

Vers une nouvelle approche de l'évaluation de la qualité audio multicanal

A. SMIMITE¹⁻², A. BEGHDAI¹ et K. CHEN¹

¹Université Paris 13, Sorbonne Paris Cité, Laboratoire de Traitement et Transport de l'Information
(L2TI, EA 3043), F-93430, Villetaneuse, France

beghdadi, ken.chen@univ-paris13.fr

²Digital Media Solutions (DMS), 45, Grande Allée du 12 Février 1934, 77186, Noisiel, France

a.smimite@dms-cinema.com

Résumé

Cette étude a pour objectif de mettre en place une méthodologie d'évaluation de la qualité du son multicanal. Ce travail est motivé principalement par deux aspects : le premier porte sur le fait que les techniques existantes, subjectives ou objectives, ne permettent pas d'analyser ou du moins pas efficacement, l'information spatiale des flux audio multicanaux. Le deuxième est que les résultats de ces techniques ne reflètent pas entièrement la réalité en n'intégrant pas totalement certains traits comportementaux. Ces derniers font qu'il y a un écart considérable entre différentes réalisations des tests subjectifs d'un côté et entre les tests menés en laboratoire et en conditions réelles d'un autre. Dans l'optique de proposer une approche aussi réaliste et complète que possible, et à travers l'étude des méthodes existantes, on présentera leurs lacunes au niveau des aspects soulignés et nos suggestions pour y remédier.

Mots clefs

Evaluation, Qualité, Audio, Objective, Subjective, Multicanal, Perceptif, Spatial, Psychoacoustique

1 Introduction

Au cours de ces dernières années, on a assisté à une remarquable révolution dans le domaine de l'audiovisuel. Des contenus de plus en plus riches ont fait leur apparition, plus réalistes, interactifs et surtout accessibles de partout et sur différents types de support. Dans ce sens, les deux dernières décennies ont vu l'émergence de nouvelles techniques de production et de restitution sonore avancées dans ce qu'on appelle *Audio 3D*, où on cherche à créer une sensation d'immersion dans la scène chez les auditeurs avec des sources sonores soigneusement spatialisées. La notion de qualité a aussi évolué. Alors que cette dernière a longtemps été délaissée au niveau de l'audio principalement pour des contraintes de bande passante, de nombreuses communautés scientifiques et industrielles commencent à s'y intéresser de nouveau. Il s'agit en effet de l'un des aspects critiques depuis la conception et jusqu'à

la commercialisation des équipements audio.

Généralement parlant, les méthodes d'évaluation de la qualité audio peuvent être classées en deux catégories : Subjectives où on fait appel à un panel d'auditeurs, experts ou non, selon l'application considérée. L'indice de qualité globale du système évalué est obtenue ensuite par l'analyse statistique (moyenne) des notes attribuées par les sujets. Bien qu'il s'agit de l'approche la plus intuitive et efficace si elle est correctement menée, on s'aperçoit rapidement de la difficulté qu'elle présente en termes de temps et de ressources. Et puis les méthodes Objectives qui interviennent à ce niveau en proposant de simuler la réponse humaine à travers des algorithmes et des mesures spécifiques. Il s'agit de méthodes rapides et moins coûteuses mais leur efficacité reste définie par leur degré de corrélation avec l'évaluation subjective. Un aspect qui est malheureusement souvent oublié.

En effet, durant nos tests subjectifs, on a constaté certains aspects chez les sujets qui risquent éventuellement de fausser les résultats, s'ils ne sont pas correctement considérés. On cite à titre d'exemple l'effet d'expérience qui se manifeste par le fait que les sujets ont tendance à être influencés par des séquences de mauvaise qualité, en attribuant une note basse aux celles écoutées après une dégradation, même si elles sont de qualité supérieure. Au lieu d'essayer de contourner cet aspect naturel, on propose donc de le modéliser et de l'intégrer dans les approches subjectives et objectives. Notre étude est aussi motivée par le fait que les méthodes actuelles, objectives ou subjectives, et bien qu'elles mentionnent parfois le son multicanal ou multivoies, n'ont pas été conçues pour analyser l'information spatiale. On propose donc une extension des standards existants à travers l'intégration d'un modèle spatial dédié.

Cet article sera organisé comme suit : A travers la première partie, on présente brièvement l'état de l'art des techniques d'évaluation subjectives et objectives de la qualité audio pour souligner les aspects qu'on estime critiques mais pas encore pris en charge. La seconde partie présente le concept de la solution proposée pour

remédier à ces limitations, principalement au niveau des aspects spatiaux et psychologiques. On expose ensuite les résultats de nos premiers tests subjectifs pour analyser l'effet d'expérience. Enfin, la conclusion décrit l'état d'avancement de nos travaux ainsi que nos perspectives d'amélioration.

2 Etat de l'art

Il existe actuellement de différentes techniques pour l'évaluation de la qualité audio selon l'application considérée. On présente essentiellement les recommandations de l'ITU (International Telecommunication Union) qui se sont démarquées comme normes au cours de la dernière décennie, ainsi que les travaux qui ont porté sur leur amélioration.

2.1 Subjective

Les méthodes subjectives consistent à effectuer des tests formels avec un panel d'auditeurs soigneusement choisis en fonction de l'application considérée. Il s'agit d'une solution intuitive et naturelle mais qui nécessite des ressources conséquentes, particulièrement en termes de temps. La recommandation ITU-R BS.1284-1 [1], présente les principales étapes à suivre lors d'un test et renvoie sur les méthodes suivantes selon le niveau de dégradation :

- Rec BS.1116-1 connu aussi sous le nom du test ABC [2]. Cette méthode est particulièrement adaptée aux dégradations de faible niveau et basée sur un principe de « doublement aveugle à triple stimulus et référence dissimulée ». La référence connue est toujours présentée comme stimulus "A". La référence dissimulée et l'objet de l'essai sont disponibles simultanément mais sont attribués au hasard aux stimulus "B" et "C". Le sujet note ensuite les séquences de 1 à 5 en fonction de la gêne induite par la dégradation.
- Rec BS.1534-1 connu aussi sous le nom du test *MUSHRA* (Multiple Stimuli with Hidden Reference) [3]. Elle est plus adaptée à des dégradations intermédiaires et consiste à effectuer des tests en double aveugle à stimuli multiples avec une référence cachée et un ou plusieurs repère(s) caché(s). On présente au sujet le signal de référence ainsi qu'un ensemble de signaux dégradés et l'utilisateur peut commuter instantanément et en permanence entre les différents signaux pour les noter. Les sujets attribuent ensuite une note sur une échelle de 0 à 100 présentée en cinq niveaux.

Le premier aspect à souligner est que ces standards recommandent principalement un panel d'experts. Il se trouve qu'il n'y ait aucun moyen objectif pour déterminer le niveau d'expertise audio des sujets (tests spécifiques) et même si c'était possible, il serait toujours plus réaliste de recourir à des non-experts pour des résultats plus représentatifs, ce qui correspond à la majorité des auditeurs.

Une des remarques principales qu'on peut noter est que les méthodes ITU dépendent de l'expression de la gêne occasionnée par les dégradations. Des études comme celle mentionnée en [4] ainsi que nos tests montrent que ce facteur est fortement personnel et difficilement exprimable. Ceci risque d'introduire des biais considérables et donc des résultats erronés.

On a souligné aussi l'absence d'une réelle considération de l'effet mémoire dans ces tests. En effet, pour un test ABC par exemple où on aura la suite Ref-Deg-Ref, toujours en mode caché, les auditeurs auront tendance à baisser la note du signal de référence la deuxième fois sous l'influence du signal dégradé. L'alternance aléatoire permet de réduire cet effet mais pas de l'éliminer. Le même constat a été mis en évidence par A. Bovik et al. dans le cadre de l'évaluation de la qualité vidéo, où ils notent la présence d'un effet hystérésis à prendre en considération [5].

L'analyse de l'information spatiale quant à elle est quasi-absente dans les protocoles subjectifs. En effet, la plupart des tests sont conçues essentiellement pour évaluer le niveau de gêne introduit par le traitement audio du système étudié. Il existe cependant et à notre connaissance, certaines études qui ont porté sur l'analyse de la dégradation de spatialisation mais avec des outils mal adaptés. En effet, elles utilisent pour cela des feuilles de réponses classiques sur lesquelles les sujets sont invités à marquer la position des sources sonores avant et après la dégradation. On note l'inefficacité de cette approche lors des tests (passage du 3D au 2D, échelle, ...) ainsi que la difficulté qu'elle présente pour traiter les résultats [6].

2.2 Objective

Les méthodes objectives d'évaluation de la qualité audio sont basées sur des mesures et algorithmes spécifiques inspirés par le système d'écoute humaine et qui permettent de le simuler. Leur fiabilité est donc définie par leur niveau de fidélité de simulation, ce qui se ramène à leur niveau de corrélation avec des tests subjectifs appropriés. Il s'agit d'une solution rapide et efficace et parfois la seule envisageable pour des applications nécessitant une évaluation en temps réel.

Actuellement, le seul standard en vigueur est la méthode PEAQ (Perceptual Evaluation of Audio Quality), connu aussi sous la référence Rec. ITU-R BS.1387-1 [7]. Cette recommandation constitue la synthèse d'un ensemble de techniques qui existaient avant et dont le principe est illustrée sur la figure 1 :

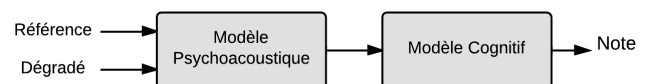


Figure 1 – Principe de la PEAQ

Il s'agit d'une méthode dite avec référence où on compare un signal de référence à sa version

potentiellement dégradée récupérée à la sortie du système évalué. Le modèle psychoacoustique transforme ensuite les deux signaux en une représentation équivalente à celle retrouvée au niveau de la membrane basilaire. Il fournit par la suite un ensemble de variables de sortie qu'on combine au niveau du modèle cognitif (réseau neuronal artificiel avec une couche cachée) pour obtenir une note globale de qualité.

La PEAQ existe en deux versions : une version basique adaptée aux applications où la vitesse et le coût du calcul sont critiques, et une version avancée, plus précise, mais quatre fois plus gourmande en termes de ressources de calcul [8]. La différence réside essentiellement au niveau du modèle psychoacoustique constitué d'une FFT pour la basique et d'une FFT plus un banc de filtres pour l'avancée. Les grandeurs analysées (MOV : Model Output Variable) par la PEAQ sont présentées ci-dessous :

Tableau 1 – MOV de la PEAQ

MOV	Signification
$WinModDiff1_B$, $AvgModDiff1_B$, $AvgModDiff2_B$	Changements de modulation (lié à la rugosité)
$RmsNoiseLoud_B$	Intensité de la distorsion
$BandwidthRef_B$, $BandwidthTest_B$	Distorsions linéaires (réponse en fréquence, etc.)
$RelDistFrames_B$, $SegmentalNMR_B$	Fréquence des distorsions audibles
$TotalNMR_B$	Rapport bruit/masque
$MFPD$, ADB_B	Probabilité de détection
EHS_B	Structure harmonique de l'erreur
$RmsNoiseLoudAsym_A$	Intensité de la distorsion
$AvgLinDist_A$	Distorsions linéaires (réponse en fréquence, etc.)

Depuis sa normalisation, différents travaux ont été menés pour améliorer les performances de la PEAQ. Le modèle psychoacoustique qui reprend le fonctionnement du système auditif humain a été révisé pour mieux le modéliser. Dans ce sens, Cave propose en [9] son propre modèle afin de remédier à l'inexactitude du modèle de masquage temporel de la méthode ITU. Il souligne notamment que le résultat de normalisation doit fournir un niveau SPL qui correspond à celui perçu au niveau du tympan, indépendamment de la résolution fréquentielle, et contrairement à ce qui se fait dans la méthode PEAQ, où on normalise par rapport à une fréquence unique. Sa solution s'est avérée efficace mais seulement pour les systèmes de codage, ne couvrant donc pas ainsi la totalité

des applications [8]. Hubert propose aussi dans sa méthode PEMO-Q-A [10] un nouveau modèle auditif qui offre une meilleure précision et gère un plus grand nombre de distorsions. Ceci en contre-partie de la complexité calculatoire qui est encore plus importante que la version avancée de la PEAQ (déjà plus lente que la version basique) [8].

Le modèle cognitif a aussi fait l'objet d'études en vue d'améliorer ses performances. Vanam et al. [11] propose un nouveau paramètre MOV résultant de leur algorithme d'égalisation d'énergie (EEA). Ce dernier est basé sur le constat que la qualité perçue de l'audio dépend de la présence de régions démarquées en temps-fréquence, principalement dans la bande 2-4 KHz, dans ce qu'ils appellent «îlots» (islands). Leur étude montre que l'intégration du nombre de ces îlots dans le modèle cognitif offre un meilleur facteur de corrélation, en particulier pour les codecs à faible débit binaire.

Suite à l'intérêt croissant que suscite le son 3D, un groupe d'étude de l'ITU (Study Group 6) a commencé à travailler depuis 2008, sur l'extension de la PEAQ pour prendre en compte le son multicanal. Des travaux comme ceux de Choi et al. [12] ont été menés dans ce sens, où ils proposent de combiner les indices interauraux avec les MOV de la PEAQ afin de la rendre compatible avec le son multicanal. Bien que leur étude soit incomplète, elle a montré des résultats encourageants.

3 Méthode proposée

Notre approche consiste à proposer une méthode aussi réaliste que possible, qui prend en compte les contraintes des nouvelles applications, en particulier le son 3D. Pour cela, on propose quelques pistes d'amélioration des protocoles subjectifs en intégrant certaines considérations d'aspects comportementaux, et puis une méthode objective globale qui prend en compte l'information spatiale.

3.1 Approche Subjective

Comme mentionné précédemment, les protocoles subjectifs ne gèrent pas certains aspects critiques susceptibles d'introduire des biais considérables, voire même incompatibles avec les nouvelles applications comme l'audio 3D. Nous nous limitons dans le cadre de cet article à l'intégration de l'effet d'expérience et l'analyse de l'information spatiale.

On rappelle que l'effet d'expérience se traduit par une chute «perçue» de la qualité suite à une dégradation, et qui a tendance à affecter le jugement des auditeurs pour une certaine durée. On propose le test ABC avec les séquences Ref-Deg-Ref pour le mettre en évidence.

Pour une dégradation de type bruit blanc additif centré (AWGN induisant un SNR de 36,5 dB), on obtient les résultats de la figure 2, où on peut constater que le niveau perçue de la qualité est plus bas que la référence (courbe de contrôle) :

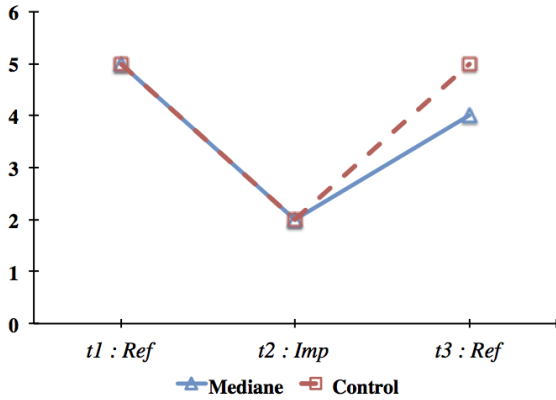


Figure 2 – Effet d'expérience

Les protocoles ITU ainsi que les travaux menés par Zielinski et al. dans [4] proposent des solutions pour réduire l'impact de cet effet : la randomisation, limiter la longueur des séquences, pause entre séquences, repères cachés, mapping des résultats, etc. Même en les intégrant, l'effet reste toujours présent, étant donné qu'il relève d'une réponse naturelle et inconsciente. En effet, cela renvoie aux travaux de H. Ebbinghaus dans ce qu'il appelle l'effet de « récence » (Recency effect). Il a souligné que pour un processus d'apprentissage, les sujets auront tendance à se souvenir d'avantage des séquences récentes et leur accorder plus d'importance que les premières.

On propose donc pour notre approche de modéliser cet effet et le compenser par une fonction de pondération comme sera expliqué dans la section 4.

L'analyse de l'information spatiale nécessite une approche différente que celle des protocoles classiques. On propose pour cela deux solutions qui sont en cours de réalisation. La première consiste à étendre les protocoles de tests grâce à une application 3D sur un dispositif de type tablette, où les sujets peuvent indiquer la position de la source sonore sur une sphère dont ils seront le centre. Dans un deuxième temps, et pour une approche plus naturelle, on propose un système de tracking à distance capable de suivre les gestes d'un bras pour estimer ensuite l'azimuth et l'élévation de la source sonore. Les résultats de cette solution feront l'objet d'un article à venir.

3.2 Approche Objective

Pour répondre aux différents besoins soulevés, on propose un concept d'évaluation avec référence comme illustré sur la figure 3 :

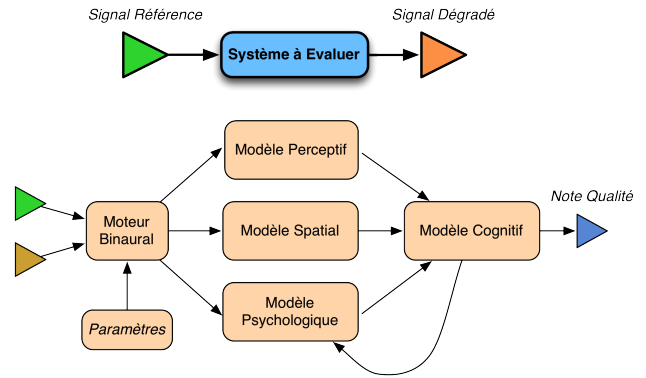


Figure 3 – Méthode proposée

Afin de rendre la PEAQ compatible avec le son multicanal, on propose d'utiliser un système de binauralisation comme suggéré par [12].

En effet, le format binaural reprend le principe physique et naturel par lequel le cerveau humain perçoit le son spatialisé. Une source sonore placée dans l'espace est captée par les oreilles gauche et droite respectivement aux instants t_l et t_r à des niveaux L_l et L_r comme illustré sur la figure 4. Une position dans l'espace correspond donc à une différence interaurale de phase (ITD) et de niveau (ILD) comme le montre les équations 2 et 3. Il s'agit de la théorie du Duplex mise en évidence par Lord Rayleigh en 1907.

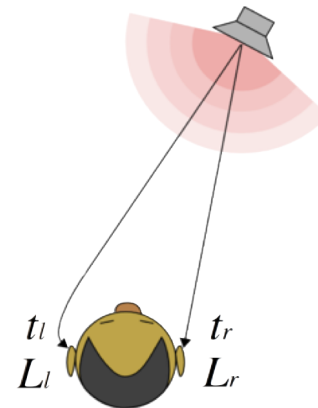


Figure 4 – Principe du Binaural

$$ITD = t_l - t_r \quad (1)$$

$$ILD = L_l - L_r \quad (2)$$

Le signal $x(t)$ de la source sonore est perçue donc au niveau de l'oreille gauche et droite autant que $x_l(t)$ et $x_r(t)$ comme le montre les équations 4 et 5, où $h_l(t)$ et $h_r(t)$ correspondent aux réponses impulsionnelles dépendant de la position de la source et la forme de la tête de l'auditeur (HRIR : Head Related Impulse Response). Ces derniers représentent la fonction de transfert (HRTF : Head Related

Transfer Function) qui relie le signal d'une source sonore à une position donnée à ses représentations au niveau des oreilles de l'auditeur.

$$x_l(t) = x(t) * h_l(t) \quad (3)$$

$$x_r(t) = x(t) * h_r(t) \quad (4)$$

Notre approche consiste à faire passer d'abord le signal de référence ainsi que sa version dégradée par notre moteur binaural. Ce dernier permet d'obtenir un flux stéréo (2 canaux) tout en gardant l'information spatiale de chaque canal. En effectuant la convolution du signal audio de chaque canal avec la HRIR correspondant à sa position dans l'espace, et en normalisant la somme de ces signaux, on obtient un flux binaural équivalent à celui perçu au niveau des oreilles.

Le modèle spatial calcule ensuite trois indices interauraux : ITD (Interaural Time Difference), ILD (Interaural Level Difference) et IACC (Interaural Cross-Correlation Coefficient) pour estimer la dégradation de l'information spatiale.

Le modèle psychologique quant à lui, permet d'intégrer l'effet d'expérience discuté précédemment en compensant par la fonction de pondération présentée dans la section 4.

Le modèle perceptif reprend le coeur de la PEAQ avec les extensions de Hubert et Cave [9, 10]. Un travail d'optimisation d'intégration permet de réduire la complexité calculatoire notamment par l'utilisation de la FFTW.

Le modèle cognitif est un réseau de neurones artificiel qui permet de combiner les variables de sortie de chaque modèle (MOV : Model Output Variable) pour obtenir un indice global de qualité du système évalué. On peut notamment retransmettre les MOV à la sortie pour des utilisateurs avancés afin de pouvoir identifier les caractéristiques de la dégradation.

4 Analyse des résultats

Etant donné que la validation de la méthode objective est en cours, ses résultats, et bien qu'ils soient très encourageants viendront par la suite. Nous présentons dans ce qui suit, les résultats de l'étude réalisée pour la modélisation de l'effet d'expérience. Comme mentionné précédemment, le test ABC en mode Ref-Deg-Ref permet de mettre en évidence cet effet. Afin d'identifier le cycle de récupération (disparition de l'effet), on propose le test Ref-Deg-Ref-Ref-..., qu'on limite à cinq séquences pour éviter de trop fatiguer les sujets.

Durant cette étude, on s'est limité à des dégradations usuelles et réalistes qu'on peut retrouver dans de nombreuses applications. Les tests ont été effectués avec la participation de 30 personnes de différents cultures audio. Les figures 5 et 6 montre respectivement les résultats du test précédent pour des dégradations de type AWGN centré avec un SNR de 36,5 dB et des pertes aléatoires de paquets à 0.4%. On souligne particulièrement que le

cycle de récupération dépend de la nature et l'intensité de la dégradation.

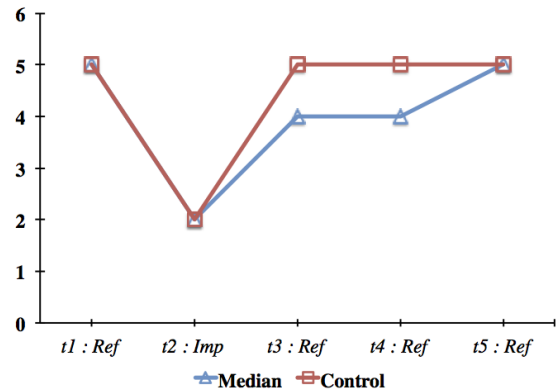


Figure 5 – Cycle de récupération suite à un AWGN

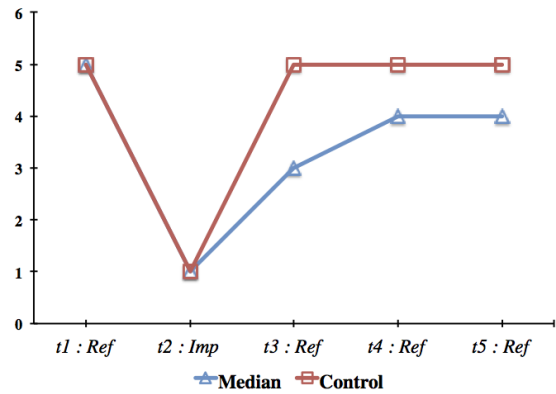


Figure 6 – Cycle de récupération suite à des pertes paquets

Cet effet peut être considéré de deux façons :

- Le modéliser et l'intégrer dans les tests subjectifs pour le lissage des données,
- L'intégrer dans une approche objective en l'appliquant sur les notes intermédiaires au niveau du modèle psychologique. Ceci permettra d'améliorer le facteur de corrélation vu qu'on s'approche d'avantage du phénomène naturel.

On peut donc noter que pour la note attribuée par un sujet à une séquence i et notée $G_{rated}(S_i)$:

$$G_{rated}(S_i) = G_{actual}(S_i) - PQD \quad (5)$$

$$PQD = a * \delta \quad (6)$$

$$\delta = f(G_{rated}(S_{i-1}), \varphi) \quad (7)$$

où $G_{actual}(S_i)$ correspond à la note « réelle » et PQD à la chute « perçue » de qualité. Cette dernière dépend d'une variable d'activation a (= 1 si la séquence précédente est dégradée, 0 sinon) et de la note attribuée à

la séquence précédente et potentiellement d'autres facteurs psychologiques (fatigue, stress, ennui, ...), représentés par la variable φ . la valeur de la PQD peut être estimée statistiquement.

On peut donc établir une fonction de pondération qui relie les notes réelles aux notes perçues comme le montre l'équation suivante :

$$\begin{pmatrix} G_{adj}(S_{i+1}) \\ G_{adj}(S_{i+2}) \\ \vdots \\ G_{adj}(S_{R-1}) \end{pmatrix} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & w_n \end{pmatrix} \times \begin{pmatrix} G_{rated}(S_{i+1}) \\ G_{rated}(S_{i+2}) \\ \vdots \\ G_{rated}(S_{R-1}) \end{pmatrix} \quad (8)$$

Les valeurs de cette matrice peuvent être obtenues par le rapport entre les données de contrôle et les données mesurées. La fonction de pondération présente un intérêt calculatoire important par sa légèreté et facilité d'implémentation. D'autres techniques peuvent être utilisées pour effectuer cette correction comme les Modèles de Markov Cachés (HMM) et leur intégration est en cours d'étude.

5 Conclusion et perspectives

Avec l'apparition d'une nouvelle génération de contenus multimédia 3D plus riches, réalistes et interactifs, le public est devenu plus conscient et exigeant par rapport à leur qualité. Notre étude présente certaines lacunes des méthodes existantes d'évaluation de la qualité, à la fois subjectives et objectives avec les solutions qu'on propose pour y remédier. On propose dans ce sens d'étendre la méthode PEAQ de l'ITU aux flux audio multicanaux grâce à l'introduction d'un système binaural qui réduit le nombre de canaux à analyser, tout en préservant l'information spatiale. Les protocoles subjectifs bénéficient aussi d'une révision pour les rendre compatible avec le son 3D et intégrer d'avantage les facteurs psychologiques. Ces derniers sont au centre de notre étude actuelle où on cherche à les modéliser d'une façon plus objective à travers les réponses physiologiques des sujets (EEG, Rythme Cardiaque, EMG au niveau du bras, ...).

Nous travaillons notamment sur l'optimisation du modèle perceptif pour rendre la solution compatible avec les applications temps réel.

Ce travail constitue la première étape pour proposer une méthodologie d'évaluation complète de la qualité du son multicanal, en proposant un outil objectif pour l'évaluation de la qualité soutenu par des protocoles subjectifs légers et efficaces.

Références

[1] ITU-R. General methods for the subjective assessment of sound quality. Rapport technique, ITU-R STANDARDS, 12 2003.

[2] ITU-R. Methods for the subjective assessment of small impairments in audio systems including

multichannel sound systems. Rapport technique, ITU-R Standards, 06 2001.

- [3] ITU-R. Method for the subjective assessment of intermediate quality levels of coding systems. Rapport technique, ITU-R Standards, 01 2003.
- [4] S/lawomir Zielinski, Francis Rumsey, et Søren Bech. On some biases encountered in modern audio quality listening tests—a review. *J. Audio Eng. Soc.*, 56(6) :427–451, 2008.
- [5] K. Seshadrinathan et Alan C. Bovik. Temporal hysteresis model of time varying subjective video quality. Dans *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 1153–1156, 2011.
- [6] Pierre Bezar. Etude de l'impact de la compression audio sur l'effet binaural. Rapport de stage non publié, 06 2013.
- [7] ITU-R. Method for objective measurements of perceived audio quality. Rapport technique, ITU-R Standards, Jun 2002.
- [8] Dermot Campbell, Edward Jones, et Martin Glavin. Audio quality assessment techniques—a review, and recent developments. *Signal Processing*, 89(8) :1489 – 1500, March 2009.
- [9] Christopher R. Cave. Perceptual modelling for low-rate audio coding. Mémoire de D.E.A., McGill University, Jun 2002.
- [10] R. Huber et B. Kollmeier. Pemo-q—a new method for objective audio quality assessment using a model of auditory perception. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(6) :1902–1911, 2006.
- [11] R. Vanam et C.D. Creusere. Evaluating low bitrate scalable audio quality using advanced version of peaq and energy equalization approach. Dans *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 3, pages iii/189–iii/192 Vol. 3, 2005.
- [12] Inyong Choi, Barbara G. Shinn-Cunningham, Sang Bae Chon, et Koeng-Mo Sung. Objective measurement of perceived auditory quality in multichannel audio compression coding systems. *J. Audio Eng. Soc.*, 56(1/2) :3–17, 2008.