

Comment ça marche ?

Les transmissions radio-numériques

15 – Le codage de la parole (1)

Par le radio-club F6KRK

Après avoir vu le formatage des données et les méthodes de corrections d'erreurs, nous allons maintenant aborder le codage de l'information pour réduire le flot de données à transmettre, en commençant par le codage de la parole (1^{ère} partie). Bien sûr, il ne s'agit pas d'exiger du lecteur une compétence en traitement du signal. Cet article n'a d'autre ambition que de donner une idée générale des processus employés.

Exposé du problème

La particularité des transmissions analogiques est d'être à "flot continu" et en "temps réel" ⁽¹⁾. Nous avons vu par ailleurs qu'elle s'adresse à nos sens, la vue et l'ouïe avec notre cerveau pour effectuer un traitement des erreurs de transmission. En numérique on pourrait faire la même chose mais cela demanderait une largeur de bande bien supérieure à l'analogique. Par exemple en numérisant en absolu chaque échantillon, une transmission audio monophonique avec une bande de 15 kHz et une dynamique de 70 dB demanderait un débit numérique de 5 Mbits/s (Horloge à 42 kHz et mots de 12 bits). Pour de la téléphonie 300-3000 Hz avec une dynamique de 40 dB, on aurait encore un débit de 64 kbits/s. C'est inacceptable et on a alors recours à des systèmes de codage qui sont tous basés sur la suppression des informations redondantes et inutiles. Ainsi, un système de codage de la parole comprend 2 parties : le codeur et le décodeur. Le codeur analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits pour archivage ou transmission. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique.

On distingue trois catégories de codages selon les plages de débits :

- Les hauts débits, supérieurs à 16 kbps, correspondant à des algorithmes de codage de la forme d'onde non spécifiques à la parole.
- Les débits moyens, de 4 kbps à 16 kbps, utilisant des méthodes de codage de la forme d'onde et prenant en compte certaines propriétés de la parole ou de la perception auditive. Le principal représentant de cette classe est le codage CELP.
- Les bas et très bas débits, de quelques dizaines de bits par seconde à 4 kbps correspondant aux vocodeurs (VOICE CODER) spécifiques au codage de la parole.

Pour que ces systèmes soient efficaces, on applique au préalable un traitement du signal pour réduire la dynamique, d'une part avec un ALC (Automatic Level Control), et d'autre part avec un compresseur, en général logarithmique.

Les "CODEC" (COdage et DECodage de l'audio en général)

Commençons avec les systèmes fonctionnant en haut débit continu et pouvant admettre quelques erreurs sans déclarer forfait. Ils sont basés sur un codage d'impulsions (PCM pour

"Pulse Coded Modulation"). Avec un PCM simple, pour une dynamique de 8 bits, le débit binaire est égal à 8 fois le rythme d'échantillonnage, ce qui ne change rien. Mais on peut améliorer l'efficacité avec des variantes.

Noter que tous ces systèmes entraînent des erreurs sur le signal, d'autant plus qu'ils sont efficaces (ils ne sont pas faits pour la Hi-Fi). Mais les erreurs restant limitées on peut les utiliser pour tout les types de signaux audio : parole, musique, DTMF, etc.

1 - DM (Delta Modulation).

Principe : Au lieu de transmettre les valeurs brutes des échantillons, on ne transmet que le delta entre deux échantillons. Voir la figure 1.

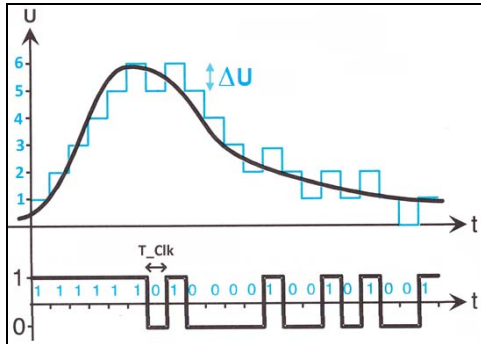


Figure 1 : Principe de la "Delta Modulation"

Par rapport au PCM simple, le débit binaire est maintenant égal au rythme d'échantillonnage, mais celui-ci doit quand même être nettement plus élevé que le critère de Nyquist. Par ailleurs la DM a des limites liées au choix de la valeur du delta U. Si celui-ci est trop faible, on aura une diminution de l'amplitude aux fréquences élevées si le rythme d'échantillonnage est trop bas. Noter qu'avec la parole, les fréquences élevées étant d'un niveau plus faible, cela permet judicieusement de réduire ce rythme d'échantillonnage. D'un autre côté, si le signal est plus faible qu'un delta U, nous aurons une suite de "1" et de "0" de valeur moyenne nulle. Il y a donc un effet de seuil.

2 - DPCM (Différential PCM).

Le DPCM est une extension de la Delta Modulation qui utilise également la redondance dans un signal analogique, spécialement dans la parole. En codage différentiel, la différence entre un échantillon du signal et une valeur prédite est quantifiée par une valeur parmi p . La complexité du système est liée à la complexité du prédicteur. Voir sur la figure 2 un synoptique général du système.

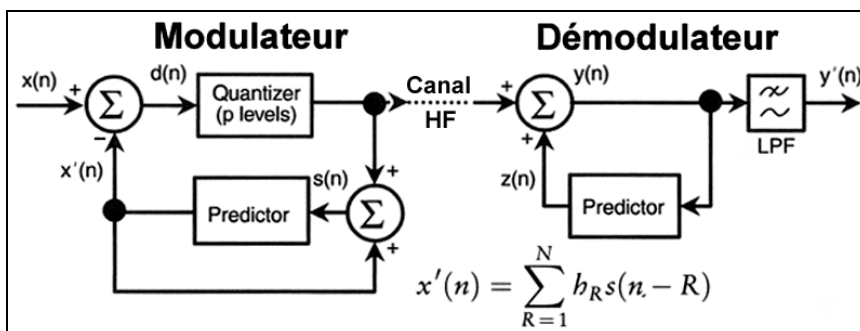


Figure 2 : Modulateur et démodulateur DPCM (h_R = coefficients du prédicteur)

Le principal avantage du DPCM sur le PCM est l'amélioration du rapport S/B. Mais elle plafonne à $N=2$ pour la parole.

3 - ADPCM (Adaptive DPCM)

La qualité d'un algorithme de codage dépend de notre connaissance statistique du signal. Si nous avons un signal bien défini nous pouvons utiliser un prédicteur à temps constant. Dans plusieurs cas, bien que les statistiques à long terme soient bien connues, le signal s'écarte significativement d'elles pour de courtes périodes. Dans ce cas un système de codage adaptatif amènera un avantage.

Le terme "adaptatif" dans l'ADPCM s'applique à deux systèmes différents : adaptation de la quantification et adaptation du prédicteur. Les valeurs estimées peuvent être obtenues de deux manières : soit par pré-estimation (DPCM-AQF), soit par post-estimation (DPCM-AQB). Voir les synoptiques des deux principes sur la figure 3.

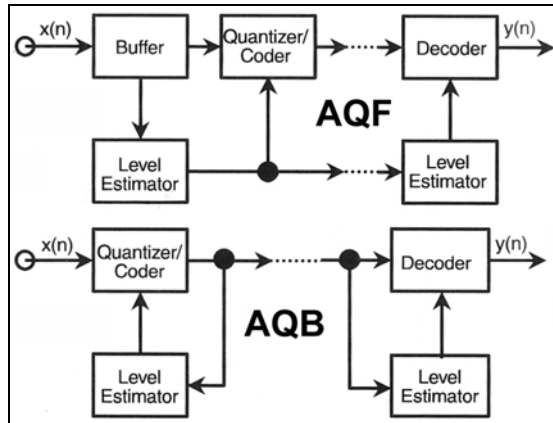


Figure 3 : Synoptiques des procédés ADPCM-AQF et ADPCM-AQB

Le procédé AQF demande plus de données et apporte un retard non négligeable, aussi le procédé AQB est le plus souvent employé.

Un ADPCM-AQB a été normalisé par le CCITT au standard G.721.

Un dérivé dénommé ADM, du type AQB est aussi employée grâce à sa facilité d'implémentation.

4 - CVSD (Continuously Variable Slope Delta modulation)

Même si l'ADM est facile à implémenter, il entraîne une dégradation significative de la parole quand il y a des erreurs dans la transmission. Ces erreurs peuvent se propager longtemps dans le flot de parole.

En CVSD, la valeur du delta dépend des deux précédentes valeurs de sortie de l'encodeur, $y(n-1)$ et $y(n-2)$. La valeur actuelle $\Delta(n)$ est donnée par les formules suivantes :

$$\Delta(n) = \beta\Delta(n-1) + D_2$$

$$\text{si } y(n) = y(n-1) + y(n-2)$$

$$\text{ou } \Delta(n) = \beta\Delta(n-1) + D_1$$

dans tous les autres cas

$$\text{avec } 0 < \beta < 1 \text{ et } D_2 \gg D_1 > 0$$

Les valeurs de D_1 , D_2 et β qui constituent les coefficients introduits sont en relation avec les deltas minimum et maximum requis par les formules :

$$\Delta_{\max} = \frac{D_2}{1 - \beta} \quad \Delta_{\min} = \frac{D_1}{1 - \beta}$$

Les valeurs de ces trois coefficients sont soigneusement choisies pour s'assurer que le codeur CVSD puisse suivre le signal entrant et dérivent d'une analyse statistique à long terme du niveau. Les premières formules montrent que les pas augmentent quand nous avons des suites de "0" et de "1". En regardant la figure 4, nous pouvons voir que c'est exactement ce qui est

nécessaire en cas de dépassement. Dans tous les autres cas, l'amplitude des pas décroît. Si nous choisissons β trop proche de 1, le changement des $\Delta(n)$ sera lent. Inversement, si nous choisissons β proche de zéro, l'adaptation sera rapide. En général, β est choisi en premier puis D_1 et D_2 sont calculés avec les équations précédentes pour Δ_{\max} et Δ_{\min} .

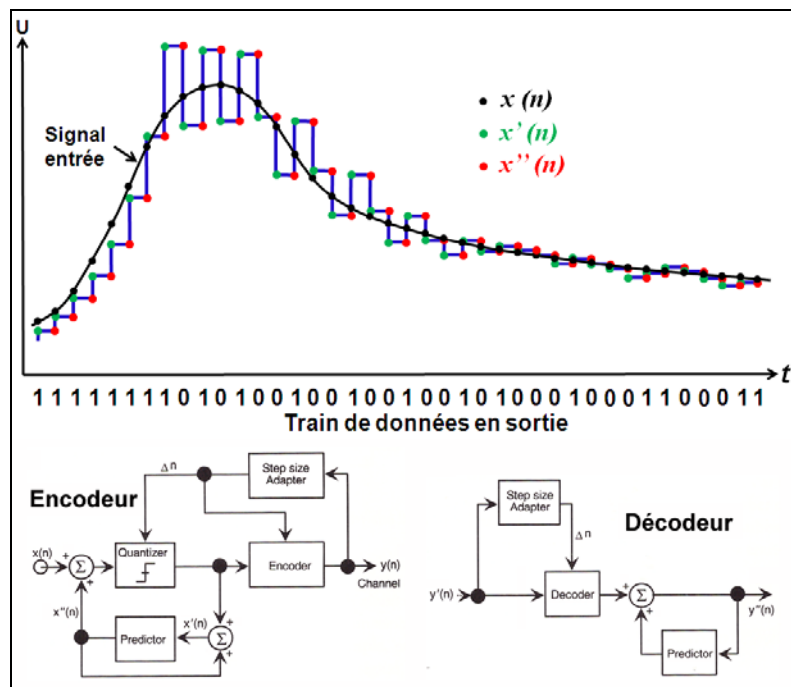


Figure 4 : Encodage et décodeur CVSD avec exemple de codage

Bien que le CVSD fournisse une phonie de qualité médiocre par rapport à l'ADM, il est bien moins sensible aux erreurs de transmission. Le CVSD fut très employé jusqu'à ce que l'ADPCM soit standardisé par le CITT.

Nous nous contenterons de ces quatre principaux codages, mais il en existe bien d'autres, comme l'APC (Adaptive Predictive Coding), le SBC (SubBandCoding), etc.

Dans le prochain "Comment ça marche ?" nous continuerons le codage de la parole avec les moyens débits (CELP) et les bas débits (VoCodeurs).

La Rubrique "Comment ça marche" est une activité collective du radio-club F6KRK (<http://www.f6krk.org>). Pour une correspondance technique concernant cette rubrique : "f5nb@orange.fr".

Bibliographie

Pour cette première partie, nous nous sommes principalement servi de cette référence : "A simple approach to DIGITAL SIGNAL PROCESSING" par Craig Marven & Gillian Ewers. Edité par Texas-Instruments.

Note :

- 1) *On considère que l'on est en temps réel quand lors d'une interruption d'une information, le cerveau la perçoit comme si elle était continue. En Audio on peut accepter une interruption de quelques millisecondes et de quelques dizaines de millisecondes pour la vision. Pour le retard de transmission, celui-ci doit être limité à quelques centaines de millisecondes pour un échange audio et peut aller à quelques secondes pour une diffusion radiophonique ou télévisuelle.*