

DYNAMIQUE DES MIXAGES POUR LA DIFFUSION TELEVISUELLE : UNE NOUVELLE DEFINITION...

Anaïs LIBOLT

LA SITUATION ACTUELLE

A l'écoute de la télévision, nous constatons tous régulièrement que certains programmes sont plus forts que d'autres... Nous constatons également que certains programmes présentent d'importantes variations de niveau qui nous poussent à ajuster le volume sonore en permanence. Tout cela présente de nombreux inconforts pour le téléspectateur. A la racine de ces phénomènes, on trouve principalement deux raisons :

1. Les instruments de mesure du niveau sonore utilisés jusqu'à présent ne reflètent pas toujours notre perception de l'intensité sonore
2. Il n'existait pas de norme ou de recommandation qui délimite la dynamique sonore acceptable pour un programme.

Pour parer à cette lacune, les diffuseurs français, la CST et la FICAM, en collaboration avec quelques mixeurs, ont établi récemment une recommandation qui a pour but de donner des repères et des guides pour la fabrication des mixages destinés à la diffusion télévisuelle. Une recommandation européenne est également à venir, mais le groupe de travail français qui s'intéresse à cette problématique a souhaité dès à présent mettre en application le fruit de ses réflexions...

La **diffusion télévisuelle** impose plusieurs **contraintes** pour assurer une continuité d'antenne. Avant tout, le diffuseur se doit de **garantir une intelligibilité et un confort d'écoute** satisfaisants pour un spectateur écoutant dans des conditions grand public, d'un bout à l'autre du programme. Les programmes qui s'enchaînent doivent également **raccorder en niveau**. Enfin, nombre de mixages étant destinés à être encodés en *Dolby E / Dolby Digital*, il faut **tenir compte de l'action des métadonnées Dolby** (*Dialogue Level, Dynamic Range Control, downmixing*). Une bonne gestion de ces métadonnées aide à répondre aux contraintes évoquées ci-dessus, en influant sur la restitution sonore chez le spectateur de façon maîtrisée.

LA DYNAMIQUE ET LE MIXAGE DES DIALOGUES

1° : Le mixage des dialogues détermine la dynamique

La gestion des dialogues par le mixeur est déterminante pour la dynamique du programme. Par exemple, on peut constater que dans un message publicitaire, la voix de l'annonceur est forte et que l'on a de ce fait peu de dynamique autour. A contrario, dans un film d'action typique, on trouvera une variété de scènes allant de la scène intimiste pour laquelle il faut tendre l'oreille à la scène de bagarre qui nous en met plein la tête!...

La valeur clé pour un programme est le **niveau moyen de ses dialogues**. Cette valeur est comme un **"centre de gravité"**, autour duquel s'organisent tous les autres éléments sonores...

NB : Par dialogue, on entend communément les voix. Cependant, d'une façon élargie, on pourrait définir ce terme comme désignant les éléments sonores qui focalisent l'attention de l'auditeur au point de lui faire modifier son niveau d'écoute si besoin. Il s'agit donc avant tout des éléments porteurs de sens pour la compréhension du programme. [En Anglais, on parle de *foreground sound* an opposition à *background sound*].

1 L'expression "*center of gravity*" est également utilisée par TC Electronics pour désigner le *loudness* moyen du programme, sans référencement au dialogue. Elle n'a donc pas exactement le même sens ici.

2°: Niveau moyen du dialogue (Dialogue Level)

Le **loudness** d'un programme se définit comme le **niveau moyen perçu** de ce programme. C'est une mesure du signal complétée par une pondération qui modélise la sensibilité de l'oreille humaine. La valeur trouvée est donc en mesure de nous indiquer la sensation de « fort » que l'on va avoir à l'écoute d'un son. Le niveau moyen ou **loudness** du programme correspond à l'intégration d'une mesure sur tout le programme (mesure *Long Term*²). On utilise maintenant la pondération R2LB définie par l'ITU dans la norme ITU-R BS.1770 pour effectuer cette mesure. La valeur obtenue s'exprime en LKFS³ (ou en LU⁴).

Le **Dialogue Level** est un cas particulier du **loudness**. C'est une **mesure de loudness Long Term qui porte uniquement sur les dialogues**. Cela donne donc le niveau moyen du dialogue dans le programme.

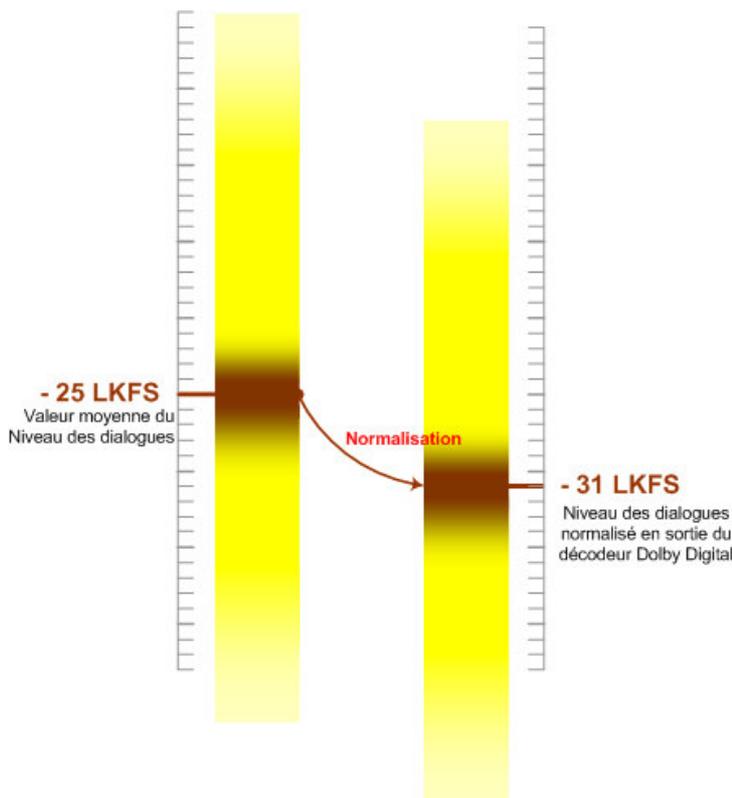


Figure 1 : Normalisation d'un programme à -31 dBFS. Le Dialogue Level est mesuré à -25 LKFS et normalisé en sortie du décodeur *Dolby Digital*.

Pour la diffusion en *Dolby Digital*, on résout le problème du raccord en niveau entre programmes par l'alignement du dialogue à -31 LKFS au décodage *Dolby Digital*. Pour que cela soit possible, la valeur charnière du *Dialogue Level* doit être correctement renseignée dans les métadonnées (voir méthode donnée dans la recommandation RT 016). Cette valeur est ensuite utilisée par le décodeur *Dolby Digital* pour traduire l'ensemble du programme et aligner le niveau de dialogue à -31 dBFS. On peut donc se poser la question de la nécessité de spécifier une plage de valeurs de *Dialogue Level* autorisées ?

Cependant, certaines **valeurs de Dialogue level** sont **symptomatiques de la dynamique du mixage**. Par exemple -31 LKFS tend à caractériser des mixages ayant une grande dynamique. À l'opposé, -20 LKFS indique une faible dynamique, et un mixage parfois un peu trop dense.

⇒ En spécifiant une plage de valeurs de référence, on peut aider à maîtriser la dynamique.

Adopter une **valeur cible de -25 LKFS** est un bon compromis. Comme tous les programmes ne nécessitent pas la même dynamique, on peut accepter que le *Dialogue Level (Long Term)* s'écarte de cette valeur, avec une tolérance de +/- 2 dB, donc **de -27 à -23 LKFS**. [Attention, cela ne signifie pas que l'on aurait 4 dB de dynamique en dialogue !]

² *Long Term* est un mode de mesure qui intègre tout l'audio reçu par le *loudness meter* entre le début et la fin de la mesure. Pour un film de 2h, la mesure *long term* fait la moyenne des 2 heures.

³ FS indique que l'on exprime la valeur sur une échelle *Full Scale*. LK indique une pondération selon la norme ITU-R BS.1770. LKFS signifie *Loudness K weighted*, relative à une échelle *Full Scale*. Une valeur en LKFS est l'expression d'une mesure pondérée K (Leq(R2LB)) sur une échelle *Full Scale*. On peut également trouver la mention LUFS...

⁴ LU : *Loudness Unit* ou unité de volume sonore. Echelle graduée en dB, décrite dans la Rec. BS.1771 comme allant de -21 à +9. Cette échelle reste encore à définir. La relation du 0 LU à l'échelle LKFS n'est pas fixée. 0 LU est considéré comme le niveau moyen cible et donc il peut être associé à sa valeur en LKFS. L'EBU travaille sur ces points actuellement.

3°: Mesurer la dynamique :

Nous avons vu que nous pouvons définir une sorte de centre de gravité pour un programme, en effectuant une mesure moyenne sur la totalité du programme. Maintenant, pour connaître et quantifier l'excursion dynamique, il faut une mesure plus courte. La mesure de *loudness* dite **Short Term**⁵ permet de visualiser la valeur du *loudness* sur les 10 dernières secondes. C'est une sorte de Vumètre qui donne un **affichage dynamique du *loudness***. Par exemple, sur un passage dialogué, les valeurs vont évoluer autour de la valeur de *Dialogue Level*, avec une plus ou moins grande excursion... Ce qui nous amènera à apprécier la dynamique des dialogues, sur la base de cette indication de *loudness Short Term*. Dans les **passages sans dialogue**, le *loudness Short Term* présente des excursions plus grandes. C'est ce qui caractérise la **dynamique du programme**.

La **dynamique du programme dépend** donc avant tout **du niveau où sont placés les dialogues**, donc du *Dialogue Level*. On appellera « **réserve** » l'excursion de niveau possible entre le niveau moyen du dialogue et le Niveau maximal permis.

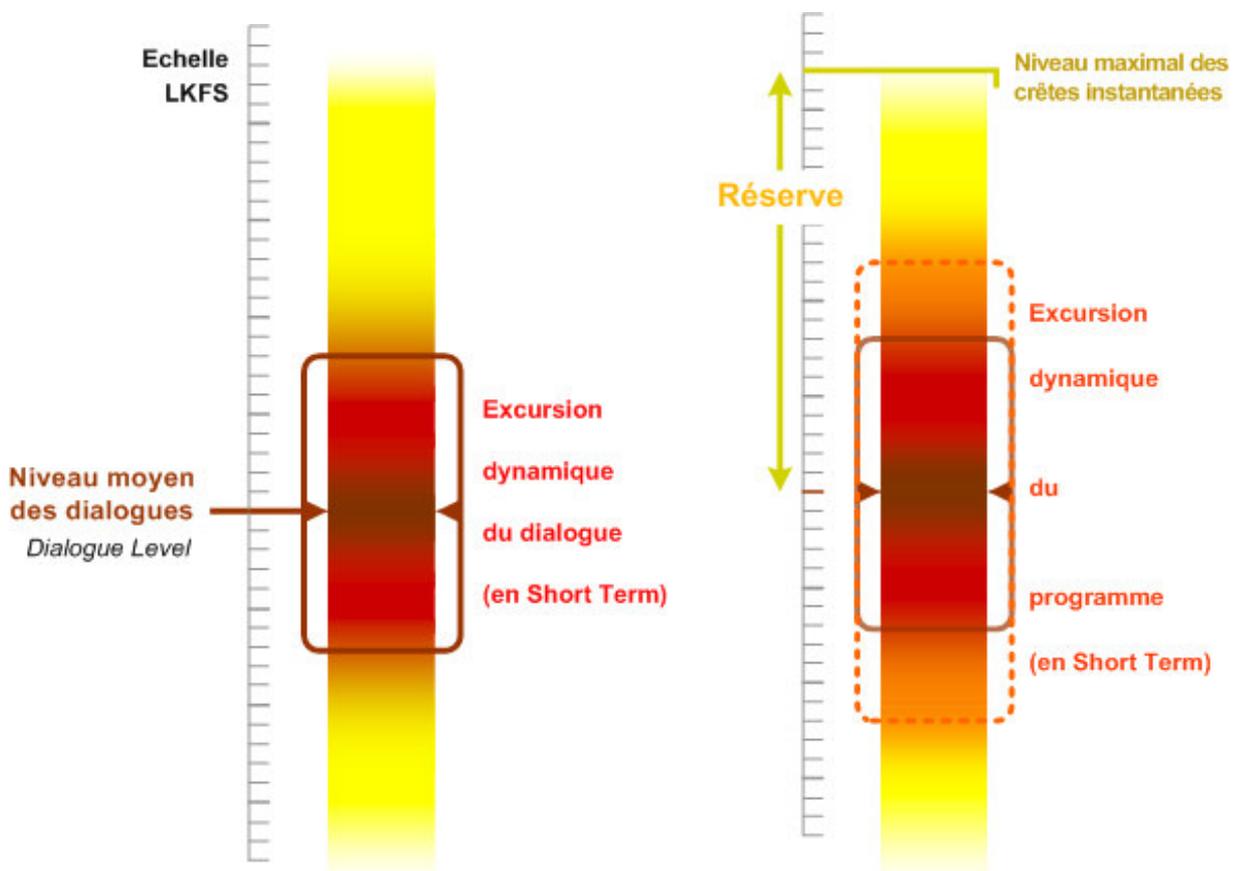


Figure 2 : La valeur moyenne du dialogue d'un programme est comme son "centre de gravité". Les excursions du dialogue (visualisées sur un *loudness meter en short term*) évoluent dans la fenêtre. Choix des couleurs : plus la teinte est foncée, plus il y a de "densité sonore" à cet endroit. La couleur la plus foncée est pour le niveau moyen et la plus claire pour les crêtes. Une double fenêtre représente les excursions 1) du dialogue 2) reste du programme.

Au final, le **mixage pourra être caractérisé** par la **valeur moyenne du dialogue** (mesurée à postériori), mais aussi par la **dynamique** donnée à ce dialogue. Cette dynamique peut être définie comme l'intervalle entre les *loudness* min et max atteints par les dialogues (à peu près, évidemment...).

⁵ *Short Term* fait référence au mode utilisé par les outils Dolby pour afficher le *loudness* moyen des 10 dernières secondes. Il s'agit donc d'une fenêtre glissante de 10 secondes. Cette appellation se normalise pour désigner la valeur de *loudness* mesurée sur une fenêtre glissante de quelques secondes.

La recommandation RT-019 définit la dynamique du mixage en s'appuyant sur des valeurs en *loudness Short Term*. On y distingue l'excursion autorisée par les dialogues de celle autorisée pour l'ensemble du programme :

- Pour les **dialogues**, une plage de **14 dB de dynamique** mesurée sur un *loudness meter* en *Short Term* suffit largement. On peut visualiser cette plage centrée sur la valeur de *Dialogue Level* compressée. [N'oublions pas qu'un indicateur de niveau à temps d'intégration plus court indiquera des excursions plus importantes.]
- Pour le **mixage, hors dialogue, 24 dB de dynamique** (+/-12) mesurés sur un *loudness meter* en *Short Term* permettent au mixage de conserver de l'ampleur tout en restant confortable pour un téléspectateur. [De même, un indicateur de niveau à temps d'intégration plus court indiquera des excursions plus importantes, d'où l'existence d'un niveau maximal pour les crêtes...]

4°: Crêtes

a) Le 5.1

A priori, il n'y aurait aucun besoin de limiter les crêtes sur le 5.1 puisque l'on n'a plus les contraintes des émetteurs de la TV analogique. Cependant, beaucoup préconisent une **limitation à - 3 dBFS pour éviter** une éventuelle **distorsion** dans la chaîne de transmission. Par ailleurs, le mixage risque de subir une réduction stéréo (*downmix*) chez le spectateur. Cette marge rend plus aisée l'opération de downmixing.

⇒ Limiter à - 3 dBFS semble être le minimum afin d'éviter tout problème de distorsion.

[Il serait également possible d'exprimer ce niveau crête en dB True Peak (dBTP), comme défini dans l'ITU-R BS.1770, mais nous attendrons les travaux conjoints de l'EBU et de l'ITU pour cela...]

b) La stéréo :

Les mixages stéréo font l'objet d'un traitement spécial car ils peuvent soit être encodés en *Dolby Digital*, soit être diffusés tels quels (ou après traitement antenne). Par soucis de compatibilité avec les anciennes recommandations sur la stéréo, les mêmes niveaux max ont été retenus, à savoir **0 dB sur un indicateur PPM 10ms** (DIN 45406). [Notons que les crêtes instantanées peuvent dépasser de 4 dB la valeur quasi-instantanée.]

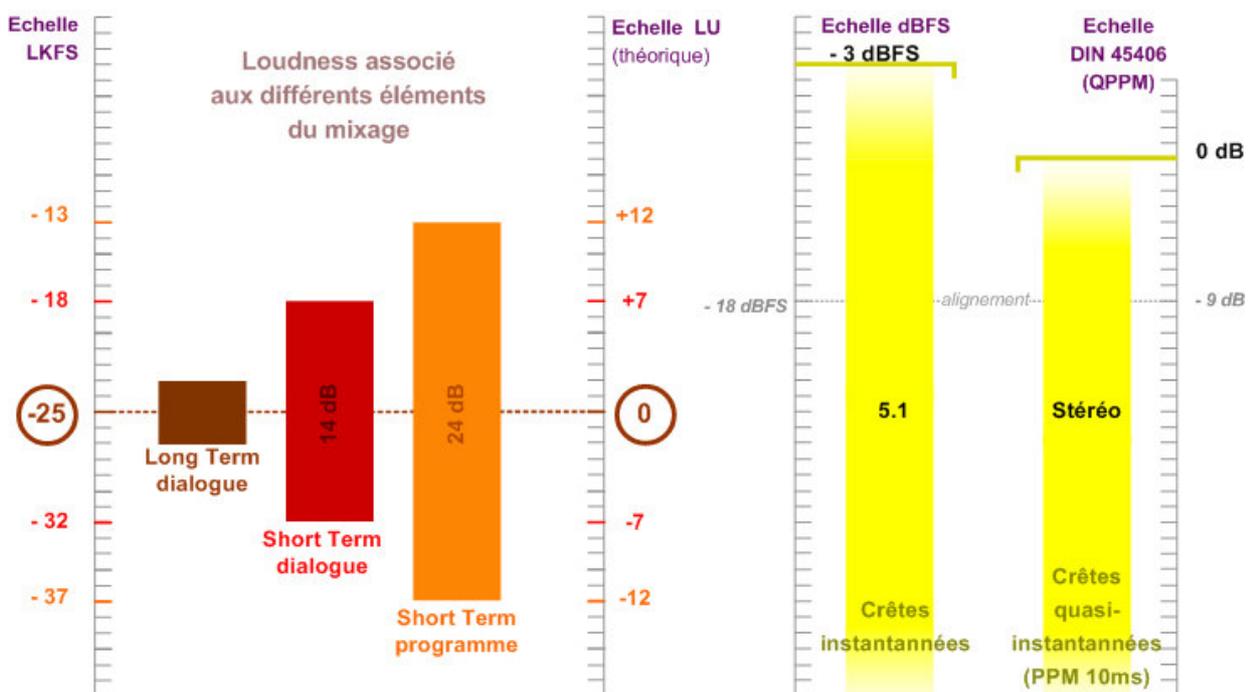


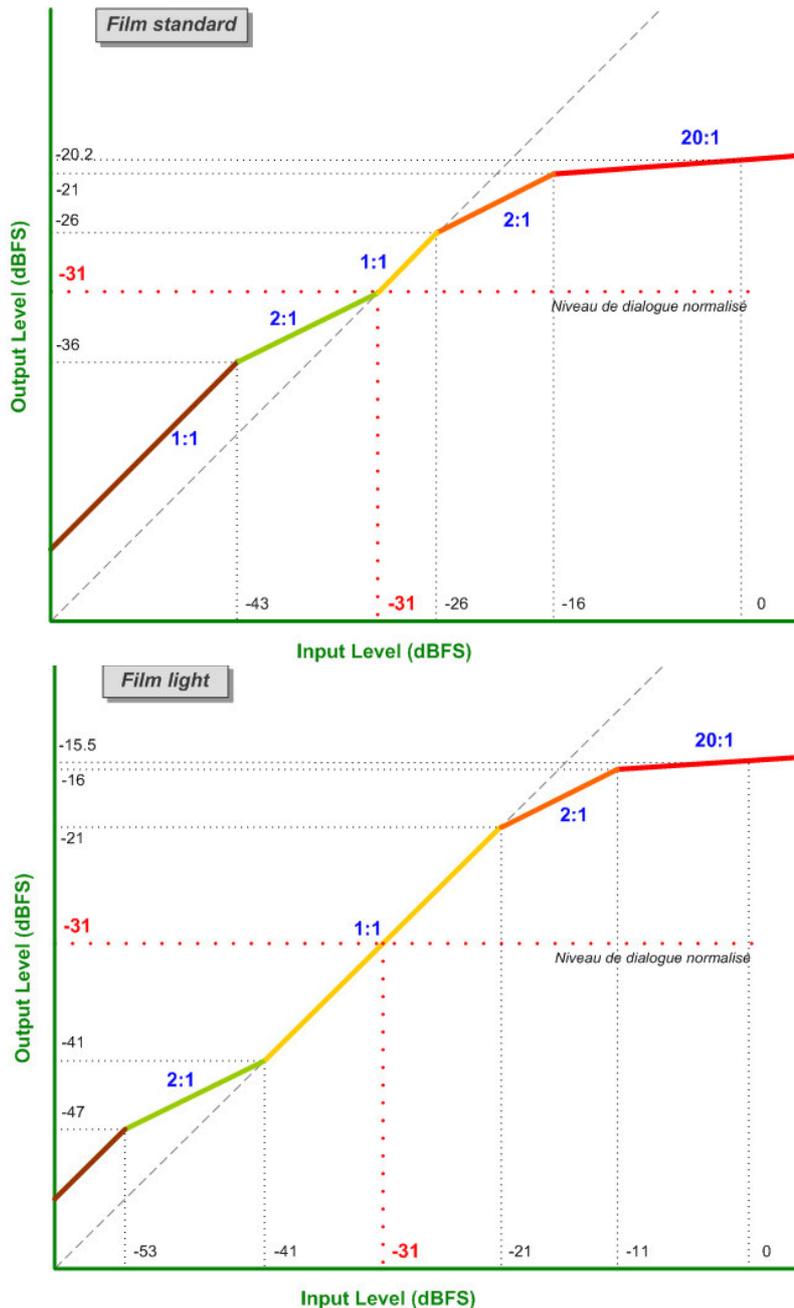
Figure 3 : Décomposition des éléments du mixage et mesure associée.
A gauche les mesures de loudness, à droite les mesures de crête.

GESTION DE LA DYNAMIQUE PAR LES METADONNEES DOLBY

1° : Les profils DRC (Dynamic Range Control)

Le **DRC⁶ (Dynamic Range Control)** est une fonction de type compresseur proposée par les décodeurs Dolby Digital, servant à assurer une **compatibilité dynamique** avec toutes sortes d'installations sonores. Seuls les amplis plutôt haut de gamme permettent de désactiver l'application des *DRC*. La plupart des auditeurs écoutent donc avec les *DRC* enclenchés. L'idéal est alors que cette compression soit là comme une « sécurité » et non comme un traitement destructeur. L'objectif inhérent à la maîtrise de la dynamique des mixages télé (5.1 mais aussi stéréo) est avant tout de préserver l'équilibre du mixage lors du décodage *Dolby Digital*. C'est-à-dire que l'on va essayer d'éviter que les *DRC* ne touchent à l'équilibre du mixage en anticipant leur comportement.

2° : Profil Dynamique



La métadonnée Dolby du DRC permet de choisir entre plusieurs profils de compression. Le **profil Film Light** permet 20 dB de dynamique sans aucune compression, plus 10 dB et 12 dB de dynamique de part et d'autre avec une « légère » compression, ce qui donne environ 40 dB de dynamique. C'est ce profil qui est **retenu pour les mixages télé**, le profil Film standard étant trop contraignant (voir figure 4).

Le mode bien connu sous le nom de *Night Mode* applique une forte compression tirée des *DRC*. C'est un mode particulier qui ne doit pas être considéré comme la normalité. Dans ce mode, fatalement, la dynamique sera fortement réduite et le mixage modifié... Nous nous pencherons donc plutôt sur le mode qui s'appelle *LINE* dans les équipements Dolby pour contrôler la compression appliquée dans la plupart des cas.

Figure 4 : Les profils DRC Film tels qu'ils s'appliquent autour du dialogue préalablement normalisé à -31 LKFS

⁶ *DRC* pour *Dynamic Range Control*. C'est la métadonnée qui permet un traitement dynamique du son par le décodeur *Dolby Digital*. Cette compression est là avant tout pour rendre tous types de mixages compatibles avec toutes sortes d'installations grand public. Il va de soi qu'il faut avoir préparé un minimum le terrain pour obtenir une restitution optimale...

3°: Le downmixing

Le *downmixing* est l'opération qui vise à faire une réduction du mixage. Généralement, on appelle *Downmix* la réduction stéréo d'un mixage 5.1. C'est une opération extrêmement fréquente dans la mesure où de nombreux auditeurs ont - au mieux - un système stéréo alors qu'ils reçoivent un mixage 5.1 en *Dolby Digital*.

Les *DRC* qui servent lors de l'écoute du 5.1, servent aussi lors de la création du *downmix* stéréo. La réduction de 5 canaux vers 2 peut se passer plus ou moins facilement selon le niveau présent sur ces canaux. Par exemple, si les 5 canaux sont au maximum, lors du *downmix* il est impératif de réduire le niveau pour éviter de saturer le signal de sortie. Les métadonnées Dolby sont là pour gérer ce genre de situation, encore faut-il les aider. Si la dynamique reste raisonnable, tout se passera bien. Dans le cas contraire, on pourrait entendre un résultat désagréable...

MISE EN PRATIQUE

1°: Niveau et conditions d'écoute :

La création d'un mixage télé et d'un mixage cinéma diffèrent sur de nombreux points. On peut citer d'une part la taille et la proximité des écoutes, mais également le niveau d'écoute. Ce dernier influe beaucoup sur le résultat final, car le mixeur va gérer différemment l'équilibre selon le niveau et les conditions d'écoute.

Au cours du mixage, chacun place naturellement les dialogues au niveau qui lui convient et construit son mixage sur cette base (ou sur tout autre élément en cas d'absence de dialogue...).

La recommandation RT-017 pour la fabrication des PAD donne comme niveau de référence pour la **vérification des programmes**, la valeur de **74 dB (C) SPL**. Cette valeur peut, bien entendu, sembler un peu faible pour permettre un travail de mixage. De l'avis de tous, les meilleurs résultats pour la création d'un mix TV sont obtenus bien plus facilement et efficacement dans une salle de petit volume avec des écoutes adaptées. Les valeurs qui se dégagent pour le **niveau d'écoute** tournent autour des **79 dB (C) SPL** et peuvent aller un peu au-dessus. Une valeur de 85 dB (C) SPL généralement utilisée en mixage film augmente les risques de sortir de la plage dynamique confortable en télévision... Quelque soit la taille de la salle, une écoute de proximité est indispensable pour se mettre dans des conditions adaptées à la télévision.

Les conditions de mixage idéales réunissent une pièce avec des dimensions et une écoute adaptées, mais également la vu-métrie et les outils de contrôle adéquats. Le PPM DIN 45406 largement répandu en télévision ne permet pas de juger du *loudness* du programme, puisqu'on peut, pour une même modulation, avoir plus de 10 dB d'écart en *loudness* moyen! Pour un mixeur qui prend ses marques dans une installation comme pour tous ceux qui souhaitent avoir un repère fiable, le ***loudness meter* devient un outil indispensable**. Des équipements permettant de simuler l'action des métadonnées sont également précieux afin d'avoir conscience du rendu final. Dans l'exemple ci-dessous, il est question du DP570⁷ de Dolby.

2°: Quelques exemples pratiques :

Lors du mixage, il sera certainement nécessaire ponctuellement d'**écouter avec les DRC enclenchés** par moment afin de **vérifier que tout se passe bien**. Hors, les *DRC* s'appliquent autour du dialogue. Il faut donc donner une valeur de *Dialogue Level* au DP570 pour pouvoir effectuer la simulation. Si on indique -25, c'est une base de travail qui permet de commencer le mixage. Si, au fil de la séance il ressort que la vraie valeur moyenne du dialogue est différente de

⁷ DP570 : C'est l'équipement de Dolby qui permet la création et le contrôle de l'application des métadonnées. Une application permet de télécommander l'appareil via un PC et de visualiser certains paramètres.

-25, on pourra ajuster. La marge de +/-2 dB (-27 à -23) fait qu'on se retrouvera certainement dans les valeurs ciblées. Par ailleurs, il est également possible d'ajuster le niveau d'écoute afin de naturellement mixer les dialogues plus haut ou plus bas...

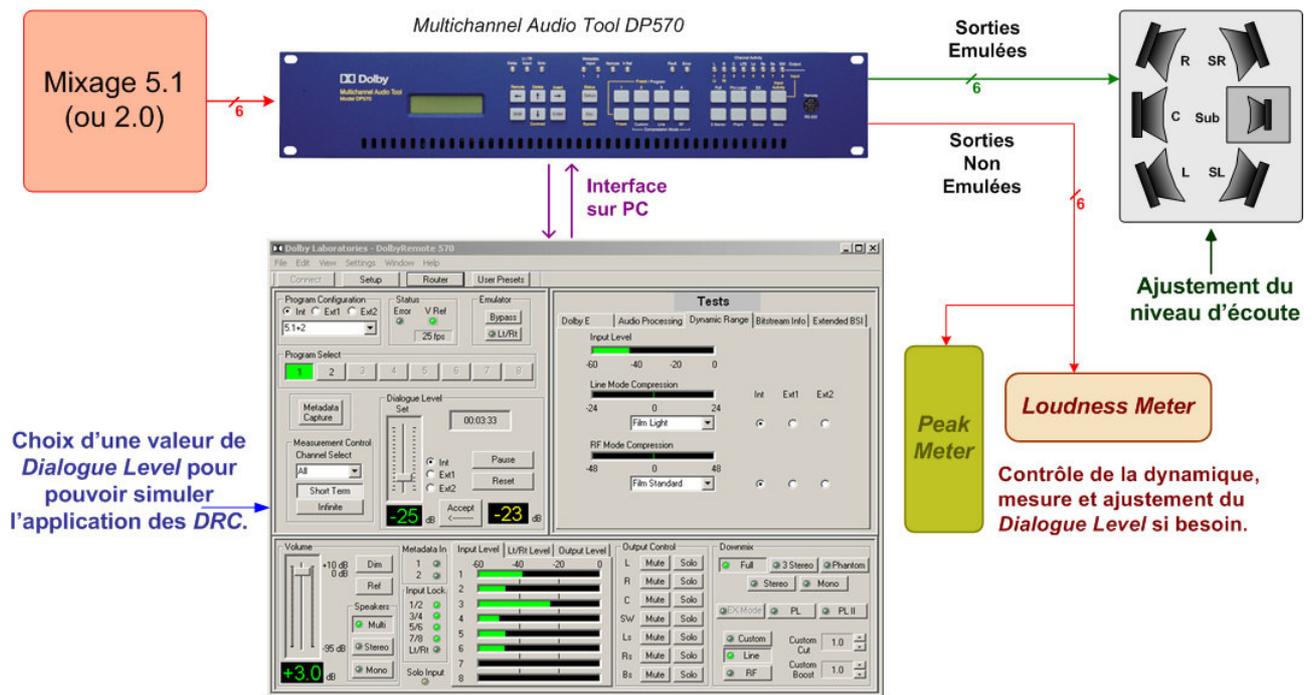


Figure 5 : Exemple de configuration de travail pour simuler un décodage Dolby Digital avec l'application des DRC. Il est nécessaire de choisir sur le DP570 un niveau moyen de dialogue pour déterminer où s'appliqueront les DRC. La valeur pourra être ajustée au cours du mixage si elle ne correspond pas au niveau réel des dialogues. L'ajustement du niveau d'écoute influence également sur le placement des dialogues et donc leur niveau moyen.

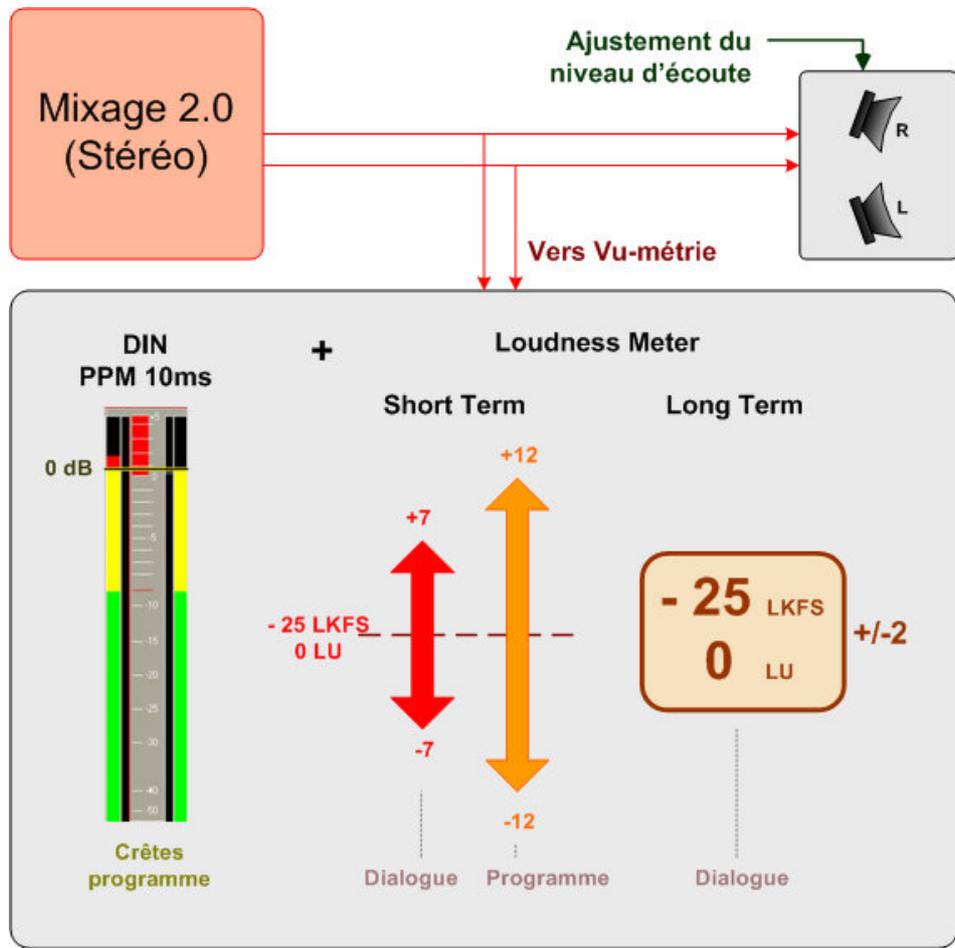


Figure 6 : Détail de la vue-métrie adéquate lors du mixage d'un programme stéréo. Cet exemple est également valable pour le 5.1 avec l'indicateur de crête correspondant.

Un point de contrôle important à ne pas oublier est la **cohérence du downmix stéréo** obtenu depuis un mixage 5.1. On rappelle qu'une grande partie des téléspectateurs écoutera ce *downmix*... La dynamique du mixage 5.1 doit donc être adaptée à une réduction stéréo, ce qui devrait être le cas avec cette recommandation.

Pour les mixages qui n'ont pas vocation à être encodés en *Dolby Digital*, nul besoin de simuler l'action des métadonnées. Par contre, la mesure du niveau moyen des dialogues reste fondamentale pour permettre un raccord en niveau lors de la diffusion. Le crête-mètre donne les limites électriques du signal, mais les indicateurs à ne pas perdre de vue sont ceux du *loudness meter*. Sans non plus mixer "à l'œil", cet indicateur se révèle utile pour prendre ses repères et s'assurer que l'on est toujours dans la bonne fenêtre lors de passages délicats.

CONCLUSION

Il est devenu indispensable d'effectuer nos mesures de niveau sonore en *loudness*, si on veut remédier à de nombreux désagréments. Cela implique de **nouvelles habitudes de travail** et il faut du temps pour se forger des **repères**. Les différents acteurs de la chaîne de production doivent donc apprivoiser cette méthode, chacun à leur niveau.

Tout ce travail profitera à la diffusion télévisuelle, mais devrait également profiter en amont à la **sortie du DVD** qui présente la même problématique. Cette version étant pérenne, c'est un gain pour tous qu'elle soit vraiment adaptée à une écoute domestique. On aura ainsi des mixages écoutables en *Home Cinema* (et pas seulement en *Cinema at Home*), mais qui conservent plus d'ampleur que ce que nous avons par le passé en TV. Bien entendu, la version pleine dynamique restera disponible pour les puristes sur les supports adaptés...

A l'échelle de l'Europe (EBU), un groupe de travail (P/Loud) se penche assidûment sur une recommandation qui définirait un niveau de *loudness* pour les programmes ainsi qu'une dynamique appropriée. (Le niveau cible tend plutôt vers -23 LKFS à l'heure actuelle.) Les réflexions se dirigent dans le sens de cette recommandation française, qui s'adaptera certainement le moment venu aux standards européens... Ce même groupe travaille également à la spécification d'un *loudness meter* universel, pour les programmes avec ou sans dialogue. Des résultats sont attendus très prochainement...

Ce document est adapté d'un article paru en juin 2009 dans Réalisason
www.realisason.com

BIBLIOGRAPHIE :

- ITU-R BS.1770-1 et 1771
- CST-RT-016, 017 et 019
- *Intelligent program loudness measurement and control*, Jeffrey Riedmiller, Charles Robinson, Steve Lyman, AES 2003
- *Levelling and loudness in radio and television broadcasting*, G. Spikofski & S. Klar, EBU TR, 2004
- *Specifying Audio for HD*, Thomas Lund, NAB 2007
- *Level and distortion in Digital Broadcasting*, T. Lund, EBU TR Avril 2007
- *Les incontournables secrets de la fabrication du Dolby E...*, Anaïs LIBOLT, Octobre 2006, Réalisason N°44.
- *Loudness Descriptors to Characterize Programs and Music Tracks*, Skovenborg & Lund, Oct 2008, AES 125th.