

# Méthodes de prétraitement pour les systèmes de codage vidéo linéaire basé SoftCast

Anthony TRIoux, François-Xavier COUDOUX, Patrick CORLAY, Mohamed GHARBI

UMR 8520 - IEMN, DOAE

Univ. Polytechnique Hauts-de-France, CNRS, Univ. Lille, YNCREA, Centrale Lille  
F-59313 Valenciennes, France

{anthony.trioux, francois-xavier.coudoux, patrick.corlay, mohamed.gharbi}@uphf.fr

**Résumé** – Un nouveau schéma de codage vidéo linéaire dénommé SoftCast a démontré ces dernières années un potentiel élevé pour la diffusion de vidéo sur des canaux sans-fil variant dans le temps. Particulièrement utile en situation de broadcast, il permet d'éviter les problèmes rencontrés dans les systèmes de diffusion vidéo sans-fil conventionnels comme le cliff-effect [1]. Dans [2], nous avons analysé et comparé deux méthodes simples de prétraitement agissant dans le domaine pixel permettant d'accroître la qualité reçue. Dans cet article, nous évaluons deux nouvelles méthodes de prétraitement agissant dans le domaine fréquentiel : OPA-SoftCast récemment proposée dans [3] et une méthode intermédiaire originale proposant le meilleur compromis performance/consommation additionnelle de bande passante. Les méthodes sont comparées et évaluées pour différents rapports signal/bruit de canal à l'aide de deux métriques : le PSNR et l'index SSIM. Les résultats de simulation mettent clairement en évidence l'importance du bloc de prétraitement dans un schéma de transmission vidéo sans-fil SoftCast. En fonction des caractéristiques de la vidéo et du prétraitement choisi, une amélioration du PSNR jusqu'à 4.75dB et du SSIM jusqu'à 0.1 peut être observée.

**Abstract** – An original wireless video transmission scheme called SoftCast has been recently proposed to deal with the issues encountered in conventional wireless video broadcasting systems such as cliff effect [1]. In [2], we analyzed and compared two simple pixel-preprocessing methods that help to increase the received video quality. In this paper, we evaluate two new preprocessing method operating in the frequency domain: OPA-SoftCast recently proposed in [3] and an intermediate original method offering the best compromise performance/additional bandwidth consumption. Methods are compared and evaluated under different channel signal-to-noise ratios using two metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM). Simulation results clearly highlight the importance of the preprocessing block in a SoftCast wireless video transmission scheme. Depending on the video input characteristics and the preprocessing method used, an improvement of the PSNR score up to 4.75dB and SSIM index up to 0.1 can be observed.

## 1 Introduction

La transmission de vidéos vers et depuis les utilisateurs mobiles est un service en plein essor. Les opérateurs de réseaux prévoient que la vidéo représentera 82% de tout le trafic de données dans les années à venir [4]. Par conséquent, un grand effort de recherche est consacré à la conception de systèmes de transmission permettant d'obtenir la meilleure qualité vidéo à la réception lorsque la bande passante est limitée. Cependant, ceci est particulièrement difficile lorsque les caractéristiques du réseau et du canal changent rapidement au fil du temps.

Les solutions de transmissions vidéo traditionnelles consistent en un codage vidéo numérique (H.264/AVC [5], HEVC [6]) sur des protocoles sans-fil appropriés (802.11, 4G, etc.). Néanmoins, cette approche nécessite une adaptation permanente des paramètres de codage source/canal par l'émetteur. En raison de l'hétérogénéité des canaux de chaque utilisateur, les récepteurs dont les conditions de canal sont dégradées seront soumis à des perturbations visuelles importantes (e.g. gel d'image) alors que les récepteurs disposant d'un meilleur canal que celui initialement prévu ne peuvent pas en tirer pleinement parti. Le codage vidéo scalable (H.264/SVC) peut en partie résoudre

ce problème, mais une approche plus radicale a été récemment proposée : le codage vidéo linéaire (CVL). SoftCast [7] représente le pionnier des systèmes de CVL où les pixels sont traités par des opérations linéaires et directement transmis sans quantification ni codage (entropique ou de canal). Ceci permet d'obtenir une qualité de vidéo reçue qui augmente directement de manière proportionnelle à la qualité du canal de transmission, sans aucune information de retour tout en évitant les mécanismes d'adaptation complexes des schémas classiques.

Dans cet article, nous comparons trois méthodes de prétraitement issues de la littérature [3, 8, 9] et proposons une méthode intermédiaire offrant le meilleur compromis entre performance et consommation additionnelle de bande passante. Par prétraitement, nous entendons la réduction de l'énergie avant l'application de l'allocation de puissance [7] dans le schéma SoftCast. Nous commençons par une description de SoftCast suivie d'une analyse des 4 méthodes de prétraitement. Celles-ci permettent d'augmenter significativement la qualité vidéo reçue et soulignent l'importance du bloc de prétraitement dans le schéma SoftCast. Le choix de la méthode de prétraitement se fera en fonction de l'application et la bande passante disponible.

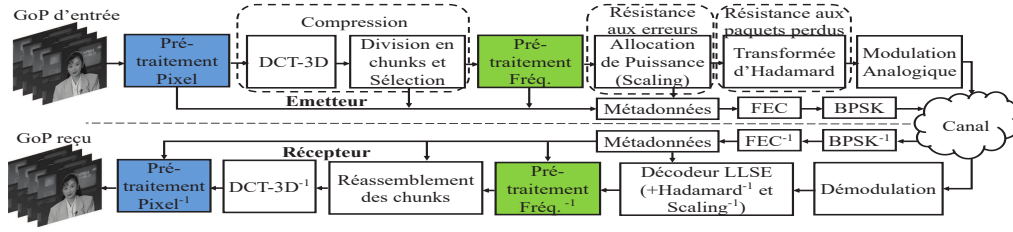


Figure 1: Diagramme bloc du schéma de transmission vidéo SoftCast

## 2 Aperçu de SoftCast

Différent des standards de transmission vidéo actuels, SoftCast est un système de codage conjoint source-canal ou Joint Source-Channel Coding scheme (CCSC/JSCC) qui transforme les pixels par des opérations linéaires successives. Ceux-ci sont ensuite transmis directement sans quantification ni codage. De fait, il permet d'obtenir une dégradation progressive de l'image (graceful degradation [7]) : la qualité reçue est une fonction linéaire du bruit du canal [10].

Le schéma de base de SoftCast [7] est présenté dans la figure 1. Les blocs colorés en bleu et vert représentent respectivement les méthodes additionnelles de prétraitement agissant soit dans le domaine pixel ou fréquentiel.

SoftCast transforme tout d'abord des groupes d'images (GoP) via une DCT en trois dimensions. Les plans DCT résultants sont ensuite divisés en blocs rectangulaires de coefficients transformés appelés *chunks* dont certains seront supprimés afin de répondre aux limitations de bande passante disponible. Une illustration de cette étape est disponible en figure 2.

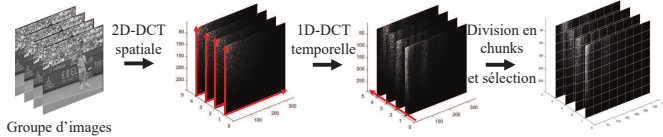


Figure 2: Étape de compression dans le schéma SoftCast

En parallèle du processus de transformation, l'émetteur SoftCast envoie des métadonnées. Celles-ci sont constituées de la moyenne et de la variance des *chunks* transmis, ainsi que d'une carte binaire, qui indique les positions des *chunks* ignorés dans le GoP.

Les métadonnées sont fortement protégées et transmises de manière robuste (BPSK par exemple [11]) pour assurer une réception correcte. En revanche, les coefficients transformés sont transmis via une modulation quasi-analogique [7].

Au récepteur, un décodeur LLSE (Linear Least Square Error) est utilisé afin d'obtenir la meilleure estimation des coefficients DCT reçus. En utilisant les métadonnées et les opérations inverses du décodeur, ces coefficients sont ensuite réassemblés pour reformer les plans DCT. Ces derniers sont ensuite transformés via un processus DCT-3D inverse.

Dans la section suivante, des méthodes de prétraitement améliorant significativement la qualité en réception sont présentées.

## 3 Méthodes de prétraitement

### 3.1 Prétraitement dans le domaine pixel

Les deux méthodes de prétraitement dans le domaine pixel (indiquées en vert dans la figure 1) sont définies comme suit :

$$f_{proc}(i, j) = f(i, j) - \bar{p} \quad (1)$$

où  $f_{proc}$  et  $f$  désignent l'image après le prétraitement et l'image d'origine respectivement,  $(i, j)$  les coordonnées pixels et  $\bar{p}$  le terme soustrait. Dans la première méthode  $\bar{p} = \lfloor mean(f) \rfloor$  et  $\bar{p} = 128$  dans la seconde.  $\lfloor \bullet \rfloor$  dénote l'opération d'arrondi au nombre entier le plus proche.

Hagag *et al.* [8] ont récemment utilisé la première solution (dénotée P1), tandis que la solution alternative (dénotée P2) a été proposée par Cui *et al.* [9]. Celle-ci consistant à soustraire un niveau de gris moyen sur 8 bits (128) évite une augmentation de la bande passante allouée pour la transmission de métadonnées. En effet, l'ajout d'un décalage de +128 après le processus de décodage ne nécessite l'envoi d'aucune information supplémentaire au récepteur.

### 3.2 Prétraitement dans le domaine fréquentiel

Récemment proposé dans [3], OPA-SoftCast (dénotée P3) fonctionne dans le domaine fréquentiel et s'insère après la 3D-DCT et la division en *chunks* comme indiqué en bleu dans la figure 1. OPA-SoftCast est un algorithme itératif qui vise à réduire l'énergie des *chunks*. Ceci est rendu possible en enlevant certaines fréquences spécifiques dénotées SFC qui seront envoyées à part en métadonnées. Les données d'entrées de l'algorithme sont représentées par une matrice  $X$  où chaque ligne représente un *chunk* de moyenne nulle. Tout d'abord, l'énergie de chaque *chunk* est calculée et la valeur la plus élevée parmi tous les *chunks* est sélectionnée. La composante fréquentielle la plus énergétique (SFC) dans ce bloc est supprimée et envoyée en métadonnée supplémentaire. Enfin, la moyenne du bloc sélectionné est ensuite ajustée pour conserver une valeur moyenne nulle. À la fin de la boucle, la nouvelle énergie du bloc sélectionné est calculée et mise à jour. La boucle est exécutée  $N_d$  fois où  $N_d$  représente le nombre de SFC sélectionnées dans le GoP. En raison de la position inconnue des *chunks* et des SFC les plus énergétiques, de la valeur des SFC retenues et des moyennes ajustées à chaque itération, OPA-SoftCast doit transmettre 4 informations pour chaque SFC.

Les auteurs ont supposé que chaque information est quantifiée avec 20 bits en moyenne. Avec une modulation BPSK et

un code FEC de redondance 1/2, le nombre total de bits à transmettre pour chaque GOP est de  $4 \cdot 20 \cdot 2 \cdot N_d = 160 \cdot N_d$  [3]. Ainsi, la largeur de bande nécessaire pour la transmission de ces métadonnées additionnelles est :  $\frac{4 \cdot 20 \cdot 2 \cdot N_d \cdot F_r}{N_I}$  où  $F_r$  et  $N_I$  désignent respectivement le nombre d'image par seconde et la taille du GoP. Afin de limiter l'augmentation de la bande passante liée aux métadonnées, les auteurs ont proposé un compromis entre amélioration de la qualité et quantité supplémentaire de métadonnées. Ils ont sélectionné un seuil équivalent à 2 SFC par image soit 16 SFC pour un GoP de 8 images. Toutefois, Fujihashi *et al.* ont montré qu'il est primordial de limiter la quantité de métadonnées, car celles-ci entraînent une dégradation de la qualité vidéo en raison de la perte de puissance et de débit pour la transmission des *chunks* eux-mêmes [11].

Plutôt que de recourir à l'algorithme OPA, nous proposons donc de sélectionner dans le domaine fréquentiel 3D uniquement la valeur de la composante DC. En effet, nous avons vérifié que la première SFC correspond toujours à la valeur de la composante DC. Ceci est normal puisqu'il est bien connu que la valeur de la composante DC, après une DCT, contient la plupart de l'énergie. La solution proposée (dénotée P4) représente le cas le plus défavorable de l'algorithme OPA, où une seule SFC est choisie dans le GoP, la valeur de la composante DC. Cependant, grâce à la position connue de cette composante DC, la métadonnée additionnelle est réduite de 160 à 40 bits (20 bits avant application du code FEC). Ceci entraîne une réduction de la bande passante additionnelle pour ces métadonnées de 9600Hz à 150Hz. Pour la méthode P1, la moyenne étant codée sur 8 bits, la consommation additionnelle de bande passante est de 480Hz. Nous montrons dans la suite que la méthode P4, bien que sous optimale à la version OPA, représente un bon compromis entre qualité de reconstruction, métadonnées additionnelles et temps d'exécution dans la section suivante.

## 4 Performances des méthodes étudiées

### 4.1 Configuration et paramètres choisis

Les séquences vidéo CIF provenant de la collection Xiph [12] sont utilisées avec une cadence de 30 images par seconde. Seule la composante de luminance est prise en compte. L'encodage est exécuté GoP par GoP avec une taille de GoP de 8 images et chaque image est divisée en 64 blocks de  $44 \times 36$  pixels, comme dans [7].

Pour une comparaison juste avec [3], la séquence *Mix* de 192 images est utilisée. La séquence est composée des 32 premières images des séquences suivantes : *Foreman*, *Akiyo*, *Coastguard*, *Flower*, *Paris* et *Bus*.

Les transmissions ont lieu via des canaux AWGN pour des rapports signal à bruit variant dans la plage [5~30dB]. La bande passante totale du canal est fixée à 1.6MHz afin de permettre la transmission de tous les coefficients analogiques [2, 11] ainsi que les métadonnées du schéma SoftCast originel. Pour une transmission fiable des métadonnées, une modulation BPSK et un code protecteur (FEC) de redondance 1/2 sont utilisés. Afin de transmettre les métadonnées additionnelles is-

sues du prétraitement P1, P3 ou P4 et assurer une comparaison équitable entre les méthodes (i.e., même bande passante), le nombre de *chunks* transmis est réduit en conséquence. Nous notons qu'un *chunk* dans la configuration actuelle représente 1584 coefficients soit 792 symboles (I et Q [2]).

Les performances sont mesurées à l'aide du PSNR qui est utilisé comme métrique purement objective ainsi que le SSIM qui fournit un indice de qualité plus corrélé avec le système visuel humain (SVH). L'index SSIM fournit des valeurs comprises entre 0 (qualité la plus basse) et 1 (qualité la meilleure). Le PSNR est donné par  $PSNR_{dB} = 10 \log_{10} \left( \frac{(2^L - 1)^2}{MSE} \right)$  où  $L$  désigne le nombre de bits représentant un pixel. La MSE vaut  $MSE = \frac{1}{PQ} \sum_{x=1}^P \sum_{y=1}^Q [(I_{ori}(x, y) - I_{rec}(x, y))^2]$  où  $I_{ori}$  désigne l'image d'origine,  $I_{rec}$  la version reconstruite et  $P, Q$  le nombre de ligne et de colonne d'une image.

### 4.2 Résultats de simulation

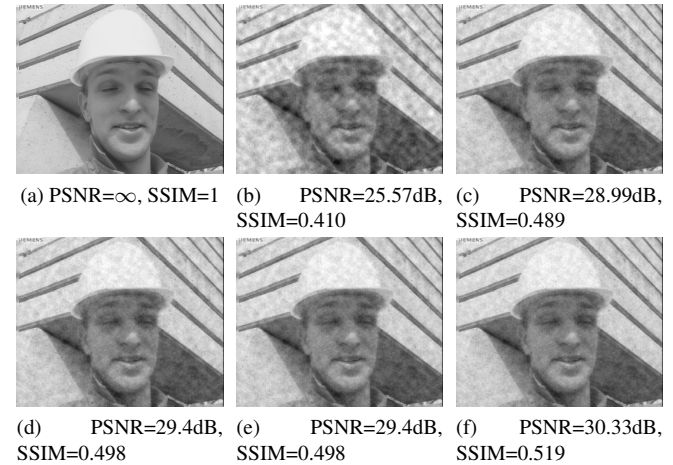


Figure 3: Comparaison visuelle de la qualité reçue dans un canal à 0dB pour la séquence *Mix* (première image). (a) Image originale, (b) SoftCast classique, (c) SoftCast (P2), (d) SoftCast (P1), (e) La méthode proposée (P4), (f) OPA-SoftCast (P3).

Un récapitulatif des gains moyens, avantages et inconvénients de chaque méthode par rapport au schéma d'origine SoftCast est donné dans le Tableau 1. En outre, le temps d'exécution et la bande passante additionnelle requise pour effectuer le prétraitement sont également indiqués. Le temps d'exécution est défini comme le temps total requis pour effectuer le prétraitement à l'émetteur et l'inverse au récepteur sur la séquence complète de 192 frames. Les différents temps ont été obtenus avec Matlab R2018b sur un ordinateur équipé d'un processeur Intel(R) Core(TM) i7-4510U CPU, 2GHz, 12G RAM. Ce tableau montre l'intérêt du prétraitement dans un contexte de transmission basé sur un CVL. En effet, un gain moyen de 2dB est observé pour les méthodes P1 et P4. Des gains similaires sont obtenus pour ces 2 méthodes, puisque le fait de soustraire la valeur moyenne de chaque image revient en effet à réduire la valeur des composantes DC de chaque plan DCT-2D [2]. Nous attirons les lecteurs sur le fait que la composante DC dans le domaine DCT-3D n'est autre qu'une moyenne des com-

Table 1: Comparaison des différentes méthodes de prétraitement étudiées

Méthode	Description	Avantages	Inconvénients	Gain Moyen	Temps d'exécution	BP requise
Valeur moyenne de l'image (P1)	Soustraction de la valeur moyenne pixel sur chaque image	Bon compromis entre performance et consommation additionnelle de bande passante [2]	Envoi de 8 bits supplémentaires en métadonnées pour chaque image	PSNR = 2.00dB SSIM = 0.041	379.9ms	480Hz
Niveau de gris moyen 8-bit = 128 (P2)	Soustraction d'un niveau de gris moyen = 128 sur chaque image	Pas de consommation additionnelle de bande passante	Les performances dépendent des caractéristiques de la vidéo [2]	PSNR = 1.58dB SSIM = 0.034	308.7ms	0Hz
OPA-SoftCast [3] (P3)	Sélection de certaines fréquences (SFC) envoyées en métadonnée et mises à zéro ensuite	Meilleure Amélioration de la qualité	Grande consommation de métadonnées additionnelles (80 bits par SFC avant FEC) et complexité plus importante	<b>PSNR = 2.47dB</b> <b>SSIM = 0.069</b>	293.6ms	9600Hz
Méthode proposée (P4)	Envoi de la DC en métadonnée (après DCT-3D) mise à zéro ensuite	Meilleur compromis entre performance et consommation de bande passante (20 bits par GoP)	Les performances dépendent de l'homogénéité de la valeur moyenne pixel dans le GoP	PSNR = 2.00dB SSIM = 0.041	<b>0.955ms</b>	150Hz

posantes DC issues du domaine DCT-2D spatial [2]. Ainsi, si les valeurs de ces composantes au sein d'un même GoP sont très différentes (un changement de plan par exemple), la solution proposée devient sous-optimale puisque l'énergie n'est pas totalement compactée sur la composante DC-3D. Toutefois, nous avons montré dans [13] qu'une détection des changements de plan liée à un encodage adaptatif permettait de s'affranchir de ce problème. La solution proposée peut donc être utilisée permettant ainsi une réduction des métadonnées additionnelles de 8 bits par image (P1), soit un total de 64 bits pour un GoP de 8 images, à 20 bits par GoP (P4) avant application du code FEC. De plus, comme indiqué dans le tableau, le temps d'exécution total de la méthode P4 sur la séquence considérée est insignifiant par rapport à P1. En effet, notre méthode n'est pas itérative et ne nécessite qu'une seule opération par GoP contrairement aux autres qui nécessitent de nombreuses opérations (une par pixel pour les méthodes P1, P2) ou sont itératives (P3).

Finalement, une comparaison visuelle est donnée dans la Figure 3. Le canal le plus défavorable est choisi (CSNR=0dB) afin d'accentuer le bruit de transmission. L'image où la différence de gain entre la méthode OPA-SoftCast (P3) et la méthode proposée (P4) est la plus importante est choisie (première image de *Foreman*). Nous pouvons observer une amélioration maximale du PSNR et du SSIM de 4.75dB et 0.1, respectivement entre la version SoftCast d'origine et la méthode P3. Au prix d'un écart maximal et moyen de PSNR entre la solution proposée et la méthode P3 de 0.93dB et 0.47dB respectivement, la bande passante additionnelle de métadonnées passe de 9600 à 150 Hz et le temps d'exécution de 380 à 1 milliseconde(s). Le choix du prétraitement se fera donc en fonction de la bande disponible ainsi que des contraintes de temps liées à l'application visée. Comme dans [2], nous avons également vérifié l'évolution du PSNR en fonction du CSNR qui reste quasi-linéaire [7, 11] qu'importe le prétraitement choisi.

## 5 Conclusion

Dans cet article, nous étudions des méthodes de prétraitement pouvant être utilisées dans un contexte SoftCast. En outre, nous proposons une version agissant dans le domaine fréquentiel permettant d'offrir le meilleur compromis entre bande passante additionnelle nécessaire aux métadonnées, temps d'exécution et qualité reçue. Le choix de la méthode de prétraitement

est effectué en fonction de la bande passante disponible. En effet, les différentes méthodes P1, P3 et P4 nécessitent une allocation plus importante de la bande passante aux métadonnées. En revanche, la méthode P2 bien que sous optimale, ne requière aucune métadonnée additionnelle, ce qui en fait une solution possible dans les environnements à bande passante très limitée. Si l'application permet l'envoi d'un nombre limité de métadonnées, la méthode proposée (P4) donne le meilleur compromis bande passante/amélioration de la qualité/temps d'exécution avec un gain moyen de 2dB sur les séquences testées.

## Références

- [1] S. Kokalj-Filipovi, E. Soljanin, *Suppressing the cliff effect in video reproduction quality*, Bell Labs Technical Journal, vol. 16, no. 4, pp. 171-185, Mar. 2012.
- [2] A. Trioux, F.-X. Coudoux, P. Corlay, M. Gharbi, *A Comparative Pre-processing Study for SoftCast Video Transmission*, in Int. Symp. on Signal, Image, Video and Commun. (ISIVC), Nov. 2018.
- [3] C. He, H. Qin, Z. He, and K. Niu, *Adaptive GoP dividing video coding for wireless broadcast based on power allocation optimization*, in Int. Conf. on Wireless Commun. & Signal Process. (WCSP), pp. 175, 2016.
- [4] Cisco, *Cisco visual networking index: Forecast and Methodology 2016-2021*, 2017.
- [5] I.E.G. Richardson, *The H.264 advanced video compression standard*, 2nd ed. Chichester: Wiley, 2010.
- [6] G.J. Sullivan, J.R. Ohm, W.J. Han, and T. Wiegand, *Overview of the High Efficiency Video Coding (HEVC) Standard*, IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
- [7] S. Jakubczak and D. Katabi, *SoftCast: Clean-slate scalable wireless video*, MIT Technical report, Feb. 2011.
- [8] A. Hagag, X. Fan, and F.E.A. El-Samie, *Satellite Images Broadcast Based on Wireless SoftCast Scheme*. Int. Journal Computer Science, vol. 44, no. 1, Feb. 2017.
- [9] H. Cui, D. Liu, Y. Han, and J. Wu, *Robust Uncoded Video Transmission under Practical Channel Estimation*, in Global Commun. Conf. (GLOBECOM), pp. 1-6, Dec. 2016.
- [10] R. Xiong, F. Wu, J. Xu, X. Fan, and al., *Analysis of decorrelation transform gain for uncoded wireless image and video communication*, IEEE Trans. Image Process., vol. 25, no. 4, pp. 1820-1833, Apr. 2016.
- [11] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P.V. Orlik, *High-Quality Soft Video Delivery With GMRF-Based Overhead Reduction*, IEEE Trans. Multimedia, vol. 20, no. 2, pp. 473-483, Feb. 2018.
- [12] *Xiph.org* Available: <https://media.xiph.org/video/derf/>
- [13] A. Trioux, F.-X. Coudoux, P. Corlay, M. Gharbi, *Etude de l'influence du contenu vidéo dans une transmission Pseudo-Analogique de type SoftCast*, in COMpression et Représentation des Signaux Audiovisuels (CORESA), Nov 2018.