Audiovisuel et la qualité perçu

1.1 Introduction

L'évolution récente des systèmes de communication numérique a conduit à une explosion de services et d'applications multimédias, tels que l'IPTV, le multimédia mobile sur les Smartphones, les réseaux sociaux (par exemple, Face book), la vidéoconférence et les présentations multimédias éducatives, ainsi que l'émergence de plusieurs codec pour audio vidéo et audiovisuelle. Donc ces applications multimédias font désormais partie intégrante (sinon indispensable) de la vie quotidienne et devraient continuer à croître de manière exponentielle.[1]

1.2 Définition audiovisuel

Le signal audiovisuel est doublement composé d'un signal vidéo et d'un signal audio, donc audiovisuel sert à désigner tout ce qui est relatif à l'image et/ou au son. Les fichiers audiovisuels s'agit de toutes les formes d'enregistrement du son et/ou des images animées et/ou des images fixes.[2]

1.3 Les domaines de l'utilisation de l'audiovisuel

1.3.1 Vidéoconférence

En 1968, la vidéoconférence a été introduite pour la premier fois et présentée comme une solution commerciale à l'exposition universelle de New York. La technologie introduite s'appelait le Picture phone d'AT& T. Les participants ont pu s'asseoir et communiquer par vidéo avec la personne de l'autre côté pendant 10 minutes à la fois pour faire l'expérience du premier appareil de visiophone conçu pour les masses. Malheureusement, cette machine particulier était ridiculement chére, maladroite et difficile à installer.[3]

1.3.2 Face time

Face time est une application de chat vidéo développée par Apple. Apple l'a développé sur un standard ouvert, ce qui signifie que techniquement (sans jeu de mots), Face time pourrait être utilisé sur une gamme de plates-formes, et d'autres fabricants peuvent tirer parti du protocole de Face time. Cependant, en réalité, Fac time reste disponible uniquement pour les utilisateurs de produits Apple.[4]

1.3.3 Dans le domaine de l'autorité

La possibilité d'enregistrer des images et du son a naturellement attiré l'attention des autorités militaires.à partir de la fin du XIXe sicle, l'audiovisuel sert à la fois pour les opérations de renseignement militaire et pour la propagande. La Section cinématographique de l'armée (SCA) est créée en 1915.[2]

1.4 Les flux de base de L'audiovisuel

1.4.1 La vidéo

Une vidéo est une succession d'images à une certaine cadence. L'oeil humain a comme caractéristique d'être capable de distinguer environ 20 images par seconde.[4]

1.4.2 Codecs

Un codec est un algorithme de compression / décompression d'un signal audiovisuel numérique, La vidéo ou l'audio brut est compressé lors de l'encodage et décompressé (décodé) lors de la lecture. MP3 est un codec audio - une norme de compression que les lecteurs MP3 savent décoder, et les encodeurs MP3 savent encoder.

1.4.3 codecs Vidéo

1.4.3.1 Format H264 /AVC

La norme de codage vidéo avancée H.264 / MPEG-4 (H.264 / AVC) est la norme de codage vidéo la plus récente développée conjointement par le Groupe d'experts du codage vidéo UIT-T (VCEG) et le groupe d'experts ISO / IEC Moving Picture (MPEG).[6]
H.264, actuellement l'un des codecs vidéo fréquemment utilisés, est une compression populaire pour la vidéo HD. étant donné que H.264 peut atteindre des vidéos de haute qualité dans des débits binaires relativement bas, il est couramment utilisé dans les caméscopes AVCHD, HDTV, Blu-ray et HD DVD. MP4 (.mp4) est l'un des formats vidéo codés H.264.[7]

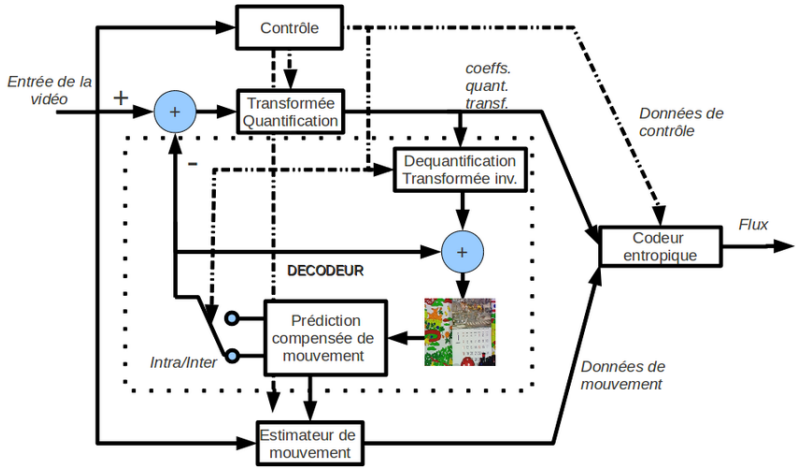

FIGURE 1.1 – Diagramme de bloc de codeur H.264.

1.4.3.2 Format MPEG-4

MPEG-4 utilise des techniques similaires à M-JPEG, en ce qui concerne la mise en séquence. Il compare essentiellement deux images compressées, enregistre l'image et enregistre uniquement la différence à partir de chaque image séquentielle supplémentaire, comme le mouvement, ce qui permet d'économiser du temps, de l'espace mémoire et une puissance de traitement.[8] Un taux de compression plus élevé fait partie des avantages de MPEG-4. Il peut synchroniser l'audio et la vidéo, et est idéal pour la visualisation en temps réel. MPEG-4 a été conçu pour prendre en charge les applications à faible bande passante.[7]

1.4.3.3 AV1(AOMedia Vidéo 1)

AV1 est actuellement candidat à la normalisation par l'Internet Engineering Task Force (IETF) en tant que vidéo Internet Codec (NetVC). La finalisation du processus de normalisation est prévue fin 2018. Il fournit à la fois un décodeur et un encodeur de référence. le codeur logiciel AV1 implémente une optimisation de codage non évidente et des algorithmes non normatifs tels que la quantification adaptative ou la structure de sous-gop dynamique. Il peut optimiser le PSNR ou les critères de qualité perceptuelle. Il prend en charge trois modes de fonctionnement RC : VBR, CBR (Rate Controle (RC) , Constant Bitrate (CBR), Variable Bitrate (VBR)et Constant (ou de façon équivalente-) Qualité.[9]

1.4.4 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.[10]

1.4.5 Codec audio

1.4.5.1 Formats audio compressés sans perte

- **RAW** : RAW est un format audio utilisé pour représenter les données de son en modulation d'impulsion codée sans en-tête ni métadonnées.[2]
- **ALAC** (Apple Lossless Audio Codec) C'est un format de codage sans perte de données, créé en 2004 par Apple.[11]

1.4.5.2 Formats audio compressés avec perte

- **AC3** : La compression AC3 permet d'utiliser jusqu'à 6 canaux sonores indépendants avec un taux échantillonneuse de 32, 44,1 ou 48 kHz et avec un taux de transfert allant de 32 à 640 kbit/s. Le Dolby Digital utilise ce principe de codage, c'est pourquoi on le désigne souvent sous ce nom. Format tr'es courant dans les DVD.[2]
- **MP3** : MP3 est l'abréviation de MPEG-1/2 Audio Layer 3, La couche (Layer) III est la couche la plus complexe. Elle est dédiée 'a des applications nécessitant des débits faibles (128 kbit/s) d'où une adhésion très rapide du monde Internet 'a ce format de compression. [2]
- **MP3PRO** : Le format mp3PRO, fruit de la collaboration entre Thomson Multimédia et l'Institut Fraunhofer, combine l'algorithme MP3 et un système améliorant la qualité des fichiers comprimés appelé (en)SBR pour Spectral Bandwidth Replication.[2]

1.4.5.3 Formats audio non compressés

- **PCM** : La modulation par impulsions et code est une représentation numérique des ondes sonores analogiques. En MIC, l'échantillonnage est utilisé pour convertir les ondes acoustiques sous forme numérique. PCM est instancié par deux capacités principales : la fréquence d'échantillonnage et la profondeur de bits. La fréquence d'échantillonnage permet de mesurer l'amplitude des vagues dans le temps et la profondeur de bits correspond au nombre de bits d'information dans chaque échantillon. PCM est largement utilisé dans la création de CD et de DVD.
- **WAV** : Le format audio Waveform, ou tout simplement WAV, est un format audio brut et non compressé développé par Microsoft et IBM et principalement utilisé sur les plates-formes Windows, WAV a perdu de son attrait en tant que format «de qualité proche de celle d'un CD», mais est toujours très populaire en raison de sa grande disponibilité.

1.5 Logiciel pour lire les audiovisuelle

VideoPad Vidéo editor : est un logiciel d'édition vidéo grâce auquel vous pourrez rapidement vous lancer sur Youtube ou Facebook. Il contient toutes les options nécessaires pour effectuer un montage rapide et efficace dont des effets de transition ou encore un effet de stabilisation. VLC media Player : VLC est un Framework et un lecteur multimédia multiplateforme gratuit et open source qui lit la plupart des fichiers multimédias ainsi que des DVD, CD audio, VCD et divers protocoles de streaming

1.6 La qualité

La qualité est généralement utilisée dans l'optique d'une ingénierie, car elle est un critére essentiel pour évaluer les systèmes, les services ou les applications au cours des phases de conception et d'exploitation. Fondamentalement, la qualité est le résultat d'un jugement humain base sur divers critères. La qualité perçue, et plus largement la QoE, devient un élément clé qu'il faut par conséquente savoir mesurer.[12]

1.6.1 qualité de l'expérience (QoE)

La QoE est une mesure du jugement personnel de l'utilisateur selon son expérience vécue, sur la qualité globale du service fourni par les opérateurs et fournisseurs de services Internet.
En effet, la notion de l'expérience utilisateur a été introduite pour la première fois par le Dr Donald Norman, évoquant l'importance de la conception d'un service centré utilisateur [13]. Gulliver et Ghinea [14] décomposent la QoE en trois composantes : l'assimilation, le jugement et la satisfaction.
La qualité d'assimilation est une mesure de la clarté du contenu d'un point de vue informatif. Le jugement de qualité reflète la qualité de présentation. La satisfaction indique le degré d'appréciation globale de l'utilisateur.

Pour évaluer la qualité audiovisuel il existe essentiellement deux catégories d'évaluation à savoir les méthodes subjectives qui impliquent des observateurs humains pour évaluer la qualité des contenus multimédias et des méthodes objectives qui calculent la qualité automatiquement à l'aide de modèles mathématiques.[2]

1.7 Evaluation de la qualité

1.7.1 Evaluation de la qualité subjective

Afin de mesurer de manière fiable la qualité perceptuelle par les systèmes auditifs et/ou visuels humains, les tests subjectifs sont effectués lorsque des groupes d'observateurs humains formés ou naïfs fournissent des cotes de qualité [15]. Cette procédure d'évaluation est connue comme évaluation de la qualité subjective qui vise à quantifier gamme d'opinions que les utilisateurs expriment quand ils voient entendre le contenu numérique et elle est généralement effectuée dans un environnement bien contrôlé 'a l'aide de recommandations normalisées.
Les modèles subjectifs d'évaluation de la qualité des multimédia les plus performants et les plus utilisés sont normalisés par l'Union Internationale des Télécommunications (UIT). Cet organisme est chargé de la normalisation et de la planification des télécommunications dans le monde. Elle établit les normes de ce secteur et diffuse toutes les informations techniques nécessaires pour permettre l'exploitation des services mondiaux de télécommunications. D'autres modèles sont aussi proposés par des laboratoires universitaires ou encore par des sociétés privées, mais ils ne sont pas validés par les normes de l'UIT.[2]

1.7.2 Evaluation de la qualité objective

Bien que l'évaluation de la qualité subjective fournisse des indices fiables de la qualité de la perception humaine, il ne peut pas être appliqué dans l'évaluation de la qualité en temps réel en service. Ainsi, les méthodes d'évaluation de la qualité objective ont été mis au point pour remplacer le panneau humain par un modèle de calcul pour prédire les résultats d'un test subjectif. A savoir, l'objectif de l'évaluation objective de la qualité est d'estimer automatiquement les valeurs MOS, qui sont aussi proches que possible de scores de qualité obtenus à partir évaluation de la qualité subjective.[16]

1.8 Conclusion

Dans ce chapitre nous avons abordé certaines définitions associées à l'audiovisuel notamment la vidéo et l'audio tout on passe par sa qualité qui contient la qualité subjective et objective et qui seront des points essentiels dans la suite de notre travail. Dans le chapitre suivant nous avons voir les différents métrique d'évaluation de la qualité audiovisuelle (QAV).

Chapitre 2

Evaluation de la qualité audiovisuelle

2.1 Introduction

Il est possible de mesurer la QoE (Quality of Experience) à l'aide de deux indicateurs : les tests subjectifs ou les métriques objectives. L'évaluation de la qualité par les tests subjectifs consiste à demander à un groupe de personnes d'attribuer une note de qualité au service qu'ils utilisent selon leur degré de satisfaction. D'autre part, le processus d'évaluation est automatisé par des algorithmes dans les mesures dites objectives. Les performances des métriques objectives sont mesurés par rapport aux résultats des tests subjectifs. [17] Dans ce chapitre nous avons présente des notions de base sur l'évaluation de la qualité audiovisuel. Cette qualité contient les tests subjectives et les mériques objective qui discuter au long du chapitre.

2.2 Les types d'évaluation de la qualité

Il existe deux types d'évaluation, subjective et objective .[2]

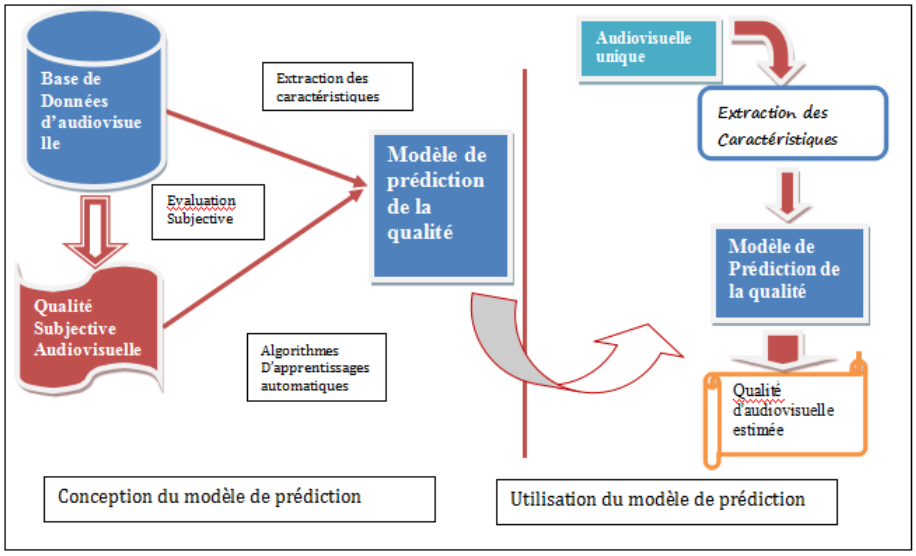

FIGURE 2.1 – Processus d'évaluation objective.[18]

2.2.1 Méthodes subjectives

Le moyen le plus naturel d'évaluer la qualité audiovisuelle est de demander à un groupe d'observateurs humains de noter l'audiovisuel suivant un protocole bien défini; puis d'effectuer la moyenne des notes données par chacun des observateurs : la note moyenne trouvée est appelée à score moyen d'opinion à(MOS : Mean Opinion Score) ou à score moyen d'opinion différentiel à (DMOS : Differential Mean Opinion Score). Cette mé thode est appelée méthode d'évaluation subjective du fait que les résultats qu'elle produit sont fortement impactés par la subjectivit à du jugement humain.
Plusieurs facteurs implicites influençant l'évaluation de la qualité des audiovisuelle par des observateurs humains, peuvent impacter la note finale donnée à l'audiovisuelle. Les facteurs examinés ci-aprés permettent de mieux les distinguer

— **L'écran d'affichage :** : le choix de l'écran d'affichage, de sa taille, de sa calibration, de sa capacité de reproduction des couleurs, de son contraste. Ainsi est-il important lors d'une évaluation subjective de la qualité d'avoir toutes les informations sur les écrans d'affichage utilisés, afin de tenir compte des art éfacts liés aux écrans et à leur calibration.

— **La distance d'observation :** la visibilité de l'audiovisuelle lors de l'évaluation dépend fortement de la distance entre l'observateur et l'écran; il est recommandé de fixer cette distance entre 4 à 6 fois la hauteur de l'image à projeter, pour permettre une visibilité optimale c'est-à-dire ni très proche, ni très éloigné de l'écran. De plus, une fois fixée, cette distance doit être maintenue tout au long de l'expérience.

— **Les conditions de visualisation :** les salles dans lesquelles sont men ées les différentes expériences, l'éclairage, ainsi que les couleurs de fond jouent un rôle important dans l'é valuation de la qualité audiovisuelle. Elles influencent fortement les notes données à audiovisuelle.

— **La durée :** il est préférable de mener des expériences de courte durée pour éviter la fatigue chez l'observateur et partant de fausser les r ésultats. L'Union Internationale des Télécommunications (IT U)recommande de ne pas faire des séances de plus de 30 minutes,afin d'éviter la fatigue chez l'observateur.

— **Les observateurs** idéalement, le choix des observateurs doit d épendre du domaine d'application visé.Aussi, avant de débuter l'expérience, une phase d'apprentissage doit être faite avec les observateurs pour leur expliquer clairement l'objectif du test ainsi que le protocole utilisé, sans pour autant influencer leur jugement.

Il importe aussi d'avoir un nombre suffisant d'observateurs qui participent à l'expérience. Il reste entendu qu'un test n'est statistiquement valide que si l'on a au moins 16 observateurs pour l'expérience.[69][70]
L'évaluation subjective dépend aussi d'autres facteurs tel que l'humeur, l'age, la culture, le niveau intellectuel, ... Les méthodes d'évaluation subjectives ainsi que le protocole D'évaluation de ces méthodes subjectives différent suivant la base de données audiovisuelle considérée. On distingue les protocoles à simple et double stimulus, ainsi que des protocoles à stimulus comparatif.[18]

2.2.1.1 Protocole d'évaluation

Il y a essentiellement trois grandes familles communes d'évaluation subjective dé finies Par l'UIT : échelle continue de la qualité sur stimulus double (DSCQS), é chelle de dégradation sur stimulus double (DSIS) et évaluation continue de la qualit é sur stimulus unique (SSCQE).[2]

1. **Le protocole d'évaluation à simple stimulus :** encore appelé à Single-Stimulus Continuous Quality-Scale à (SSCQS), il permet de juger de la qualité audiovisuelle dégradée par un stimulus unique[18].La méthodologie à stimulus unique est plus utile dans un environnement de test réaliste, comme les tests conversationnels dans lesquels deux sujets à coutent et parlent de maniére interactive via le systéme de transmission en cours d'évaluation pour fournir une qualité.[2]
   Le processus d'évaluation est le suivant : un audiovisuelle s'affiche, puis l'observateur a un temps de latence pour donner une note à cet audiovisuelle en fonction de l'échelle des notes proposée dans l'expérience, et ensuite on passe à l'audiovisuelle suivante. Des échelles à 5, 6, 10 ou même 100 niveaux peuvent être utilisées.[18] ,En générale, on utiliser la méthode suivant :
   - Méthode ACR : ABSOLUTE CATEGORY RATING
   La méthode ACR ou méthode d'évaluation par catégories absolues consiste a attribuer une note de qualité aprés chaque séquence AV visualisé e/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 caté gorielle 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 a bas débit.[20]

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

5	Excellent
4	Bon
3	Satisfaisant
2	Médiocre
1	Mauvais

TABLE 2.1 – Echelle d'évaluation de qualité a 9 et 5 niveaux

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

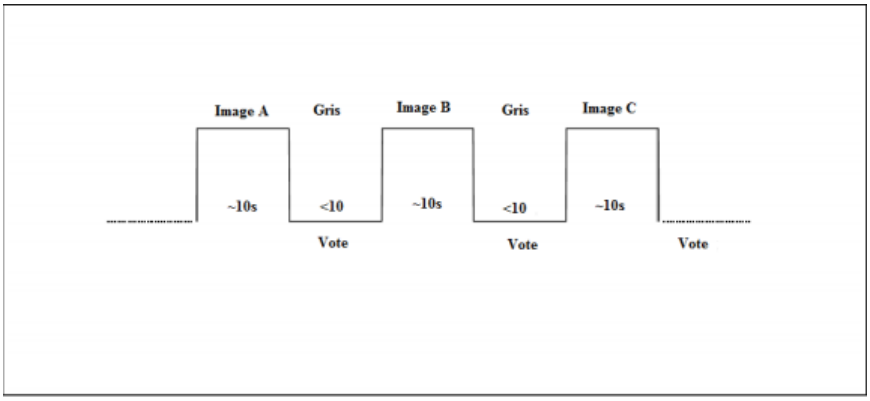

FIGURE 2.2 – Chronogramme de la méthode ACR.[24]

2. **Le protocole d'évaluation à double stimulus :** également appelé à Double Stimulus Continuous Quality-Scale à (DSCQS).[18] Le but principal de la méthode DSCQS est de mesurer la qualité des systémes par rapport à une référence. Les personnes qui sont montrées paires de sé quence audiovisuelle (la séquence de référence et la séquence altérée) dans un ordre aléatoire. Il est largement accepte comme une méthode de test précis avec peu de sensibilité aux effets de contexte, en tant que spectateurs sont présentes deux fois la séquence. Les téléspectateurs sont invites à é valuer la qualité de chaque séquence de la paire aprés la deuxiéme projection. Il est également utilise pour mesurer la qualité du codage audiovisuelle stéréoscopique. [2], la mé thode proposée dans le cadre de cette norme est :
- METHODE DCR : DEGRADATION CATEGORY RATING
La méthode DCR ou méthode par évaluation de catégories de dégradations propose une présentation des séquences AV de test par paires. Les séquences constituant la paire sont identiques a la différence que la premiére est toujours présentée sans dégradations (référence) tandis que la seconde est traitée par le systéme à évaluer (donc susceptible de comporter des dégradations). La séquence traitée est toujours présentée aprés la référence.[24]

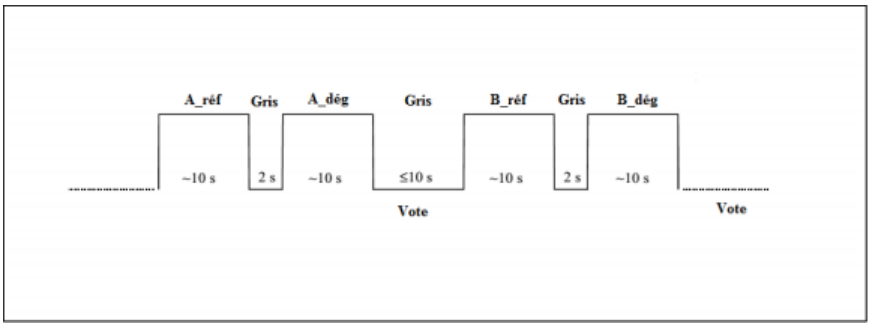

FIGURE 2.3 – Chronogramme de la méthode DCR.[24]

Seule la séquence traitée est évaluée par les participants en comparaison avec la condition de référence. L'échelle d'évaluation correspond ici a une é chelle de perceptibilité de la dégradation comme présente par le tableaux 2.2.

5	Imperceptible
4	Perceptible
3	Peu dégradée
2	Dégradée
1	Trés dégradée

TABLE 2.2 – Echelle de dégradation à cinq niveaux

3. **Le protocole d'évaluation à stimulus comparatif :**les mé thodes comparatives permettent d'évaluer la qualité audiovisuelle en fonction d'une ou plusieurs autres audiovisuelle, venant toutes de la méme audiovisuelle de réfé rence.[18],comme la méthode suivante :
METHODE PC : PAIR COMPARISON
La méthode des comparaisons par paires implique que les séquences d'essai soient pr ésentées en paires. Chaque paire est formé de la même séquence, pré sentée d'abord au moyen d'un systéme l'essai puis au moyen d'un autre systéme. La séquence de référence (sans dégradation) peut être incluse et sera trait é comme un systéme 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 aprés la présentation de chaque paire. Cette méthode est notamment préconise 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 inferieure ou égale à dix secondes :[2]

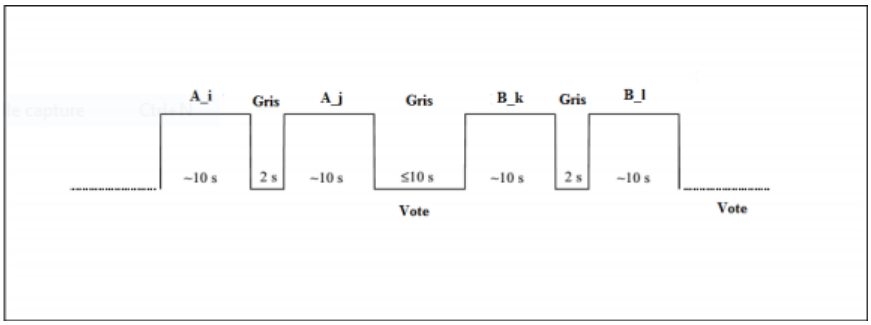

FIGURE 2.4 – Chrono-gramme de la méthode PC.[24]

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

Lors du choix d'une méthode d'essai, un critère important est la différence fondamentale entre méthodes faisant appel à des références explicites (par exemple DCR) et méthodes ne faisant pas appel à des références explicites (par exemple ACR, PC). Cette deuxième classe de méthodes ne contrôle ni la transparence ni la fidélité.
Il convient d'utiliser la méthode DCR lorsque l'on contrôle la fidélité de transmission par rapport au signal de source. Ce facteur présente souvent de l'importance pour l'évaluation de systèmes de haute qualité. D'autres méthodes peuvent être utilisées pour évaluer les systèmes de haute qualité. Les observations spécifiques de l'échelle DCR (dégradation imperceptible/perceptible) sont pré cieuses lorsque la détection d'une dégradation par l'observateur est un facteur important.

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

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

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

2.2.2 Métriques objectives

Bien qu'une évaluation subjective de la qualité indices de qualité de la perception humaine, il ne peut pas être appliqué dans l'évaluation en temps r éel de la qualité en service. Ainsi, des méthodes objectives d'évaluation de la qualité ont été développées pour remplacer le panel humain par un modèle informatique pour prédire les résultats d'un test subjectif. à savoir, le but de l'évaluation objective de la qualité est d'estimer automatiquement les valeurs MOS (mean opinion score), qui sont aussi proches que possible des scores de qualité obtenus à partir de l'évaluation subjective de la qualité. Les mesures numériques de la qualité obtenues à partir de la méthode objective (é galement appelées MOS objectives ou prédites) devraient mieux correspondre à la subjectivité humaine. Il existe différentes mesures pour mesurer la relation entre le MOS subjectif et le MOS prédit. Les deux paramètres statistiques les plus couramment utilisés pour rendre compte des performances des méthodes d'évaluation objective de la qualité sont à l'erreur quadratique moyenne (RMSE)à et la corr élation de Pearson un algorithme d'évaluation de la qualité objective ayant une forte corrélation (généralement supérieure à 0,8) est considéré comme efficace.[23]

Deux principaux avantages d'une évaluation objective de la qualité utilisation dé finissent la signification du MOS pour une application donnée (c.-à-d. que les gens savent ce qu'un MOS de 3 signifie en termes de qualité) et la prédiction reproductible du MOS (c.-à-d.les personnes utilisant l'outil pour les mêmes é chantillons de test obtiennent les mêmes résultats). Les techniques objectives de mesure de la qualité peuvent être classées en cinq groupes, en fonction du type de données d'entre utilisées par les paramétres thématiques :

i. Modles de couche média : les modèles de cette caté gorie ne exiger des informations sur le système en question. En particulier, ces modèles n'utilisent que des échantillons audio ou vidéo pour estimer la qualité et peuvent être appliqués à des applications telles que l'optimisation et la comparaison de codecs.

ii. Modèles de couches de paquets paramétriques : Les solutions pour prédire la qualité dans ce groupe sont légères, car les modéles de couches de paquets paramétriques doivent uniquement traiter les informations d'en-tête de paquet sans traiter avec les médias signaux.

iii. Modèles de planification paramétrique : Ces modèles utilisent encodage et paramétres réseaux pour prédire la qualité. Ils demandent donc une connaissance a priori du système Dans la question.

iv. Modèles de couche binaire : Ces modèles prédisent la qualité à l'aide des informations codées de la couche binaire et de la couche paquet qui sont utilisées dans les modèles paramétriques de la couche paquet.

v. Modéles hybrides : les modèles de cette classe inté grent généralement deux ou plusieurs des modèles mentionnés ci-dessus.

D'autre part, les techniques objectives d'évaluation de la qualit é peuvent également être classées en trois catégories : référence compléte (FR), référence réduite (RF) et sans référence (NR) selon la disponibilité de la référence (originale / idéale), des informations partielles sur la référence, ou aucune référence pour évaluer la qualité, respectivement.[21]

2.2.2.1 Evaluation objective de la qualité audiovisuel :

Les métriques de qualité objective peuvent être classées en trois catégories principales en fonction de la disponibilité du signal de réf érence non déformé : référence complète (FR), référence réduite (RR) et sans

référence (NR). Les métriques FR comparent un signal de réfé rence à un signal déformé afin de calculer la différence de qualité entre les deux. Les algorithmes FR sont généralement les plus précis et relativement simples ce qui contribue à leur utilisation généralisée. Cependant, dans de nombreuses applications réelles (par exemple, vidéoconférence, IPTV, etc.), les modèles FR ne peuvent pas être utilisés car le signal de référence n'est tout simplement pas disponible pour la comparaison. Dans de tels cas, les métriques NR sont g énéralement utilisées. Les méthodes NR sont une mesure absolue des caractéristiques et des caractéristiques d'un signal dégradé et sont souvent axées sur un type de dégradation spécifique (par exemple, flou, bloc) et l'analyse des r églages des paramètres de codage. En raison de l'absence d'un signal de référence, elles peuvent être moins précises que d'autres approches, mais sont plus efficaces à calculer. Les algorithmes RR, au lieu d'une r éférence complète, utilisent des caractéristiques de qualité extraites de la référence et des signaux déformés. Ces caractéristiques sont ensuite comparées afin de générer un seul score de qualité. Les modèles RR sont généralement adoptés dans les cas ou le signal de référence complet ne peut pas être utilisé (par exemple dans une transmission avec une bande passante limitée).

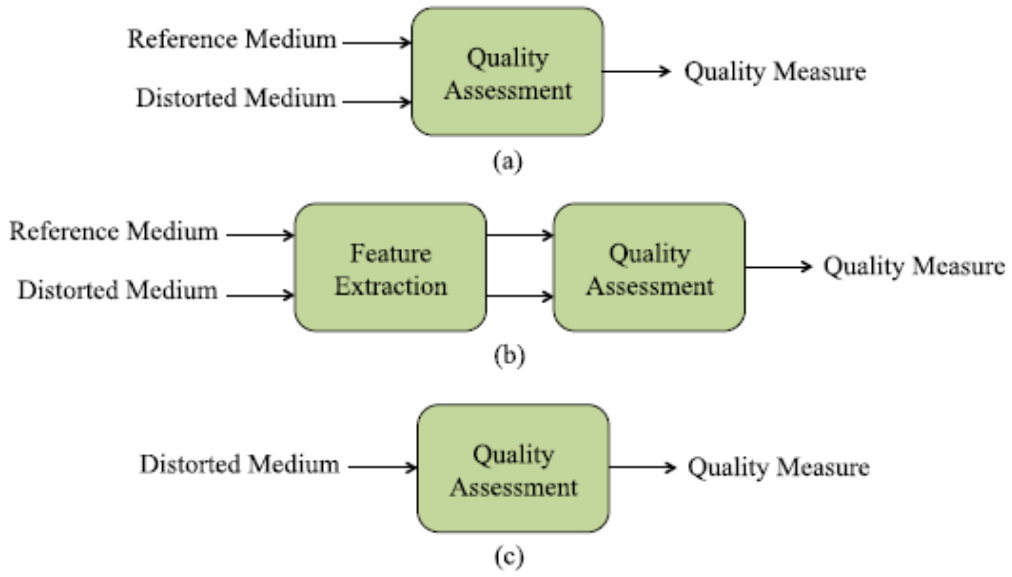

FIGURE 2.5 – aperçu (a) la méthode de référence complète,(b) la m éthode de référence réduite, (c)la méthode de non-référence.[21]

2.2.2.2 L'approche de prévision de la qualité audiovisuelle :

L'UIT-T a proposé certains modèles normalisés de prévision de la qualité audiovisuelle, par exemple, les Recommandations UIT-T P.1201[58], UIT-T G.1070[59] et UIT-TG.1071[60] :

— Le modèle UIT-T P.1201 : C'est un modèle non intrusif d'information en tête de paquet visant la surveillance de service et le benchmarking du streaming UDP. Le modèle prend en charge les applications à basse résolution comme la télévision mobile et les applications à plus haute résolution comme l'IPTV. Il utilise les informations extraites de l'entête du paquet et les informations fournies hors bande. Il fournit des prédictions distinctes de la qualité audio, vidéo et audiovisuelle sous forme de résultat en terme du MOS à 5 points. Le modèle a été validé pour la compression, la perte de paquets et le buffering des altérations de l'audio et de la

vidéo avec des débits différents. Le modèle ITU-T Rec. P.1201 après des testes . Les valeurs de corrélation RMSE et de Pearson pour la modélisation audiovisuelle ont été évaluées respectivement à 0,470 et 0,852 pour les applications à résolution inférieure et à 0,435 et 0,911 pour les applications à plus haute ré solution.

— Le modèle UIT-T G.1070 : ce modèle propose un algorithme a estimé la qualité de l'expérience et la qualité de services. il contient trois fonctions principales d'évaluation de la qualité audio, de la qualité vidéo et de la qualité multimédia globale. La fonction d'estimation de la qualité de la audio prend comme paramètres d'entrée le type de codec vocal, le taux de perte de paquets, le débit binaire et le niveau sonore d'écho de la parole. La fonction vidéo prend comme paramètres d'entrée le format vidéo, la taille d'affichage, le type du codec, le taux de perte de paquets, le débit binaire, l'intervalle d'images et le taux d'images. La fonction multimédia intégre séparément la qualité audio et la qualité vidéo en incluant l'asynchronisme audiovisuel (audiovisual asynchrony) et le délai de bout en bout. Sur des ensembles de données précis, la précision du modèle d'évaluation de la qualité des communications multimédias en terme de corrélation de Pearson est de 0,83 pour QVGA et de 0,91 pour la résolution QQVGA. L'application du modèle est limitée à la planification de la QoE et de la QoS.

— le modèle UIT-T G.1071 :est recommandé pour la planification ré seau des services de diffusion audio et vidéo. Cette recommandation concerne les domaines d'application à plus haute résolution (HR) comme l'IPTV et les domaines d'application de résolution inférieure (LR) comme la TV mobile. L'application des modèles est limitée à la planification de la qualité d'expé rience (QoE)/qualité de service (QoS). Le modèle prend en entrée les hypothèses de planification de réseau comme la résolution vidéo, les types et profils de codecs audio et vidéo, les débits audio et vidéo et le taux de perte de paquets. Il fournit en sortie des prédictions distinctes de la qualité audio, vidéo et audiovisuelle définies sur l'échelle MOS à 5 points. Notre note que les tests ont montré que les applications à basse résolution utilisant les bases de données d'apprentissage et le test ITU-T P.1201.1. Les valeurs de corrélation RMSE et de Pearson pour la modélisation audiovisuelle ont été évaluées respectivement à 0.5 et 0.83 Pour les applications à haute ré solution et 0,51 et 0,87 pour des bases de données d'apprentissage et de validation ITU-T P.1201.2.[2]

2.2.2.3 L'approche d'apprentissage automatique :

Le monde de l'apprentissage automatique consiste un nombre incalculable d'algorithmes ainsi que de leurs implémentations dans diverses bibliothèques. Certaines de ces méthodes sont destinées uniquement aux problèmes de classification. Cependant, plusieurs algorithmes sont adaptés à des problèmes de classification et de régression. Ce sont quelques-uns méthodes d'apprentissage automatique

- Méthodes d'ensemble basées sur l'arbre de décision : Les arbres de décision (DecisionTrees, DT) sont des structures de données hiérarchiques utilise pour des problèmes de classification et de régression par stratégie de **Diviser-et-conquérir** (divide-and-conquer). Pour évaluer la QAV demribilek a généré deux modèles basés sur les forêts d'arbres décisionnels et deux modèles basés sur les techniques de bootstrap. Ces modèles utilisent comme caractéristiques soit les paramètres indépendants (5 caractéristiques), soit tous les paramètres extraits (34 caractéristiques). Et par la suite il a comparé les résultats obtenus. Pour des besoins de simplification, il nomme le modèle de forêt d'arbres décisionnels qui utilisent tous les paramètres (paramètres indépendants et supplémentaires) par le modèle RF1. Le modèle RF2 va référer au modèle de forêt d'arbres décisionnels qui utilise uniquement les paramètres indépendants. Avec la même logique, il appellera le modèle basé sur les techniques de bootstrap et utilisant tous les paramètres (indépendants et supplémentaires) le modèle BG1. Le modèle basé également sur les techniques de bootstrap et qui utilise uniquement les paramètres indépendants sera nommé par le modèle BG2.[61]

- Régression symbolique et programmation génétique : La programmation génétique est une technique de calcul, permet de trouver une solution á un problème sans connaitre la forme de la solution. baser par l'évolution d'une population de programmes informatiques ou les populations sont transformées aléatoirement á nouvelles populations génération par génération. Pour évaluer la QAV et découvrir le meilleur modèle demirbilek a généré deux modèles comme Apprentissage profond, un modèle utilise uniquement les variables indépendantes et un modèle utilise tout les variables et après il a comparé les résultats.[61] .

- Apprentissage profond :
L'apprentissage profond remonte aux années 1940, reflétant l'influence de différents chercheurs et de différentes perspectives. Cette appellation spécifique est très récente, tells que typique d'un modèle d'apprentissage profond est **feed forward Deep Network** ou le perceptron multicouche (Multi-Layer perceptron, MLP) [62]. Dans cette apprentissage demirbilek a généré deux modèles, un modèle utilise uniquement les variables indépendantes et un modèle utilise tout les variables et après il a comparé les résultats, qui découvrira que le modèle utilise uniquement les variables indépendantes ont obtenu de meilleurs résultats que le modèle utilise toutes les variables.

2.2.2.4 L'approche de la fusion des deux modalités :

Les études empiriques montrent que les domaines auditif et visuel ont une influence mutuelle 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 [61].
Les modèles de calcul susmentionnés prédisent automatiquement la qualité perceptuelle à l'aide d'opérations mathématiques. De telles opérations sont souvent faites à l'usage, d'un modèle du système visuel humain (HVS) et du système auditif.
La plupart des mesures de qualitè objectives existantes se concentrent uniquement sur une seule modalité, audio ou vidéo, et ne tiennent pas compte de la forte influence mutuelle des deux dans le processus d'évaluation de la qualité. En cas de stimulus multimodales, notre cerveau utilise plusieurs sources d'informations sensorielles drivés de plusieurs modalité différentes.
La perception multimodale n'est pas une simple combinaison linéare de perceptions de modalités uniques. La plupart des recherches indiquent qu'à un moment donné du traitement perceptuelle, toutes ces différentes sources d'information s'intègrent pour former un percept cohérent et robuste (fusion perceptuelle). Au cours de ce processus, une modalité peut modifier et compléter la perception dérivée d'une autre modalité. En cas d'évaluation de la qualité multimodale, un tel effet transmodal de l'intégration multi sensorielle peut fortement influencer (positivement ou négativement) la QoE.
La qualité audiovisuelle est donc décrite comme une fusion de deux dimensions (qualités audio et vidéo), comme illustré à la Figure 2.6 :

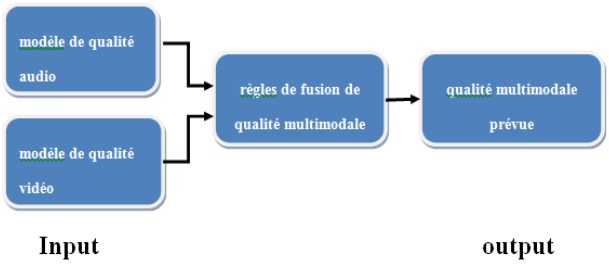

FIGURE 2.6 – Modèle d'estimation de la qualité multimédia de base

Le modèle de fusion le plus couramment utilisé et adopté dans plusieurs études est celui présenté dans l'équation suivante [61][[62][63] :

$$QAV = \alpha + \beta QA + \gamma QV + \delta QAQV \tag{2.1}$$

Ou QAV, QA, QV et α, β, γ, δ sont respectivement la qualité audiovisuelle, la qualité audio, la qualité vidéo et les poids prédits. Les valeurs rapportées dans la littérature vont de α = [- 3.34, 4.26], β = [- 0.19, 0.85], γ= [0, 0.89], δ = [- 0.01, 0.26]. Peu d'études suggèrent que les canaux audio et vidéo pourraient être int égrés dans une phase précoce de la formation de la perception humaine. Sur cette base, plusieurs chercheurs [61], ont proposé des modèles de Qualité audiovisuelle comme une multiplication de qualité audio et vidéo d'égale importance, comme le montre l'équation suivante :

$$QAV = \alpha + \beta QAQV \tag{2.2}$$

De même, Martineza et al [64], ont proposé trois mesures de la qualité perçue audiovisuelle. Le premier modèle est un modèle linéaire simple donné par l'é quation suivante :

$$QAV = \alpha + \beta QA + \gamma QV \tag{2.3}$$

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

$$QAV = ((\beta QA)P + (\gamma QV)P)1/P \tag{2.4}$$

Oú les valeurs de la puissance de Minkowski (P) sont toutes comprises entre 1 et 1,2. Sur la base de ces résultats, nous avons varié la valeur de P dans la plage de 0,9 à 1,3 et répété la procédure d'ajustement pour chacune de ces valeurs. Comme certaines études, suggèrent que la modalité visuelle peut ê tre plus dominante que l'audio dans la formation de la qualité audiovisuelle perçue, en particulier pour les vidéos avec des données de mouvement élevées, ainsi les auteurs dans pr ésentent ainsi l'équation suivante :

$$QAV = \alpha + \beta QV + \gamma QAQV \tag{2.5}$$

malgré les modèles dans les équations prétendants atteint assez précise la qualité audiovisuelle prévue dans certaines études lorsque des durées de qualité audio et vidéo sont les mêmes, il ne reflète pas les différences de l'influence de seulement audio et vidéo uniquement des stimuli sur la qualité globale.

2.2.3 Performance des modèles d'évaluation objective de la qualit é audiovisuelle

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 pr écise qu'un sujet de test moyen. L'incertitude des votes subjectifs est calculée par l'é cart-type et l'intervalle de confiance (IC) correspondant. Ces paramètres statistiques visent à dé terminer l'incertitude des sujets par fichier, ou par condition de test[74]. La performance d'un modèle est évaluée via trois métriques statistiques, utilisées pour informer de la pré cision du modèle, de sa consistance et de sa linéarité/monotonie[74][75] :

- **précision :**saisit la capacité du modèle à prédire les évaluations de qualité subjectives avec de faibles erreurs.
- **consistance :**reflète le degré auquel le modèle maintient l'exactitude des prévisions sur la plage des séquences de test.
- **la monotonie :**correspond au degré auquel les prédictions du modèle conviennent avec l'ampleur relative

des évaluations subjectives de la qualité.

Lorsque les données sont tirées de données de test avec une distribution proche de la normale, ces critéres sont obtenus en calculant respectivement l'erreur de pré diction, le rapport de valeurs aberrantes (outlier ratio) et le coefficient de corrélation de Pearson[75]. 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 litté rature au lieu du coefficient de corrélation de Pearson comme mesure de la monotonie [74][75].
il est recommandé d'utiliser l'erreur de prédiction pour la pré cision, le rapport de valeurs aberrantes (OR) ou la distribution d'erreur résiduelle pour la cohé rence et le coefficient de corrélation de Pearson pour la linéarité[74].

2.2.3.1 Exactitude du modèle (précision) :

L'erreur de prédiction (c'est-à-dire l'exactitude) est obtenue à l'aide de l'erreur quadratique 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 é chelle 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 [74].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Actual_i)^2}{N}} \tag{2.6}$$

Ou i est l'index de la séquence, et N est le nombre de séquences utilis ées 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%.
Le rmse est approximativement caractérisé par un $\chi 2(n)$, ou n représente les degrés de liberté et est défini par l'équation : n=N - d, oú d = 4 indique les degr és de liberté de la fonction de cartographie (fonction polynomiale de 3 e ordre). En utilisant la distribution $\chi 2(n)$, l'intervalle de confiance de 95 % pour la rmse est donné par l'équation[74] :

$$\frac{(rmse\sqrt{N-1})}{\sqrt{\chi^4_{0.975}(N-d)}} < rmse < \frac{(rmse\sqrt{N-1})}{\sqrt{\chi^4_{0.025}(N-d)}} \tag{2.7}$$

Cette mesure de rmse dépend de l'échelle de notation utilisée lors des tests subjectifs. Par conséquent, pour comparer deux valeurs rmse, les scores de qualité doivent d'abord être convertis à la même échelle. rmse est toujours positif, et des valeurs rmse plus faibles indiquent une plus grande précision. Pour tenir compte du degré d'incertitude des jugements des sujets, les valeurs de qualité dites epsilon-modifiées Root-Mean-Square-Error (rmse*) entre les valeurs pr édites et subjectives peuvent être calculées à la place de rmse. Rmse* est similaire à rmse, mais avec :

$$Perror(i) = max(0.|Mos(i) - Mosp(i)| - ci95(i)) \tag{2.8}$$

Oú ci95 est l'intervalle de confiance à 95 % de la séquence i.
Cette métrique pour comparer les performances des modèles sur la base de bases de données de tests subjectifs avec des intervalles de confiance très variables [75] pour l'évaluation de la qualité à référence complète.

2.2.3.2 Consistance du modèle (cohérence et) :

La consistance du modèle est obtenue en calculant soit le rapport des valeurs aberrantes (Outlier Ratio OR), soit la distribution des erreurs résiduelles [74][75].

$$OR = \frac{TotalNoOutliers}{N} \tag{2.9}$$

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.

$$|Perror(i)| > \frac{z\sigma(MOS(i))}{\sqrt{Nsup}} \tag{2.10}$$

$$\sigma(MOS(i)) = \sqrt{\frac{MOS(i)(1 - MOS(i))}{N}} \tag{2.11}$$

Ou $\sigma(MOS(i))$ 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 d épend du Nsubj, respectivement le degré de liberté df=Nsubj -1[74][75].

2.2.3.3 Modèle monotonie (linéarité) :

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 corré lation 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[74][75].

$$R = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(\sum_{i=1}^{n}(X_i - \bar{X})) * (\sum_{i=1}^{n}(Y_i - \bar{Y}))} \tag{2.12}$$

Xi indique le score subjectif MOS et Yi le score objectif (MOSp). N repré sente le nombre total d'échantillons pris en compte dans l'analyse.
Le coefficient de corrélation de Spearman est défini comme suit [74][75] :

$$R_s = \frac{\sum_{i=1}^{n}(RO(i) - \bar{RO})(RO_e(i) - \bar{RO_e})}{(\sum_{i=1}^{n}(RO_i - \bar{RO})^2) * (\sum_{i=1}^{n}(RO_e - \bar{RO_e})^2)} \tag{2.13}$$

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 pr édites (ROe(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 [74][75] :

$$z = 0.5ln(\frac{1+R}{1-R}) \tag{2.14}$$

$$\sigma_z = \sqrt{\frac{1}{N-3}} \tag{2.15}$$

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 donn ée par l'équation [74][75] :

$$z \pm 1.96 * \sigma_z \tag{2.16}$$

2.3 Conclusion

Dans ce chapitre nous avons présenté les méthode d'évaluation de la qualité audiovisuel qui se divise aux deux familles importantes : les méthodes subjectives qui se divisent sur trois protocoles d'évaluation : protocole d'évaluation simple stimulus qui contient la méthode ACR, protocole d'évaluation double stimulus qui contient la méthode DCR, protocole d'évaluation stimulus comparatif qui contient la méthode PC et les méthodes objectives basée trois approches : L'approche de prévision de la qualité audiovisuelle, L'approche d'apprentissage automatique, L'approche de la fusion des deux modalités .