, 1999.

[42] E. Styles, *The psychology of attention*. Psychology Press, 2006.

[43] H. B. Martinez and M. C. Farias, "Full-reference audio-visual video quality metric," *Journal of Electronic Imaging*, vol. 23, no. 6, p. 061108, 2014.

[44] D. S. Hands, "A basic multimedia quality model," *IEEE Transactions on multimedia*, vol. 6, no. 6, pp. 806–816, 2004.

[45] D. S. Hands, "A basic multimedia quality model," *IEEE Transactions on multimedia*, vol. 6, no. 6, pp. 806–816, 2004.

[46] D. S. Hands, "A basic multimedia quality model," *IEEE Transactions on multimedia*, vol. 6, no. 6, pp. 806–816, 2004.

[47] "Technologie eyevinn. evaluation de la qualité vidéo." https://medium.com/@eyevinntechnology/video-quality-assessment-34abd35f96c0, 2018.

[48] M. L. Fentazi, O. Bouzit, *et al.*, *Évaluation aveugle de qualité des images fi xes*. PhD thesis, University of Jijel, 2020.

[49] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan, "Modeling packet-loss visibility in mpeg-2 video," *IEEE transactions on Multimedia*, vol. 8, no. 2, pp. 341–355, 2006.

[50] T.-L. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. C. Cosman, and A. R. Reibman, "A versatile model for packet loss visibility and its application to packet prioritization," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 722–735, 2009.

[51] S. Paluri, K. K. Kambhatla, B. A. Bailey, P. C. Cosman, J. D. Matyjas, and S. Kumar, "A low complexity model for predicting slice loss distortion for prio-

ritizing h. 264/avc video," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 961–985, 2016.

[52] F. Tommasi, V. De Luca, and C. Melle, "Packet losses and objective video quality metrics in h. 264 video streaming," *Journal of Visual Communication and Image Representation*, vol. 27, pp. 7–27, 2015.

[53] A. R. Reibman and D. Poole, "Predicting packet-loss visibility using scene characteristics," in *Packet Video 2007*, pp. 308–317, IEEE, 2007.

[54] M. A. Aabed and G. AlRegib, "Reduced-reference perceptual quality assessment for video streaming," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 2394–2398, IEEE, 2015.

[55] M. Carnec, P. Le Callet, and D. Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference," *Signal Processing : Image Communication*, vol. 23, no. 4, pp. 239–256, 2008.

[56] Z. Lu, W. Lin, E. Ong, X. Yang, and S. Yao, "Pqsm-based rr and nr video quality metrics," in *Visual Communications and Image Processing 2003*, vol. 5150, pp. 633–640, International Society for Optics and Photonics, 2003.

[57] B. Karacali and A. Krishnakumar, "Measuring video quality degradation using face detection," in *2012 35th IEEE Sarnoff Symposium*, pp. 1–5, IEEE, 2012.

[58] W. Zhou, G. Jiang, M. Yu, F. Shao, and Z. Peng, "Reduced-reference stereoscopic image quality assessment based on view and disparity zero-watermarks," *Signal Processing : Image Communication*, vol. 29, no. 1, pp. 167–176, 2014.

[59] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Human vision and electronic imaging X*, vol. 5666, pp. 149–159, International Society for Optics and Photonics, 2005.

[60] R. Soundararajan and A. C. Bovik, "Rred indices : Reduced reference entropic differencing framework for image quality assessment," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1149–1152, IEEE, 2011.

[61] A. A. Abdelouahad, M. El Hassouni, H. Cherifi, and D. Aboutajdine, "Image quality assessment measure based on natural image statistics in the tetrolet domain," in *International Conference on Image and Signal Processing*, pp. 451–458, Springer, 2012.

[62] P. CHAVEL, D. KUAN, A. SAWCHUK, and T. STRAND, "Techniques de réduction de speckle," in *1° Colloque Image : traitement, synthèse, technologies et applications, FRA, 1984*, GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 1984.

[63] J.-H. Seo, S. B. Chon, K.-M. Sung, and I. Choi, "Perceptual objective quality evaluation method for high quality multichannel audio codecs," *Journal of the Audio Engineering Society*, vol. 61, no. 7/8, pp. 535–545, 2013.

[64] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference image quality assessment in reorganized dct domain," *Signal Processing : Image Communication*, vol. 28, no. 8, pp. 884–902, 2013.

[65] M. Liu, K. Gu, G. Zhai, P. Le Callet, and W. Zhang, "Perceptual reduced-reference visual quality assessment for contrast alteration," *IEEE Transactions on Broadcasting*, vol. 63, no. 1, pp. 71–81, 2016.

[66] Y. Zhang, J. Wu, G. Shi, and X. Xie, "Reduced-reference image quality assessment based on entropy differences in dct domain," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2796–2799, IEEE, 2015.

[67] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of dwt coefficients of locally weighted gradient magnitudes," *IEEE Transactions on image processing*, vol. 25, no. 11, pp. 5293–5303, 2016.

[68] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal processing letters*, vol. 17, no. 5, pp. 513–516, 2010.

[69] P. CHAVEL, D. KUAN, A. SAWCHUK, and T. STRAND, "Techniques de réduction de speckle," in *1° Colloque Image : traitement, synthèse, technologies et applications, FRA, 1984*, GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 1984.

[70] J.-H. Seo, S. B. Chon, K.-M. Sung, and I. Choi, "Perceptual objective quality evaluation method for high quality multichannel audio codecs," *Journal of the Audio Engineering Society*, vol. 61, no. 7/8, pp. 535–545, 2013.

[71] C. D. Creusere, K. D. Kallakuri, and R. Vanam, "An objective metric of human subjective audio quality optimized for a wide range of audio fidelities," *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 129–136, 2007.

[72] Z. Akhtar and T. H. Falk, "Audio-visual multimedia quality assessment : A comprehensive survey," *IEEE access*, vol. 5, pp. 21090–21117, 2017.

[73] D. Câmpeanu and A. Câmpeanu, "Peaq–an objective method to assess the perceptual quality of audio compressed files," in *Proceedings of the International Symposium on System Theory, SINTES*, vol. 12, pp. 487–492, 2005.

[74] M. H. Pinson, W. Ingram, and A. Webster, "Audiovisual quality components," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 60–67, 2011.

[75] H. B. Martinez and M. C. Farias, "Full-reference audio-visual video quality metric," *Journal of Electronic Imaging*, vol. 23, no. 6, pp. 1 – 12, 2014.

[76] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio–visual services : A survey," *Signal Processing : Image Communication*, vol. 25, no. 7, pp. 482–501, 2010.

[77] H. B. Martinez and M. C. Farias, "Full-reference audio-visual video quality metric," *Journal of Electronic Imaging*, vol. 23, no. 6, p. 061108, 2014.

[78] M. Hayashi, "The information revolution and the rules of jurisdiction in public," *The resurgence of the state : Trends and processes in cyberspace governance*, p. 59, 2007.

[79] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio–visual services : A survey," *Signal Processing : Image Communication*, vol. 25, no. 7, pp. 482–501, 2010.

[80] P. Gray, M. Hollier, and R. Massara, "Non-intrusive speech-quality assessment using vocal-tract models," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 147, no. 6, pp. 493–501, 2000.

[81] J. G. Beerends, K. v. Nieuwenhuizen, and E. L. Broek, "Quantifying sound quality in loudspeaker reproduction," *Journal of the Audio Engineering Society*, vol. 64, no. 10, pp. 784–799, 2016.