

RECOMMANDATION UIT-R BS.1770-1

**Algorithmes de mesure de l'intensité sonore des programmes audio
et des niveaux de crête vrais des signaux audio**

(Question UIT-R 2/6)

(2006-2007)

Domaine de compétence

La présente Recommandation spécifie des algorithmes de mesure des signaux audio permettant de déterminer l'intensité sonore subjective des programmes et le niveau de crête vrai des signaux.

L'Assemblée des radiocommunications de l'UIT-R,

considérant

- a) que les techniques modernes de transmission sonore numérique offrent une plage dynamique extrêmement large;
- b) que les techniques modernes de production et de transmission sonores numériques offrent une combinaison de formats mono, stéréo et multicanal et que des programmes sonores sont produits dans tous ces formats;
- c) que les auditeurs souhaitent que l'intensité sonore subjective des programmes audio soit uniforme pour différentes sources et différents types de programmes;
- d) qu'il existe de nombreuses méthodes permettant de mesurer les niveaux sonores, mais que les méthodes de mesure actuellement employées pour la production des programmes ne donnent pas d'indication de l'intensité sonore subjective;
- e) qu'il est indispensable, pour l'échange de programmes, de disposer d'un seul et même algorithme recommandé pour l'estimation objective de l'intensité sonore subjective;
- f) que de futurs algorithmes complexes fondés sur des modèles acoustiques fourniront peut-être des mesures objectives améliorées de l'intensité sonore pour une grande diversité de programmes audio;
- g) que la surcharge des supports numériques se fait brusquement et qu'il convient donc d'éviter les surcharges, même lorsqu'elles sont temporaires,

considérant en outre

- h) que les niveaux de crête des signaux peuvent augmenter en raison de processus couramment utilisés tels que le filtrage ou la réduction du débit binaire;
- j) que les techniques de mesure actuelles ne reflètent pas le niveau de crête vrai contenu dans un signal numérique, étant donné que la valeur de crête vraie peut être observée entre les échantillons;

k) que, compte tenu de l'état actuel du traitement des signaux numériques, il est possible de mettre en œuvre un algorithme permettant d'avoir une bonne estimation du niveau de crête vrai d'un signal;

l) que l'utilisation d'un algorithme indiquant le niveau de crête vrai permettra d'avoir une indication précise de la marge entre le niveau de crête d'un signal audionumérique et le niveau d'écrêtage,

recommande

1 que, lorsqu'il est nécessaire de procéder à une mesure objective de l'intensité sonore d'un canal ou d'un programme audio pour faciliter la diffusion et l'échange de programmes, l'algorithme spécifié dans l'Annexe 1 soit utilisé;

2 que les méthodes employées pour la production et la postproduction de programmes afin d'indiquer l'intensité sonore des programmes puissent être fondées sur l'algorithme spécifié dans l'Annexe 1;

3 que, lorsqu'il est nécessaire d'indiquer le niveau de crête vrai d'un signal audionumérique, la méthode de mesure soit fondée sur les lignes directrices exposées dans l'Annexe 2 ou sur une méthode donnant des résultats comparables ou meilleurs,

NOTE 1 – Les utilisateurs doivent savoir que l'intensité sonore mesurée est une estimation de l'intensité sonore subjective et entraîne certaines différences en fonction des auditeurs, des données audio et des conditions d'écoute.

recommande en outre

1 que des travaux complémentaires soient effectués pour développer l'algorithme décrit dans l'Annexe 1, afin qu'il fournisse une indication de l'intensité sonore à court terme;

2 que l'on envisage la possibilité de mettre à jour la présente Recommandation, au cas où il s'avérerait que de nouveaux algorithmes de mesure de l'intensité sonore permettent d'obtenir une qualité de fonctionnement nettement meilleure que celle de l'algorithme décrit dans l'Annexe 1.

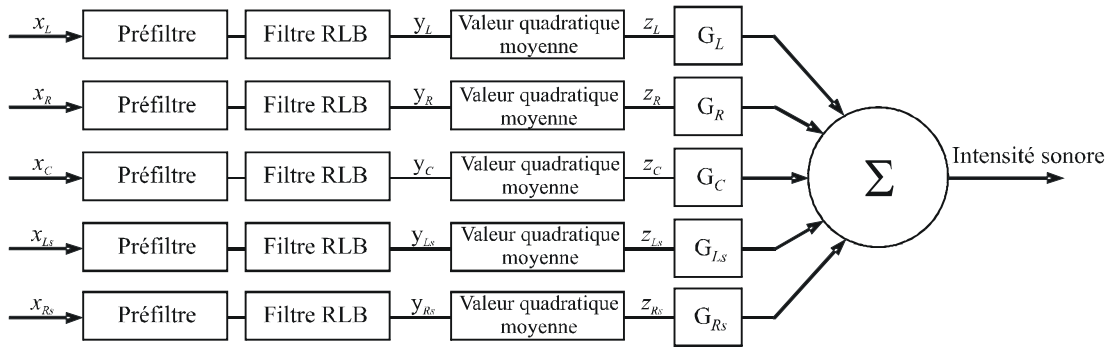
Annexe 1

Spécification de l'algorithme de mesure de l'intensité sonore objective multicanal

La présente Annexe décrit l'algorithme de mesure de l'intensité sonore multicanal. La Fig. 1 est un schéma fonctionnel des diverses composantes de cet algorithme. Des étiquettes sont fournies en différents points le long du trajet de flux du signal, pour faciliter la description de l'algorithme. Le schéma fonctionnel indique les entrées pour cinq canaux principaux (canal de gauche, canal central, canal de droite, canal ambiophonique gauche et canal ambiophonique droit), ce qui permet de contrôler les programmes contenant un à cinq canaux. Lorsqu'un programme comporte moins de cinq canaux, certaines entrées ne seront pas utilisées. Le canal d'effets basses fréquences (LFE, *low frequency effect*) n'est pas pris en compte dans les mesures.

FIGURE 1

Schéma fonctionnel de l'algorithme de mesure de l'intensité sonore multicanal

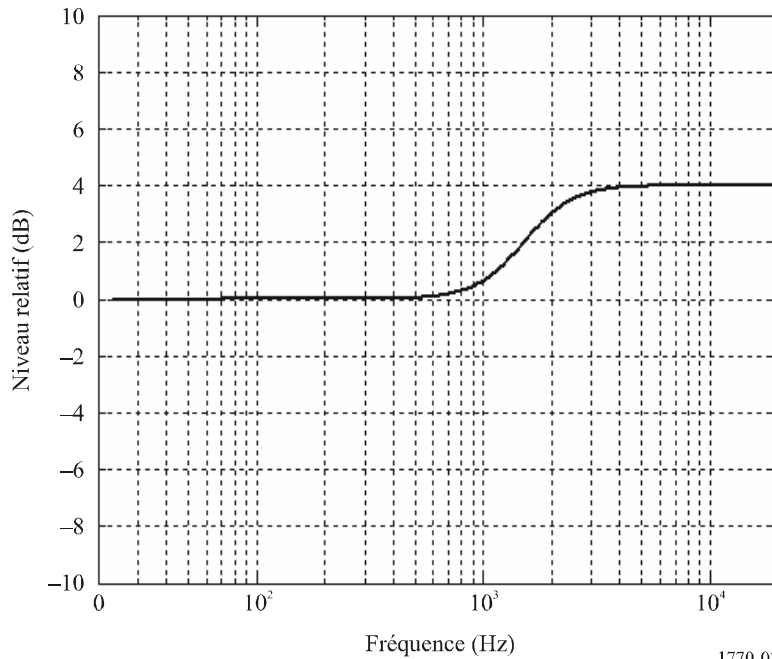


1770-01

La première étape de l'utilisation de l'algorithme consiste à effectuer un préfiltrage du signal avant la mesure $Leq(RLB)$, comme indiqué dans la Fig. 2. Le préfiltrage tient compte des effets acoustiques de la tête, qui est modélisée sous la forme d'une sphère rigide.

FIGURE 2

Réponse du préfiltre utilisé pour tenir compte des effets acoustiques de la tête

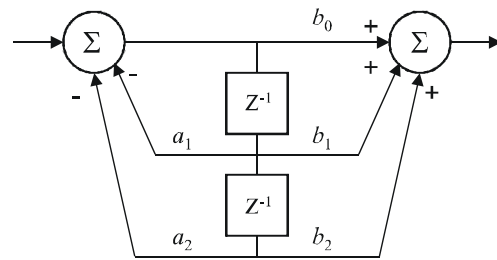


1770-02

Le préfiltre est défini par le filtre indiqué dans la Fig. 3, avec les coefficients donnés dans le Tableau 1.

FIGURE 3

Diagramme de flux de signaux sous la forme d'un filtre de deuxième ordre



1770-03

TABLEAU 1

Coefficients de filtre pour que le préfiltre modélise une tête sphérique

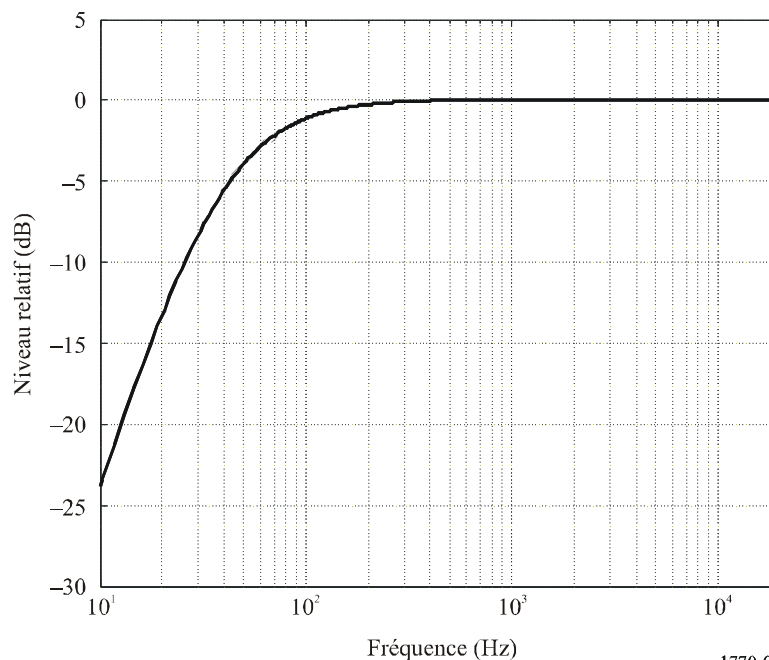
		b_0	1,53512485958697
a_1	-1,69065929318241	b_1	-2,69169618940638
a_2	0,73248077421585	b_2	1,19839281085285

Ces coefficients de filtre sont prévus pour une fréquence d'échantillonnage de 48 kHz. Les mises en œuvre à d'autres fréquences d'échantillonnage nécessiteront des valeurs de coefficient différentes, qui devraient être choisies de manière à fournir la même réponse en fréquence que celle que le filtre spécifié fournit à 48 kHz. Il peut être nécessaire de quantifier les valeurs de ces coefficients en raison de la précision interne du matériel disponible. Des essais ont montré que la qualité de fonctionnement de l'algorithme n'est pas sensible à de légères variations de ces coefficients.

La deuxième étape de l'utilisation de l'algorithme consiste à appliquer la courbe de pondération RLB, qui comprend un filtre passe-haut simple, comme indiqué dans la Fig. 4.

FIGURE 4

Courbe de pondération RLB



1770-04

La courbe de pondération RLB est spécifiée sous la forme d'un filtre de deuxième ordre, comme indiqué dans la Fig. 3, avec les coefficients spécifiés dans le Tableau 2.

TABLEAU 2
Coefficients de filtre pour la courbe de pondération RLB

		b_0	1,0
a_1	-1,99004745483398	b_1	-2,0
a_2	0,99007225036621	b_2	1,0

Ces coefficients de filtre sont prévus pour une fréquence d'échantillonnage de 48 kHz. Les mises en œuvre à d'autres fréquences d'échantillonnage nécessiteront des valeurs de coefficient différentes, qui devraient être choisies de manière à fournir la même réponse en fréquence que celle que le filtre spécifié fournit à 48 kHz.

Une fois que le préfiltre et le filtrage RLB ont été appliqués, on mesure la valeur quadratique moyenne de l'énergie, dans l'intervalle de mesure T , à l'aide de la formule:

$$z_i = \frac{1}{T} \int_0^T y_i^2 dt \quad (1)$$

où y_i est le signal d'entrée filtré par le préfiltre, pour modéliser les effets de tête et la courbe de pondération RLB ($i = L, R, C, Ls, Rs, N$, où N est le nombre de canaux).

Une fois que la moyenne quadratique pondérée, z_i , a été calculée pour chaque canal, la dernière étape consiste à additionner les N canaux de la façon suivante:

$$\text{Intensité sonore} = -0,691 + 10 \log_{10} \sum_i^N G_i \cdot z_i \quad \text{LKFS} \quad (2)$$

Dans cette mesure, la pondération de fréquence, qui est la concaténation du préfiltre et de la pondération RLB, est appelée pondération, K. Le résultat numérique pour la valeur de l'intensité sonore calculée dans la formule (2) devrait être suivi de l'indication LKFS, qui signifie intensité sonore, pondérée par, K, par rapport à une pleine échelle nominale. L'unité LKFS équivaut à un décibel dans la mesure où une augmentation du niveau d'un signal de 1 dB entraînera une augmentation d'une unité LKFS de la valeur de l'intensité sonore lue sur l'instrument de mesure.

Si l'on insère un signal sinusoïdal de 1 kHz et de 0 dB FS à l'entrée du canal de gauche, du canal central ou du canal de droite, l'intensité sonore indiquée sera égale à -3,01 LKFS.

Les coefficients de pondération pour les différents canaux sont indiqués dans le Tableau 3.

TABLEAU 3

Coefficients de pondération pour les différents canaux audio

Canal	Coefficient de pondération, G_i
Gauche (G_L)	1,0 (0 dB)
Droit (G_R)	1,0 (0 dB)
Central (G_C)	1,0 (0 dB)
Ambiophonique gauche (G_{Ls})	1,41 (~ +1,5 dB)
Ambiophonique droit (G_{Rs})	1,41 (~ +1,5 dB)

Il convient de noter que même s'il s'est révélé efficace pour les programmes audio caractéristiques des contenus de radiodiffusion, cet algorithme ne se prête généralement pas à l'estimation de l'intensité sonore subjective pour les tonalités pures.

Appendice 1 à l'Annexe 1

Description et conception de l'algorithme de mesure multicanal

Le présent Appendice décrit un nouvel algorithme conçu pour mesurer objectivement l'intensité sonore perçue des signaux audio. Cet algorithme peut servir pour mesurer avec précision l'intensité sonore des signaux mono, stéréo et multicanaux. L'un de ses principaux avantages est sa simplicité, de sorte qu'il peut être mis en œuvre à un coût très modique. Le présent Appendice expose également les résultats d'essais subjectifs formels effectués pour créer une base de données subjective, qui a servi à évaluer la qualité de fonctionnement de l'algorithme.

1 Introduction

Dans de nombreuses applications, il faut mesurer et contrôler l'intensité sonore perçue des signaux audio. On peut citer, par exemple, les applications de radiodiffusion sonore et télévisuelle, dans lesquelles la nature et le contenu des données audio varient fréquemment. Dans ces applications, le contenu audio peut constamment varier entre des signaux musicaux, des signaux vocaux et des effets sonores ou une combinaison d'entre eux. Ces variations de contenu des éléments de programmes peuvent entraîner de profondes modifications de l'intensité sonore subjective. De plus, on applique fréquemment diverses formes de traitement dynamique aux signaux, ce qui peut avoir des incidences importantes sur l'intensité sonore perçue du signal. Bien entendu, la question de l'intensité sonore subjective revêt également beaucoup d'importance pour l'industrie musicale, qui utilise couramment le traitement dynamique pour augmenter au maximum l'intensité sonore perçue d'un enregistrement.

Ces dernières années, le Groupe de travail 6P des radiocommunications s'est efforcé d'identifier un moyen objectif permettant de mesurer l'intensité sonore perçue de programme type pour les applications de radiodiffusion. Dans un premier temps, l'UIT-R a examiné exclusivement les algorithmes de mesure de l'intensité sonore objective monophonique et a constaté qu'une mesure quadratique moyenne pondérée, $Leq(RLB)$, donnait les meilleurs résultats pour les signaux monophoniques [Soulodre, 2004].

On sait qu'un appareil de mesure de l'intensité sonore pouvant fonctionner sur des signaux mono, stéréo et multicanaux est nécessaire pour les applications de radiodiffusion. Le présent document propose un nouvel algorithme de mesure de l'intensité sonore qui fonctionne de façon satisfaisante sur des signaux audio, mono, stéréo et multicanaux. L'algorithme proposé est fondé sur une extension simple de l'algorithme $Leq(RLB)$. De plus, le nouvel algorithme multicanal conserve la très faible complexité de calcul de l'algorithme monophonique $Leq(RLB)$.

2 Considérations générales

Au cours de la première phase de ses travaux, l'UIT-R a mis au point une méthode d'essai subjective pour examiner la perception de l'intensité sonore dans le cas de programmes monophoniques types [Soulodre, 2004]. Des essais subjectifs ont été effectués sur cinq sites à travers le monde, afin de créer une base de données subjective visant à évaluer la qualité de fonctionnement d'algorithmes potentiels de mesure de l'intensité sonore. Les auditeurs ont comparé l'intensité sonore de diverses séquences audio monophoniques à une séquence de référence. Les séquences audio provenaient de programmes de radiodiffusion réels (radiodiffusion sonore et télévisuelle).

Parallèlement à ces tests, sept entités différentes ont soumis un total de dix appareils ou algorithmes de mesure de l'intensité sonore monophonique, en vue de leur mise sur le marché, pour évaluation au Laboratoire de perception sonore du Centre de recherches sur les communications du Canada.

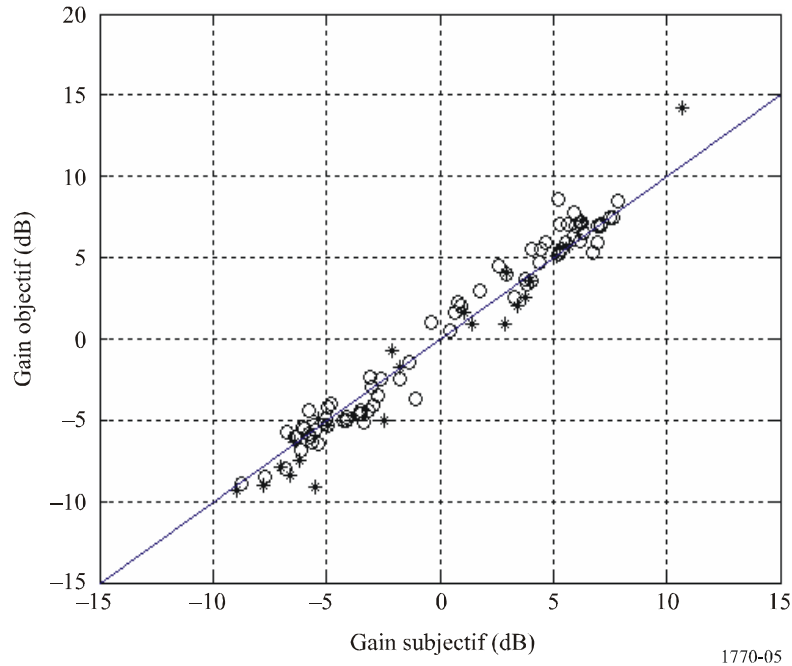
En outre, Soulodre a présenté deux autres algorithmes de base de mesure de l'intensité sonore destinés à servir de référence en matière de qualité de fonctionnement [Soulodre, 2004]. Ces deux mesures objectives comprenaient une fonction simple de pondération en fréquence, suivie d'un bloc de mesure de la valeur quadratique moyenne. L'une des deux mesures, $Leq(RLB)$, utilise une courbe de pondération en fréquence passe-haut appelée courbe B basse fréquence révisée (RLB).

L'autre mesure, Leq , est simplement une mesure non pondérée de la valeur quadratique moyenne.

Les résultats de l'étude initiale menée par l'UIT-R pour l'appareil de mesure de l'intensité sonore $Leq(RLB)$ sont présentés sur la Fig. 5. L'axe horizontal indique l'intensité sonore subjective relative provenant de la base de données subjective, tandis que l'axe vertical indique l'intensité sonore prévue par la mesure $Leq(RLB)$. Chaque point du graphique correspond au résultat obtenu pour l'une des séquences d'essai audio de l'essai. Les cercles ouverts représentent les séquences audio fondées sur la parole, tandis que les étoiles correspondent à des séquences qui ne sont pas fondées sur la parole. On constate que les points de données sont regroupés de façon très dense autour de la diagonale, ce qui témoigne de la très bonne qualité de fonctionnement de l'appareil de mesure $Leq(RLB)$.

FIGURE 5

Appareil de mesure de l'intensité sonore monophonique $Leq(RLB)$
et résultats subjectifs ($r = 0,982$)



Il est apparu que parmi tous les appareils de mesure évalués, l'appareil $Leq(RLB)$ était celui qui donnait les meilleurs résultats (encore que du point de vue statistique, certains appareils de mesure psychoacoustique donnaient également de bons résultats). On a également constaté que l'appareil de mesure Leq fonctionnait presque aussi bien que l'appareil RLB. Ces conclusions montrent que pour un programme de radiodiffusion monophonique type, une mesure simple de l'intensité sonore fondée sur l'énergie offre la même fiabilité que des mesures plus complexes pouvant comporter des modèles perceptuels détaillés.

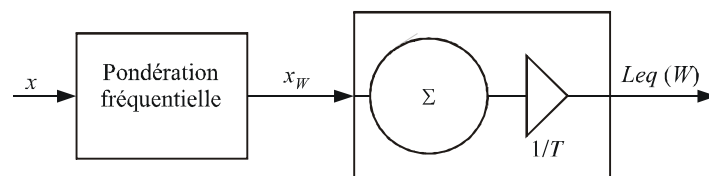
3 Conception de l'algorithme $Leq(RLB)$

L'algorithme de mesure de l'intensité sonore $Leq(RLB)$ a été spécialement conçu pour être très simple. Un schéma fonctionnel de cet algorithme est reproduit dans la Fig. 6. Il comprend un filtre passe-haut suivi d'un moyen permettant d'établir la moyenne de l'énergie dans le temps. La sortie du filtre va vers un bloc de traitement qui additionne l'énergie et calcule la moyenne dans le temps.

Le filtre vise à fournir une pondération du contenu spectral du signal perçue comme étant significative. L'un des avantages de l'utilisation de cette structure de base pour les mesures de l'intensité sonore est que toutes les opérations de traitement peuvent être effectuées à l'aide de blocs temporels simples nécessitant très peu de calculs.

FIGURE 6

Schéma fonctionnel des mesures simples de l'intensité sonore fondées sur l'énergie



L'algorithme $Leq(RLB)$ représenté sur la Fig. 6 constitue simplement une version pondérée en fréquence d'une mesure du *niveau sonore équivalent* (Leq). Leq est défini à l'aide de la formule suivante:

$$Leq(W) = 10 \log_{10} \left[\frac{1}{T} \int_0^T \frac{x_W^2}{x_{Ref}^2} dt \right] \quad \text{dB} \quad (3)$$

où:

- x_W : signal à la sortie du filtre de pondération
- x_{Ref} : niveau de référence
- T : longueur de la séquence audio.

Dans la formule $Leq(W)$, le symbole W représente la pondération fréquentielle, qui, en l'espèce, était la courbe B basse fréquence révisée (RLB).

4 Essais subjectifs

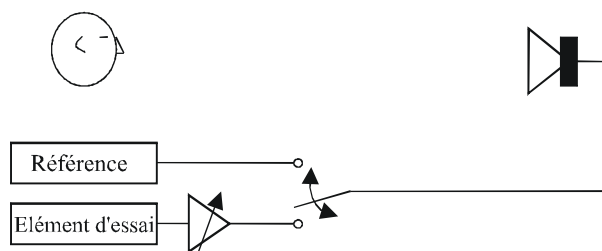
Afin d'évaluer les mesures potentielles de l'intensité sonore multicanal, il a fallu procéder à des essais subjectifs formels pour créer une base de données subjective. On a pu ainsi évaluer les algorithmes de mesure potentielle de l'intensité sonore du point de vue de leur capacité de prévoir les résultats des essais subjectifs. La base de données a fourni des équivalents pour la sonie perçue pour une grande diversité de programmes mono, stéréo et multicanaux. Les éléments de programmes utilisés lors des essais provenaient de programmes réels de radiodiffusion sonore et télévisuelle du monde entier ainsi que de CD et de DVD. Les séquences comprenaient des séquences musicales et des séries télévisées, des films dramatiques, des manifestations sportives, des bulletins d'information, des effets sonores et des messages publicitaires. Elles comportaient aussi des segments vocaux dans plusieurs langues.

4.1 Dispositif d'essai subjectif

Les essais subjectifs comportaient une opération d'adaptation de l'intensité sonore. Les participants écoutaient une large gamme de programmes types et réglaient le niveau de chaque élément d'essai, jusqu'à ce que l'intensité sonore perçue de cet élément corresponde à celle d'un signal de référence (voir la Fig. 7).

Le signal de référence était toujours reproduit à un niveau de 60 dBA, considéré par Benjamin comme un niveau d'écoute type pour les téléspectateurs à domicile [Benjamin, 2004].

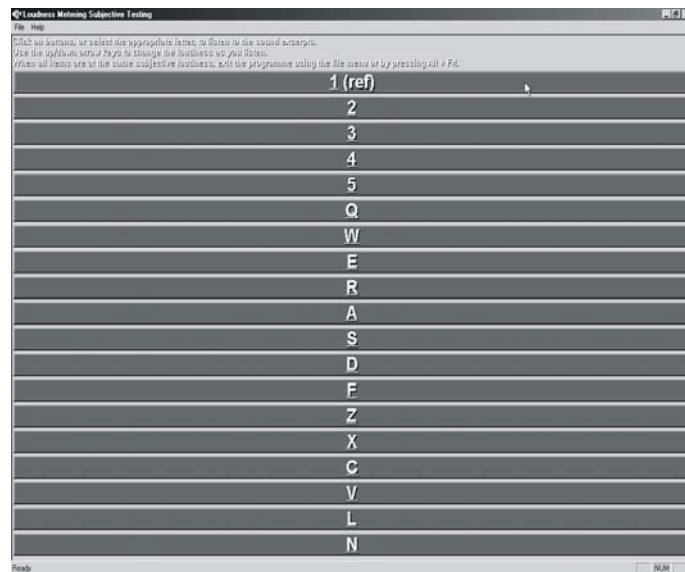
FIGURE 7
Méthode d'essai subjectif



Un système d'essai subjectif multicanal basé sur un logiciel, conçu et fourni par l'Australian Broadcasting Corporation, permettait à l'auditeur de passer instantanément d'un élément de test à l'autre et de régler le niveau (intensité sonore) de chaque élément. Une vue sur écran du logiciel utilisé pour l'essai est représentée sur la Fig. 8. Le niveau des éléments d'essai pouvait être réglé par pas de 0,25 dB. En sélectionnant le bouton «1», on pouvait accéder au signal de référence. Le niveau du signal de référence était maintenu sur une valeur fixe.

FIGURE 8

Interface utilisateur du système de test subjectif



1770-08

A l'aide du clavier de l'ordinateur, l'auditeur choisissait un élément d'essai donné et en réglait le niveau jusqu'à ce que son intensité sonore corresponde au signal de référence. Les auditeurs pouvaient instantanément passer de l'un de ces éléments d'essai à l'autre en choisissant la touche appropriée. Les séquences étaient diffusées en permanence (boucle) pendant les essais. Le logiciel enregistrait les réglages du gain pour chaque élément d'essai, tels qu'ils avaient été fixés par l'auditeur. En conséquence, les essais subjectifs ont permis d'obtenir une série de valeurs de gain (décibels) nécessaires pour adapter l'intensité sonore de chaque séquence d'essai à la séquence de référence. Ainsi, on a pu déterminer directement l'intensité sonore relative de chaque élément d'essai.

Avant d'effectuer les tests aveugles formels, chaque auditeur avait suivi une séance de formation au cours de laquelle il a pu se familiariser avec le logiciel d'essai et les tâches à accomplir lors de l'expérience. Etant donné que bon nombre des éléments d'essai contenaient un ensemble de signaux vocaux et d'autres signaux sonores (musique, bruits de fond, etc.), les auditeurs avaient reçu pour consigne expresse d'adapter l'intensité sonore du signal global, et pas seulement la composante vocale des signaux.

Pendant les tests aveugles formels, l'ordre dans lequel les éléments d'essai étaient présentés à chaque auditeur a été rendu aléatoire. En conséquence, les éléments d'essai n'étaient jamais présentés dans le même ordre à deux auditeurs, de façon à supprimer tout risque d'impartialité due aux effets de la séquence.

4.2 Base de données subjective

La base de données subjective utilisée pour évaluer la qualité de fonctionnement de l'algorithme proposé comprenait en fait trois ensembles de données distincts. Ceux-ci ont été créés à partir de trois essais subjectifs indépendants effectués sur quelques années.

Le premier ensemble de données comprenait les résultats provenant de l'étude initiale de l'UIT-R au cours de laquelle des auditeurs mettaient en correspondance l'intensité sonore perçue de 96 séquences audio monophoniques. Pour cet ensemble de données, des essais subjectifs ont été effectués sur cinq sites distincts dans le monde avec la participation de 97 auditeurs. Un groupe de trois personnes, composé de membres du Groupe spécial du Rapporteur du GT 6P des radiocommunications, a sélectionné les séquences d'essai ainsi que l'élément de référence. Au cours de cette expérience, le signal de référence était composé d'une voix de femme anglaise. Les séquences étaient reproduites par l'intermédiaire d'un seul haut-parleur placé directement en face de l'auditeur.

A la suite de l'étude monophonique initiale menée par l'UIT-R, certaines entités ayant proposé l'algorithme pensaient que la gamme et le type des signaux utilisés dans les essais subjectifs n'étaient pas suffisamment larges et que, de ce fait, l'algorithme simple fondé sur l'énergie du signal *Leq*(RLB) donnait de meilleurs résultats que tous les autres algorithmes.

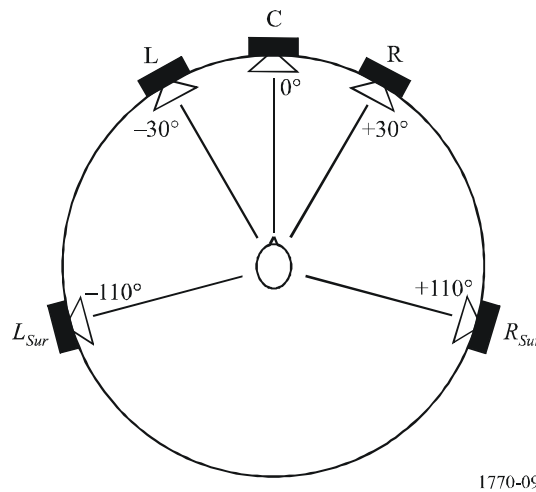
Pour remédier à ce problème, on a demandé aux initiateurs de soumettre de nouvelles séquences audio pour une nouvelle série d'essais subjectifs et on les a encouragés à soumettre des séquences monophoniques qui, selon eux, seraient relativement plus difficiles à reproduire avec l'algorithme *Leq*(RLB). Parmi les entités ayant proposé des instruments de mesure, seules deux ont présenté de nouvelles séquences.

Des essais subjectifs formels ont été réalisés au Laboratoire de perception sonore du Centre de recherches sur les communications (Canada) à l'aide de ces nouvelles séquences. Vingt auditeurs ont fourni des équivalents pour la sonie pour 96 séquences monophoniques. Pour les essais, on a utilisé la même méthode subjective que celle qui avait été employée pour mettre au point le premier ensemble de données ainsi que le même signal de référence. Les résultats de ces essais ont constitué le deuxième ensemble de données de la base de données subjective.

Le troisième ensemble de données était composé d'équivalents pour la sonie pour 144 séquences audio. Les séquences d'essai comprenaient 48 éléments monophoniques, 48 éléments stéréo et 48 éléments multicanaux. De plus, les éléments monophoniques ont été reproduits pour moitié via le canal central (mono), et pour moitié via les haut-parleurs gauche et droit (double mono), afin de tenir compte des deux différentes manières permettant d'écouter un signal monophonique. Aux fins de cet essai, le signal de référence était composé d'une voix de femme anglaise, avec musique d'ambiance stéréo et musique de fond de faible niveau. Vingt auditeurs ont participé à cet essai, qui utilisait la configuration de haut-parleurs spécifiée dans la Recommandation UIT-R BS.775 (voir la Fig. 9).

FIGURE 9

Configuration de haut-parleurs utilisée pour le troisième ensemble de données



Etant donné que les deux premiers ensembles de données étaient limités aux séquences d'essai monophoniques, les images n'entraient pas en ligne de compte. Dans le troisième ensemble de données, qui comportait aussi des séquences stéréo et multicanal, les images constituaient un facteur important qu'il fallait traiter. On a considéré que, dans une séquence, les images et l'ambiance pouvaient vraisemblablement avoir des conséquences importantes sur l'intensité sonore perçue de cette séquence. En conséquence, on a choisi des séquences stéréo et multicanal comprenant une large gamme de types d'images (par exemple, panoramique centrale/extrême gauche/droite, sources de face/sources tout autour) et différents types d'ambiances (sécheresse de la salle/avec réverbération).

Le fait que les auditeurs devaient simultanément mettre en correspondance l'intensité sonore des signaux mono, double mono, stéréo et multicanaux, signifiait que cet essai était par nature plus difficile que les ensembles de données précédents, limités à des signaux mono. La situation était encore compliquée par les différents types d'images et les différents effets d'ambiance. On craignait qu'en raison de ces facteurs, les auditeurs ne soient dépassés par la tâche à accomplir. Heureusement, des essais préliminaires ont donné à penser que cette tâche était gérable et les 20 auditeurs ont effectivement été en mesure de fournir des résultats homogènes.

5 Conception de l'algorithme de mesure de l'intensité sonore multicanal

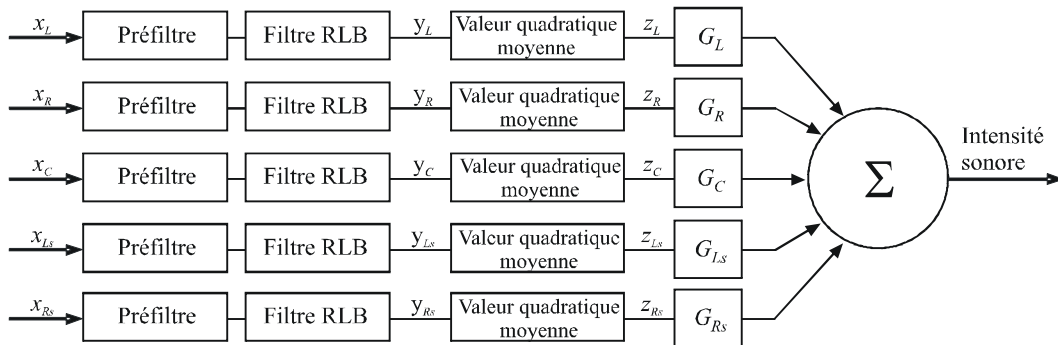
Comme indiqué plus haut, l'algorithme *Leq*(RLB) a été conçu pour fonctionner avec des signaux monophoniques et une étude menée précédemment a montré qu'il donnait de très bons résultats à cet égard. La conception d'un algorithme de mesure de l'intensité sonore multicanal pose d'autres problèmes. Pour qu'un algorithme multicanal fonctionne de manière satisfaisante, il est indispensable qu'il fonctionne bien également avec des signaux mono, double mono et stéréo. Autrement dit, ces formats doivent être considérés comme des cas particuliers d'un signal multicanal (encore que ces cas soient très fréquents).

Dans la présente étude, nous partons du principe que les signaux multicanaux sont conformes à la configuration normalisée de canaux types 5.1, décrite dans la Recommandation UIT-R BS.775. Il n'est pas tenu compte du canal LFE.

Avec l'appareil de mesure de l'intensité sonore multicanal, l'intensité sonore de chacun des différents canaux audio est mesurée de manière indépendante par un algorithme monophonique $Leq(RLB)$, comme indiqué sur la Fig. 10. Toutefois, on applique un préfiltrage à chaque canal avant la mesure $Leq(RLB)$.

FIGURE 10

Schéma fonctionnel de l'appareil de mesure de l'intensité sonore multicanal proposé



1770-10

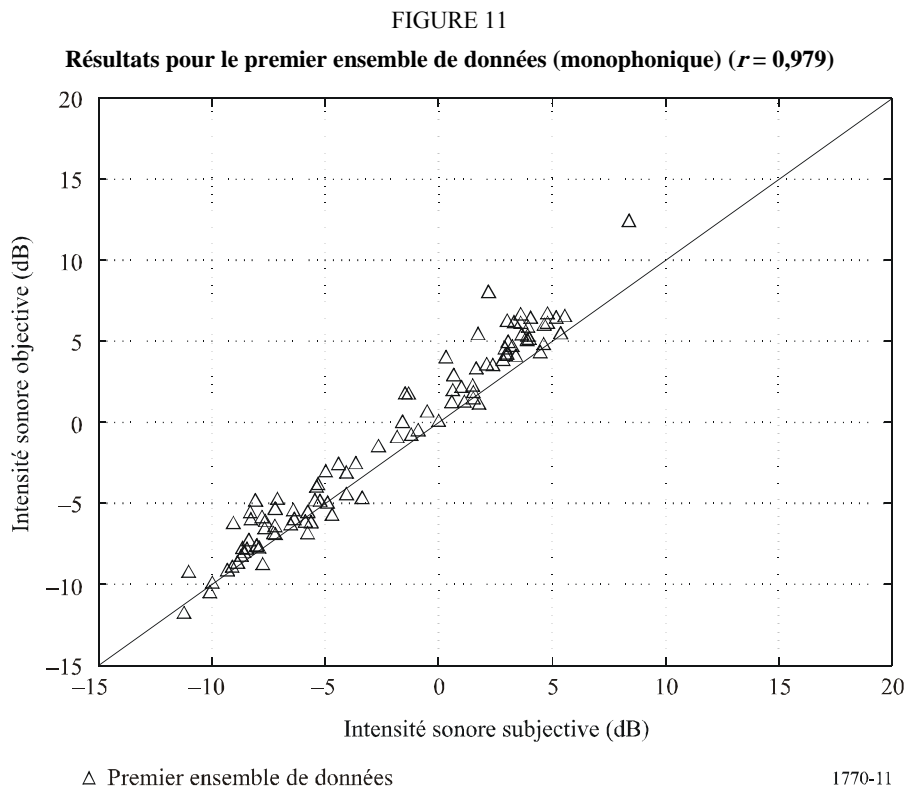
Le préfiltre vise à tenir compte des effets acoustiques de la tête sur les signaux entrants. Dans le cas considéré, la tête est modélisée sous la forme d'une sphère rigide. On applique le même préfiltre à chaque canal. Les valeurs de l'intensité sonore ainsi obtenues sont alors pondérées (G_i) en fonction de l'angle d'arrivée du signal, puis additionnées (dans le domaine linéaire) pour fournir une mesure composite de l'intensité sonore. Les pondérations sont utilisées pour tenir compte du fait que, pour un auditeur, les sons qui arrivent de l'arrière peuvent être perçus comme étant plus forts que ceux qui viennent de face.

L'un des principaux avantages de l'algorithme de mesure de l'intensité sonore multicanal proposé réside dans sa simplicité. Cet algorithme est entièrement composé de blocs de traitement des signaux très simples qui peuvent être facilement mis en œuvre dans le domaine temporel sur du matériel peu coûteux. Autre avantage de l'algorithme: son extensibilité. Étant donné que le traitement appliqué à chaque canal est identique, on peut aisément mettre en œuvre un instrument de mesure capable de prendre en charge un nombre de canaux compris entre 1 et N . De plus, étant donné que les contributions des différents canaux sont additionnées sous la forme de valeurs de l'intensité sonore et non pas au niveau du signal, l'algorithme ne dépend pas de la phase ou de la corrélation entre canaux. En conséquence, la mesure proposée de l'intensité sonore est beaucoup plus générique et fiable.

6 Évaluation de l'algorithme multicanal

On a traité les 336 séquences audio utilisées dans les trois ensembles de données au moyen de l'algorithme multicanal proposé et on a enregistré les équivalents pour la sonie prévus. Par la suite, on a pu évaluer la qualité de fonctionnement globale de l'algorithme sur la base de la concordance entre les équivalents pour la sonie prévus et les équivalents subjectifs obtenus lors des essais subjectifs formels.

Les Fig. 11, 12 et 13 présentent les résultats obtenus avec l'appareil de mesure de l'intensité sonore proposé pour les trois ensembles de données. Sur chaque Figure, l'axe horizontal correspond à l'intensité sonore subjective de chaque séquence audio de l'ensemble de données. L'axe vertical correspond à l'intensité sonore objective prévue par l'appareil de mesure de l'intensité sonore proposé. Chaque point du graphique représente le résultat correspondant à une séquence audio particulière. A noter qu'avec un algorithme objectif parfait, tous les points de données se trouveraient sur la diagonale ayant une pente de 1 et passant par l'origine (comme indiqué sur les figures).



Il ressort de la Fig. 11 que l'algorithme proposé pour la mesure de l'intensité sonore multicanal fonctionne très bien pour ce qui est de la prévision des résultats provenant du premier ensemble de données (monophonique). La corrélation entre les équivalents pour la sonie subjective et la mesure de l'intensité sonore objective est $r = 0,979$.

Comme il ressort de la Fig. 12, la corrélation entre les équivalents pour la sonie subjective et la mesure de l'intensité sonore objective pour le deuxième ensemble de données est également très bonne ($r = 0,985$). Il est intéressant de noter qu'environ la moitié des séquences de cet ensemble de données sont des séquences musicales.

FIGURE 12

Résultats pour le deuxième ensemble de données (monophonique) ($r = 0,985$)

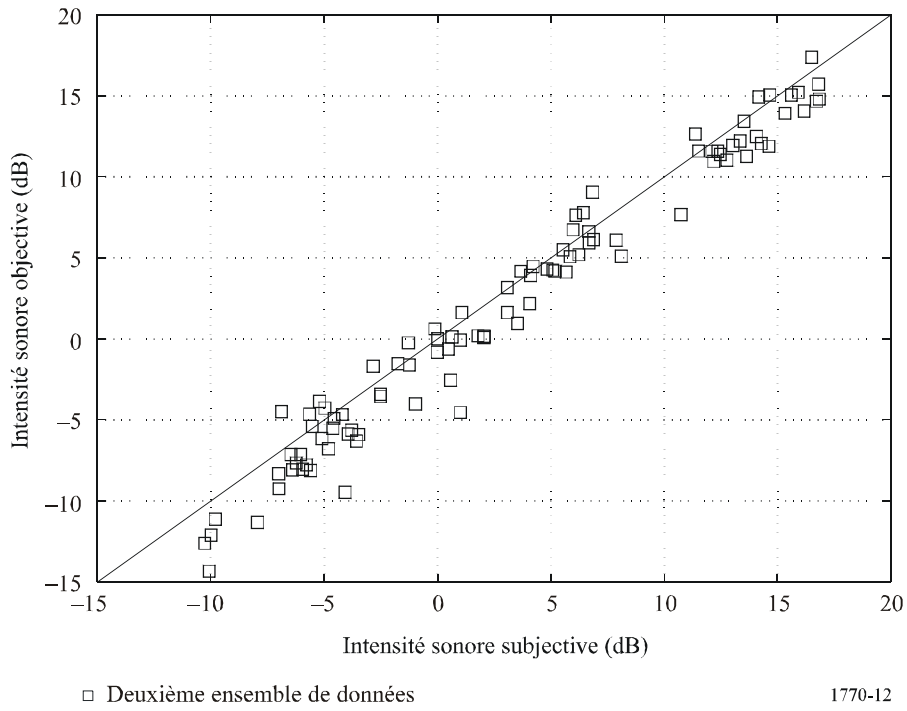
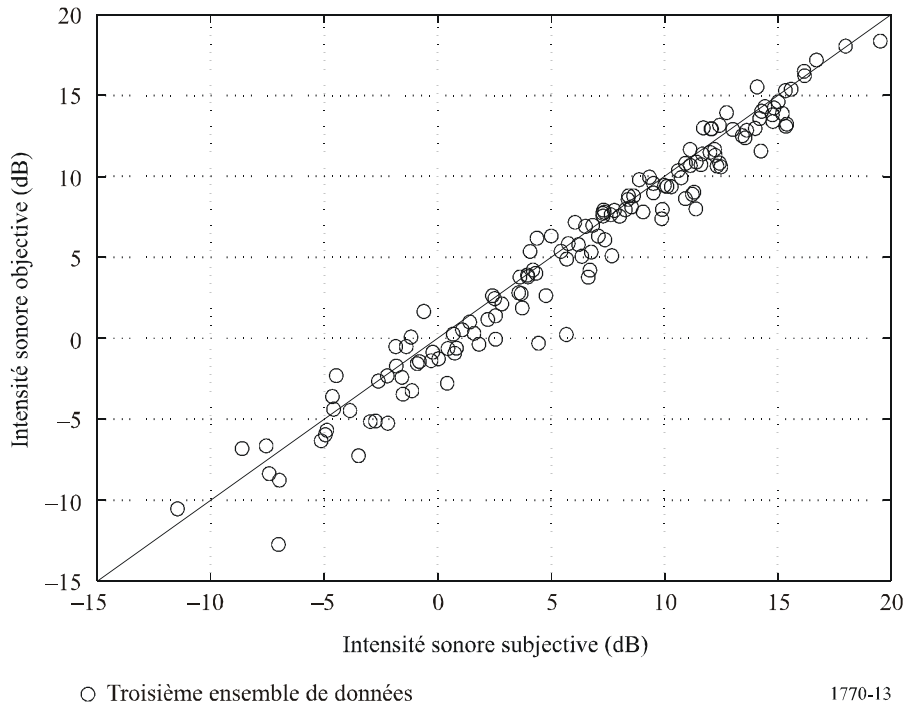


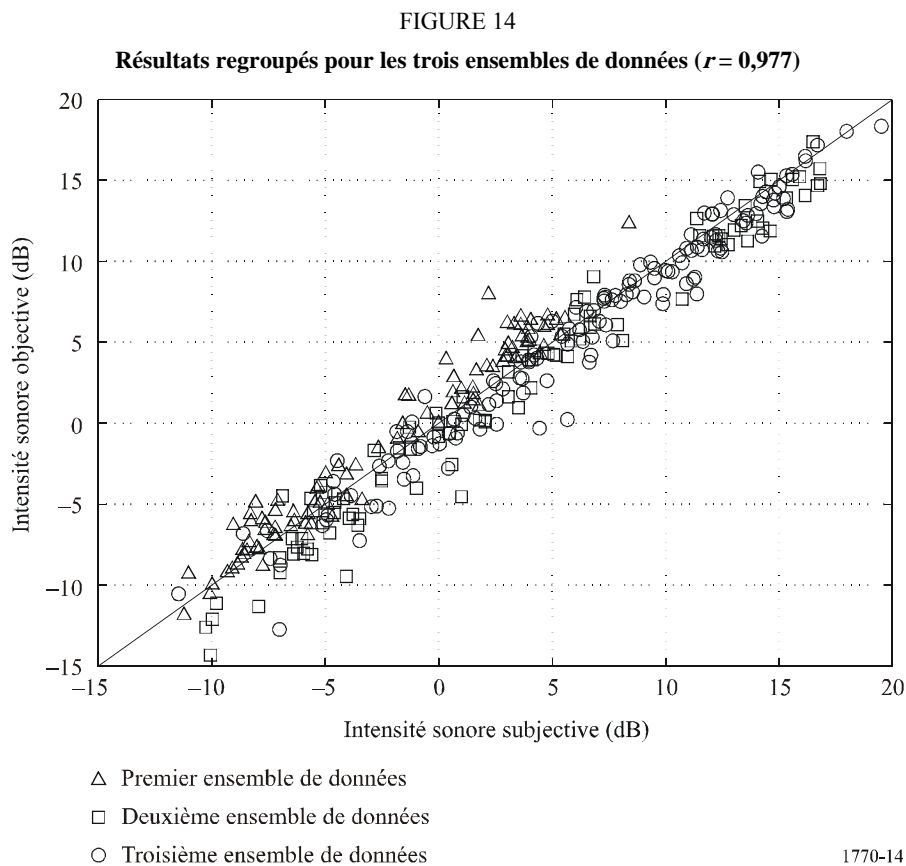
FIGURE 13

Résultats pour le troisième ensemble de données (mono, stéréo et multicanal) ($r = 0,980$)



La Fig. 13 indique les résultats correspondant au troisième ensemble de données, qui comprenait des signaux mono, double mono, stéréo et multicanaux. Les résultats pour les signaux multicanaux présentés sur les Fig. 13 et 14 correspondent à l'algorithme spécifié, mais sans les pondérations du canal ambiophonique fixées à 4 dB (proposition initiale) au lieu de 1,5 dB (spécification finale). On a vérifié que le remplacement de la valeur 4,0 dB par la valeur 1,5 dB n'avait aucune incidence notable sur les résultats. A nouveau, la qualité de fonctionnement de l'algorithme est très satisfaisante, avec une corrélation de $r = 0,980$.

Il est utile d'examiner la qualité de fonctionnement de l'algorithme pour la totalité des 336 séquences audio qui constituaient la base de données subjective. En conséquence, la Fig. 14 regroupe les résultats des trois ensembles de données. On constate que la qualité de fonctionnement est très bonne pour l'intégralité de la base de données subjective, avec une corrélation globale de $r = 0,977$.



Les résultats de cette évaluation montrent que l'algorithme de mesure de l'intensité sonore multicanal, qui repose sur la mesure de l'intensité sonore $Leq(RLB)$, donne de très bons résultats pour les 336 séquences de la base de données subjective. La base de données subjective comprenait une grande diversité de programmes, notamment des séquences musicales, des séries télévisées, des films dramatiques, des manifestations sportives, des bulletins d'information, des effets sonores et des messages publicitaires. Les séquences comportaient aussi des segments vocaux dans plusieurs langues. De plus, les résultats montrent que l'appareil de mesure de l'intensité sonore proposé fonctionne très bien avec des signaux mono, double mono, stéréo et multicanaux.

Références

- SOULODRE, G.A. [mai 2004] – Evaluation of Objective Loudness Meters, 116th Convention of the Audio Engineering Society, Berlin, Preprint 6161.
- BENJAMIN, E. [octobre 2004] – Preferred Listening Levels and Acceptance Windows for Dialog Reproduction in the Domestic Environment, 117th Convention of the Audio Engineering Society, San Francisco, Preprint 6233.

Annexe 2

Lignes directrices relatives à la mesure précise du niveau de «crête vrai»

La présente Annexe décrit un algorithme permettant d'estimer le niveau de crête vrai dans un signal audionumérique MIC linéaire monocanal. Dans les lignes qui suivent, on prend pour hypothèse une fréquence d'échantillonnage de 48 kHz. Le niveau de crête vrai est la valeur maximale (positive ou négative) de la forme d'onde du signal dans le domaine temporel continu; cette valeur peut être supérieure à la valeur la plus élevée de l'échantillon dans le domaine temporel échantillonné à 48 kHz. L'algorithme permet d'estimer le signal tel qu'il est et, à titre facultatif, tel qu'il serait au cas où des équipements situés en aval devaient supprimer la composante DC du signal. A titre facultatif, une légère préaccentuation haute fréquence dans le trajet du signal de mesure de la valeur de crête peut permettre à l'algorithme d'indiquer un niveau de crête plus élevé, pour les signaux haute fréquence, que ce n'est le cas en réalité. En effet, les déplacements de phase des étapes ultérieures du traitement des signaux (l'utilisation de filtres de Nyquist, par exemple) pourraient entraîner une augmentation des niveaux de crête des signaux haute fréquence, fonction qui pourrait être utile dans certaines applications pour assurer une meilleure protection contre l'écrêtage en aval.

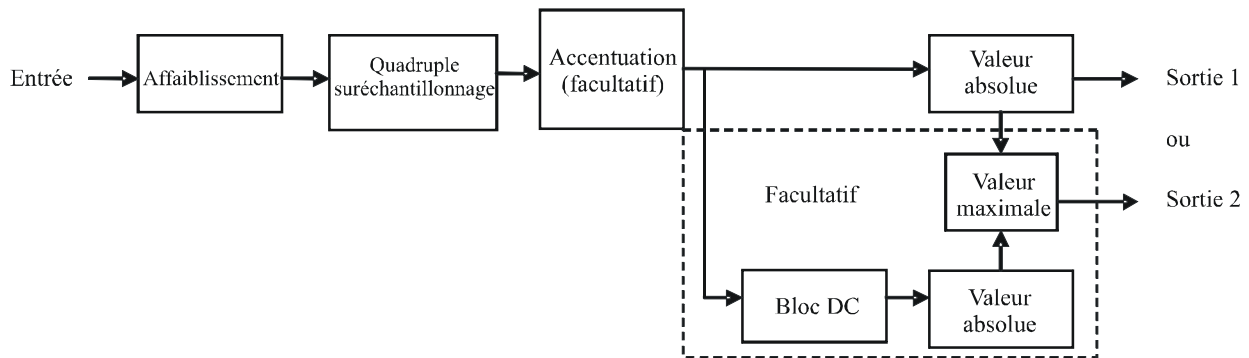
1 Résumé

Les étapes du traitement sont les suivantes:

- 1 Affaiblissement: affaiblissement de 12,04 dB
- 2 Quadruple suréchantillonnage
- 3 Accentuation: filtre de préaccentuation en plateau, zéro à 14,1 kHz, pôle à 20 kHz (facultatif)
- 4 Bloc DC (facultatif)
- 5 Valeur absolue
- 6 Valeur maximale: détection de la valeur la plus élevée (facultatif, inclus si le bloc DC est inclus).

La détection de la valeur absolue avant et après le bloc DC permet d'estimer le niveau de crête du signal au point de mesure actuel ainsi que le niveau de crête en cas de suppression de la composante DC du signal au niveau d'un dispositif situé en aval.

2 Schéma fonctionnel



1770-15

3 Description détaillée

La première étape consiste à imposer un affaiblissement de 12,04 dB (décalage de 2 bits), afin de prévoir une marge pour le traitement ultérieur des signaux au moyen de l'arithmétique d'entiers. Cette étape n'est pas nécessaire si les calculs sont effectués en virgule flottante.

Le filtre de quadruple suréchantillonnage fait passer la fréquence d'échantillonnage du signal de 48 kHz à 192 kHz. Cette version plus élevée de la fréquence d'échantillonnage du signal indique avec davantage de précision la forme d'onde réelle qui est représentée à l'intérieur du signal. La préférence est donnée à des fréquences d'échantillonnage et à des fréquences de suréchantillonnage (voir l'Appendice 1 à la présente Annexe). Les signaux entrants qui se trouvent à des fréquences d'échantillonnage plus élevées nécessitent un suréchantillonnage proportionnellement moins important (par exemple, pour un signal entrant à une fréquence d'échantillonnage de 96 kHz, une valeur de double suréchantillonnage serait suffisante).

Le filtre facultatif de préaccentuation en plateau a pour conséquence que l'algorithme indique un niveau de crête plus élevé pour les composantes du signal aux fréquences supérieures. Cette opération peut être effectuée sans tenir compte du fait qu'il est plus difficile de mesurer et de contrôler les valeurs de crête des composantes du signal aux fréquences supérieures en raison des effets de dispersion (déphasage) qui se produisent dans les nombreux filtres de Nyquist fréquemment utilisés sur l'ensemble de la chaîne du signal de radiodiffusion.

Le filtre de blocage DC facultatif assure une couverture dans le cas où le signal est fortement asymétrique ou contient un décalage de la composante DC. Ce mécanisme facultatif permet non seulement de mesurer la valeur de crête du signal actuel (y compris l'asymétrie ou le décalage de la composante DC), mais aussi de mesurer le signal tel qu'il serait si un équipement situé en aval devait mettre en œuvre un filtre de blocage DC.

On obtient la valeur absolue des échantillons en inversant les échantillons à valeur négative. À ce stade, le signal est unipolaire et les valeurs négatives sont remplacées par des valeurs positives de la même grandeur. La sortie 1 est le flux des valeurs de sortie si le bloc DC facultatif n'est pas mis en œuvre.

Si le bloc DC facultatif est mis en œuvre, le bloc «MAX» choisit l'échantillon le plus grand de chaque échantillon provenant des deux trajets du signal. En l'espèce, la sortie choisie provient de la sortie 2.

Les blocs ultérieurs du système (qui ne sont ni indiqués, ni spécifiés dans le présent document) peuvent comparer les valeurs de l'échantillon en sortie au niveau nominal de crête du signal à 100% (1/4 du signal pleine échelle si on a appliqué à l'entrée un affaiblissement de 12 dB), ce qui permet d'estimer le niveau de crête vrai par rapport au signal numérique en pleine échelle.

Les appareils de mesure qui suivent ces lignes directrices et qui utilisent une fréquence d'échantillonnage suréchantillonnée d'au moins 192 kHz devraient indiquer le résultat en unités dB TP, c'est-à-dire en décibels par rapport à une mesure de niveau de crête vrai pleine échelle de 100%.

Appendice 1* **à l'Annexe 2**

Facteurs à prendre en compte pour mesurer avec précision les valeurs de crête des signaux numériques

Quel est le problème?

Dans les systèmes numériques, les crêtemètres enregistrent souvent une «valeur d'échantillon de crête» plutôt qu'une «valeur de crête vraie».

En général, un appareil de mesure d'une valeur d'échantillon de crête fonctionne en comparant la valeur absolue (modifiée) de chaque échantillon entrant à la valeur actuelle affichée sur l'appareil de mesure. Si le nouvel échantillon est plus grand, il remplace la valeur actuelle indiquée; dans le cas contraire, on multiplie la valeur actuelle indiquée par une constante légèrement inférieure à l'unité pour obtenir une décroissance logarithmique. Si ces appareils de mesure sont très répandus, en raison de leur simplicité d'emploi, ils n'enregistrent pas toujours la valeur de crête vraie du signal audio.

En conséquence, l'utilisation d'un instrument de mesure de valeurs d'échantillons de crête dans les cas où il est important de mesurer avec précision les valeurs de crête des programmes peut poser des problèmes. Malheureusement, la plupart des crêtemètres numériques sont des appareils de mesure de valeurs d'échantillons de crête, même si cela n'est généralement pas manifeste pour les opérateurs.

Ce problème tient au fait que les valeurs de crête vraies d'un signal échantillonné apparaissent généralement entre les échantillons, plutôt que de manière précise à un instant d'échantillonnage, de sorte qu'elles ne sont pas correctement enregistrées par l'appareil de mesure de la valeur d'échantillon de crête.

Il en résulte plusieurs anomalies en ce qui concerne les appareils de mesure de valeurs d'échantillons de crête:

- *Manque d'homogénéité des valeurs de crête affichées:* on constate fréquemment que la lecture répétée d'un enregistrement analogique dans un système numérique avec un appareil de mesure de valeurs d'échantillons de crête donne des valeurs très différentes des valeurs de crête obtenues à chaque lecture. De même, si l'on reproduit plusieurs fois un enregistrement numérique par l'intermédiaire d'un convertisseur de fréquences d'échantillonnage avant les mesures, les valeurs de crête enregistrées sont différentes à chaque lecture. En effet, les instants d'échantillon peuvent correspondre à différentes parties du signal vrai à chaque lecture.

* NOTE – Le texte reproduit ci-après à titre d'information a été présenté par l'AES Standards Working Group SC-02-01, par l'intermédiaire du Rapporteur pour le GT 6J des radiocommunications sur la mesure de l'intensité sonore.

- *Surcharges imprévues*: étant donné que les signaux échantillonnés peuvent contenir des surcharges même lorsqu'ils ne contiennent pas d'échantillons au niveau du signal numérique à pleine échelle, voire à proximité de ce niveau, l'indication de surcharge par un appareil de mesure des valeurs d'échantillons de crête n'est pas fiable. Les surcharges peuvent provoquer un écrêtage lors des processus ultérieurs, par exemple dans certains convertisseurs numériques/analogiques ou lors de la conversion de la fréquence d'échantillonnage, même si elles n'ont pas été enregistrées précédemment par l'appareil de mesure de valeurs d'échantillons de crête (et si elles étaient inaudibles lors du contrôle effectué à ce stade).
- *Affichage de valeurs inférieures et battement des tonalités mesurées*: les tonalités pures (les tonalités d'alignement, par exemple) proches de multiples entiers de la fréquence d'échantillonnage indiquent parfois une valeur inférieure ou peuvent indiquer une valeur variable en permanence, même si l'amplitude de la tonalité est constante.

Ce problème peut-il être grave?

En général, plus la fréquence du signal mesuré de l'échantillon de crête est élevée, plus le risque d'erreur est grand.

Pour les tonalités pures continues, on peut aisément démontrer, par exemple, que la valeur affichée est inférieure de 3 dB dans le cas d'une tonalité mise en phase par inadvertance au quart de la fréquence d'échantillonnage. La valeur inférieure ainsi indiquée pour une tonalité située à la moitié de la fréquence d'échantillonnage pourrait être pratiquement infinie, mais la plupart des signaux audionumériques ne contiennent pas beaucoup d'énergie à cette fréquence (celle-ci est en effet en grande partie exclue par les filtres de protection contre le repliement de spectre au point de conversion numérique/analogique et les sons (réels) ne sont généralement pas dominés par des signaux continus à hautes fréquences).

Les tonalités continues qui ne sont pas situées à proximité de petits multiples entiers de la fréquence d'échantillonnage ne donnent pas de valeurs inférieures sur les appareils de mesure des valeurs d'échantillons de crête, parce que la fréquence de battement (c'est-à-dire la différence entre $n \cdot f_{\text{tone}}$ et f_s) est élevée par rapport à la réciproque du taux de décroissance de l'appareil de mesure. Autrement dit, l'instant d'échantillonnage est suffisamment proche de la valeur de crête vraie de la tonalité pour que l'instrument de mesure n'affiche pas de valeur inférieure.

Toutefois, pour différentes transitoires, les valeurs inférieures indiquées ne sont pas masquées par ce mécanisme, si bien que plus le contenu en fréquence de la transitoire est important, plus le risque d'indication d'une valeur inférieure est grand. Il est normal, dans des conditions sonores «réelles», que les transitoires se produisent avec un contenu haute fréquence important, et il arrive fréquemment que les valeurs indiquées soient inférieures de plusieurs dB.

Etant donné que les sons réels ont généralement un spectre qui décroît en direction des fréquences les plus élevées et que cette situation n'évolue pas en fonction de l'augmentation de la fréquence d'échantillonnage, la valeur inférieure affichée sur l'appareil de mesure de la valeur d'échantillon de crête est moins importante aux fréquences d'échantillonnage initiales les plus élevées.

Quelle est la solution?

Afin de mesurer la valeur de crête vraie d'un signal échantillonné, il est nécessaire de «suréchantillonner» ou de «sous-échantillonner» le signal en recréant, pour l'essentiel, le signal d'origine entre les échantillons existants et en augmentant ainsi la fréquence d'échantillonnage du signal. Bien que cette proposition, à savoir comment recréer des informations qui ont apparemment déjà été perdues, semble peu judicieuse, la théorie de l'échantillonnage montre qu'il est possible de la mettre en œuvre, parce que nous savons que le signal échantillonné ne contient aucune fréquence au-dessus de la moitié de la fréquence d'échantillonnage initiale.

Quel rapport de suréchantillonnage faut-il prévoir? A cette fin, il faut répondre aux deux questions suivantes:

- Quelle est l'erreur maximale acceptable concernant l'affichage d'une valeur inférieure?
- Quel est le rapport entre la fréquence la plus élevée à mesurer et la fréquence d'échantillonnage («fréquence normalisée» maximale)?

Si nous connaissons ces critères, il est possible de calculer le rapport de suréchantillonnage dont nous avons besoin (même sans prendre en compte la question de la mise en œuvre du suréchantillonnage) à l'aide d'une méthode simple dite du «papier millimétré». Nous pouvons simplement examiner quelle valeur inférieure sera indiquée avec une paire d'échantillons à la fréquence suréchantillonnée qui se situe symétriquement de part et d'autre de la valeur de crête d'un signal sinusoïdal sur notre fréquence normalisée maximale. Il s'agit de la valeur inférieure affichée correspondant «au cas le plus défavorable».

En conséquence, pour:

un rapport de suréchantillonnage, n

une fréquence normalisée maximale, f_{norm}

une fréquence d'échantillonnage, f_s

nous constatons que:

la période d'échantillonnage à la fréquence suréchantillonnée est $1/n.f_s$

la période de la fréquence normalisée maximale est $1/f_{norm} . f_s$

et qu'en conséquence:

la valeur inférieure maximale indiquée (dB) est $20.\log(\cos(2.\pi.f_{norm}.f_s/n.f_s.2))$

(2 en dénominateur car nous pouvons omettre une valeur de crête correspondant au plus à la moitié de la période de suréchantillonnage)

ou:

la valeur inférieure maximale inférieure indiquée (en dB) = $20.\log(\cos(\pi.f_{norm}/n))$.

Cette formule a été utilisée pour élaborer le tableau suivant, qui s'applique probablement à la gamme de valeurs concernées:

Rapport de suréchantillonnage	Valeur inférieure maximale indiquée (dB) $f_{norm} = 0,45$	Valeur inférieure maximale indiquée (dB) $f_{norm} = 0,5$
4	0,554	0,688
8	0,136	0,169
10	0,087	0,108
12	0,060	0,075
14	0,044	0,055
16	0,034	0,042
32	0,008	0,010

Comment mettre en œuvre un instrument de mesure du niveau de crête vrai?

On procède au suréchantillonnage en insérant des échantillons de valeur nulle entre les échantillons d'origine, de façon à créer un flux de données au niveau de la fréquence suréchantillonnée voulue et en appliquant un filtre «d'interpolation» passe-bas pour exclure les fréquences supérieures à la valeur maximale souhaitée f_{norm} . Si nous utilisons à présent l'algorithme de calcul de valeurs d'échantillons de crête sur le signal suréchantillonné, nous disposons d'un appareil de mesure de la valeur de crête vraie ayant la valeur inférieure maximale voulue.

Il est intéressant d'examiner la mise en œuvre de ce suréchantillonneur. En général, on utilise le filtre passe-bas comme filtre FIR symétrique. Lorsque ces filtres servent à transmettre des sons de haute qualité, par exemple dans des convertisseurs numérique/analogique *avec* suréchantillonnage (obsolètes) ou dans des convertisseurs de fréquence d'échantillonnage, il est nécessaire de calculer un grand nombre de «valeurs» pour maintenir une très faible ondulation sur la bande passante et assurer un affaiblissement extrême dans la bande d'arrêt ainsi qu'une bande de transition étroite. Il faut également maintenir une longueur de mot importante pour préserver la plage dynamique et réduire au minimum les distorsions.

Toutefois, étant donné que nous n'allons pas écouter à la sortie de notre suréchantillonneur, mais uniquement l'utiliser pour afficher une valeur ou établir un diagramme à barres, nous n'aurons probablement pas les mêmes contraintes en matière de précision. Il s'agit de faire en sorte que l'ondulation sur la bande passante, associée aux composantes non essentielles provenant de la bande d'arrêt, ne dégradent pas la précision de lecture au-delà de l'objectif fixé. Cela réduit considérablement le nombre requis de mesures, même si l'on peut encore être amené à prévoir une bande de transition étroite, en fonction de l'objectif maximal fixé concernant la fréquence normalisée. De même, il faudra peut-être que la longueur de mots suffise juste à garantir la précision recherchée jusqu'à la base du diagramme à barres, sauf si une sortie numérique précise est nécessaire aux faibles amplitudes.

En conséquence, il se peut qu'un suréchantillonneur approprié (éventuellement pour de nombreux canaux) puisse être facilement mis en œuvre dans un processeur DSP ou un réseau logique programmable (FPGA) classique de coût modique, voire dans un processeur encore plus simple. Par ailleurs, on a mis en œuvre des appareils de mesure de suréchantillonnage en utilisant des puces de suréchantillonnage de haute précision destinées à être utilisées pour la conversion numérique/analogique. Bien qu'ils comportent un grand nombre de circuits intégrés et consomment beaucoup d'énergie, ces dispositifs sont bon marché et facilement accessibles.

La meilleure manière de déterminer le nombre requis de mesures et leurs coefficients pour une spécification donnée d'un appareil de mesure consiste à utiliser un programme de conception de filtre FIR récursif, par exemple celui de Remez ou Meteor.

Par ailleurs, il peut être nécessaire que le crêtemètre exclue les effets d'une composante DC à l'entrée, étant donné que cette composante est généralement bloquée à l'entrée de l'audiomètre. De plus, si l'on veut connaître la valeur du signal de crête vrai pour éliminer toute surcharge, il faut maintenir et mesurer le contenu DC. Le cas échéant, on peut exclure la composante DC moyennant une puissance de calcul réduite, en plaçant un filtre passe-haut IIR de second ordre à l'entrée de l'appareil de mesure.

On est parfois amené à mesurer l'amplitude du signal de crête après application d'un filtre de pondération, afin d'accentuer les effets de certaines parties de la bande de fréquences. La réalisation dépend de la nature du filtre de pondération considéré.