

Codage et compression de l'audio

INF4710 Introduction aux technologies
multimédia

Plan

- ▶ Signal audio
- ▶ Codage PCM
- ▶ Loi μ
- ▶ Codage Golomb–Rice
- ▶ LPC
- ▶ MDCT

Codage et compression de l'audio

▶ Signal audio:

- Par rapport à une image ou à une séquence vidéo, un signal audio est originalement sous forme analogique (c'est un signal continu). En effet, un microphone émet un signal continu, alors que pour les images, de nos jours, les capteurs fonctionnent en mode numérique, et transmet un signal numérique.
- Pour l'image, le nombre de photons reçus à chaque photosite est quantifié par le capteur (numérisé). L'image est une matrice de valeurs numériques.
- Pour l'audio, une onde sonore est convertie en signal électrique.

Codage et compression de l'audio

▶ Signal audio:

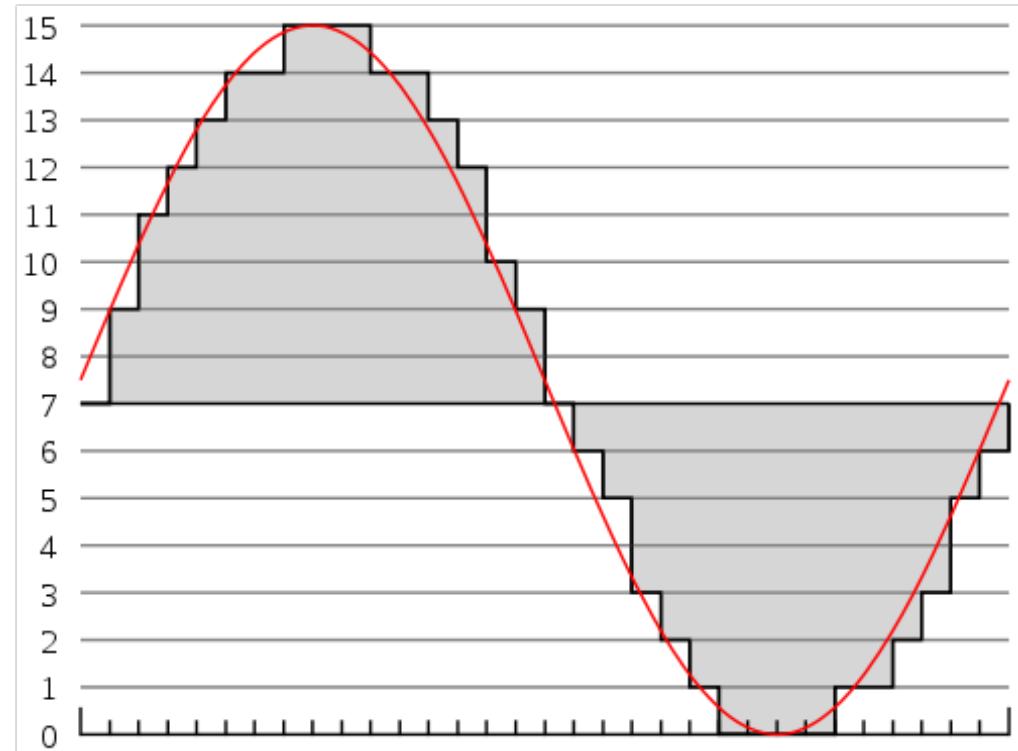
- Donc, pour stocker ou transmettre un signal audio via un ordinateur, il faut d'abord le convertir en signal numérique.
- On utilise un convertisseur analogique-numérique qui générera un signal PCM (pulse-code modulation).
- Le signal PCM est le signal de base pour représenter le son. Il est d'ailleurs utilisé sur les blu-ray, CD, et DVD, et pour le format wav.

Codage PCM

- ▶ La fidélité du codage PCM est définie par la fréquence d'échantillonnage et le nombre de bits utilisé.
- ▶ Le signal PCM est généré de la façon suivante:
 1. Le signal est échantillonné selon la fréquence d'échantillonnage choisie;
 2. Pour chaque échantillon, on approxime sa valeur par la valeur numérique la plus proche selon le nombre de bits choisi.
- ▶ La fréquence d'échantillonnage doit être choisie selon le théorème de Nyquist–Shannon qui stipule que la fréquence d'échantillonnage doit être plus de deux fois la fréquence la plus haute du signal.

Codage PCM

- ▶ Note: il y aura des erreurs de quantification. Ici, le signal est codé sur 4 bits.



- ▶ Exemple MATLAB:
 - *CodagePCM.m*

Codage PCM

▶ Codage PCM

- La quantification peut être faite de différentes façons. Si elle est linéaire, on a un signal LPCM (Linear PCM). -> CD, wav, DVD, blu-ray, HDMI
- Certaines formes de PCM combinent la compression et le codage: DPCM, ADPCM, modulation Delta

Codage et compression de l'audio

- ▶ Codage DPCM (Differential PCM)
 - C'est une méthode de codage prédictif.
 - On encode seulement la différence entre deux échantillons PCM.
 - En audio, cela permet une compression de 25% environ.

Codage DPCM

- ▶ Codage DPCM (Differential PCM)
 - ▶ Soit le signal suivant à coder:
 - Exemple: 7, 9, 11, 12, 13, 14, 15
 - Trouver le code DPCM, si la prédiction est la valeur précédente.

Codage et compression de l'audio

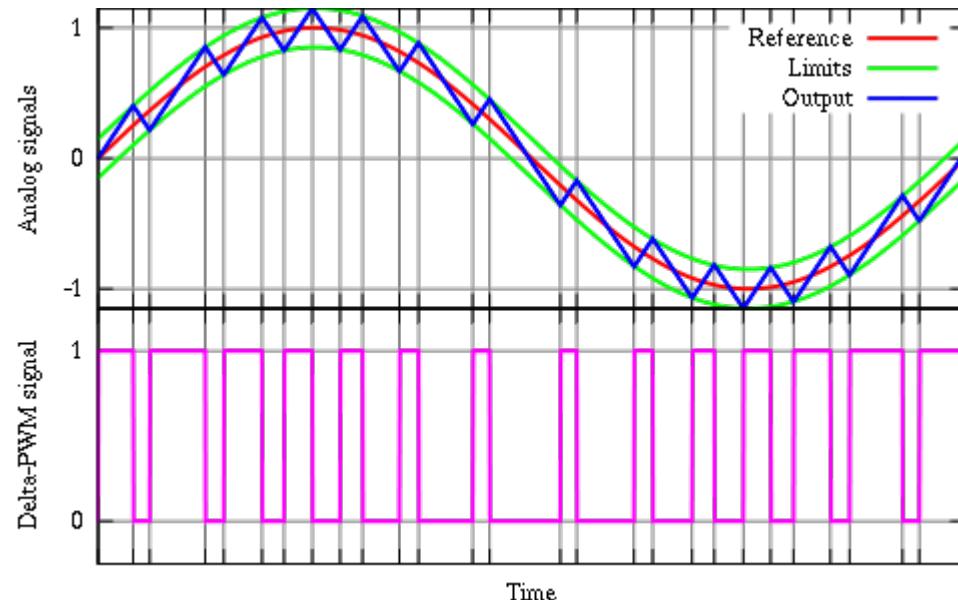
- ▶ Codage ADPCM (Adaptive DPCM)
 - Dans ce cas, le pas de quantification est modifié dynamiquement selon la grandeur des erreurs.
- ▶ Exemple MATLAB:
 - *CodageADPCM.m*
- ▶ Codage modulation Delta
 - Dans ce cas, la différence est encodée à l'aide d'un seul bit: 0 baisse d'amplitude, 1 hausse d'amplitude.
 - La qualité n'est pas très bonne. On peut sur-échantillonner pour avoir une meilleure fidélité (les variations instantanées étant alors plus petites).

Codage modulation Delta

- ▶ Codage modulation Delta
 - ▶ Soit le signal suivant à coder:
 - Exemple: 7, 9, 11, 12, 13, 14, 15, 11, 11, 11, 11
 - Trouver le code modulation delta avec des augmentations et baisses de 1.
 - Trouver le code modulation delta avec des augmentations et baisses de 2.

Codage modulation Delta

- ▶ Un autre moyen d'obtenir une meilleure fidélité avec la modulation Delta est de considérer l'enveloppe autour du signal à coder pour déterminer quand augmenter ou baisser l'amplitude.



Codage et compression de l'audio

▶ μ -law (loi μ)

- La loi mu est une méthode de quantification logarithmique qui réduit la plage dynamique et permet:
 - d'augmenter le rapport signal sur bruit;
 - De compresser le signal pour un rapport signal sur bruit donné. C'est-à-dire qu'au lieu d'augmenter le rapport signal sur bruit, on peut garder le rapport original et réduire la bande passante.
- Plage dynamique: rapport entre la plus grande et la plus petite valeur.
- Rapport signal sur bruit: grandeur d'un signal par rapport au bruit.

loi μ

- ▶ La loi mu est calculée de la façon suivante pour l'encodage:

$$F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)} \quad -1 \leq x \leq 1$$

- ▶ Et pour le décodage:

$$F^{-1}(y) = \text{sgn}(y)(1/\mu)((1 + \mu)^{|y|} - 1) \quad -1 \leq y \leq 1$$

- ▶ Exemple MATLAB:
 - *CodageLoiMU.m*

loi μ

- ▶ On peut compresser le signal en allouant plus de bits pour les signaux faibles (moins fréquents) après l'application de la loi μ .
- ▶ L'idée est d'ajuster l'erreur en fonction de la force du signal (e.g. chercher à obtenir toujours 5% d'erreur).
- ▶ L'erreur peut être plus grande pour un signal fort, car le signal fort cache mieux les erreurs qu'un signal faible.
- ▶ Exemple d'application: Format Au (Sun microsystems), Téléphonie.

Codage et compression de l'audio

▶ Codage Golomb–Rice

- Méthode de compression sans perte adaptée au codage de signaux où les petites valeurs sont fréquentes. C'est le cas d'un signal codé avec un codage différentiel tel que DPCM.
- Le principe est de séparer un nombre x en son quotient (q) et son reste (r). x sera codé par (q,r) . q est codé par un codage unaire et r est codé par un codage binaire tronqué.

Codage Golomb–Rice

- ▶ Codage unaire
 - Le codage unaire est un codage entropique qui représente un nombre x , par x « un » suivi d'un « zéro ».
 - Donc:
 - $0 \rightarrow 0$
 - $1 \rightarrow 10$
 - $2 \rightarrow 110$
 - $3 \rightarrow 1110$
 - etc.
 - Optimal pour les distributions de probabilité géométriques de type $P(n)=2^{-n}$

Codage Golomb–Rice

▶ Codage binaire tronqué

- Méthode de codage entropique utilisée pour les distributions de probabilité uniformes avec un alphabet prédéterminé.
- Généralise le codage binaire quand le nombre de symboles (n) n'est pas une puissance de 2. n est défini comme étant $n=2^k+b$
- Ce codage assigne aux premiers 2^k-b symboles des codes de longueur k , et aux autres $2b$ symboles des codes de longueur $k+1$;
- C'est un code préfixé, car les $2b$ derniers codes utilisent des codes inutilisés de longueur k avec l'ajout de 0 et 1.

Codage binaire tronqué

- ▶ Si $n=7$ (7 symboles à coder)
 - Donc $n=2^2+3$. $k=2$, et $b=3$
 - Les premiers 2^2-3 symboles sont de taille 2. Les autres (2^*3) de taille 3.
 - Donc:
 - 0-> 00
 - 1-> 010 (01 inutilisé)
 - 2-> 011
 - 3-> 100 (10 inutilisé)
 - 4-> 101
 - 5-> 110 (11 inutilisé)
 - 6-> 111

Codage binaire tronqué

- ▶ Pour le décodage:
 - On lit les k premiers bits.
 - Si la valeur encodée est plus petite que u ($u = 2^{k+1} - n$), le symbole est complet. u représente le nombre de codes inutilisés sur $k+1$ bits.
 - Sinon, on lit un autre bit.
- ▶ Pour l'exemple précédent:
 - $u = 2^3 - 7 = 1$

Codage Golomb–Rice

- ▶ Tel que mentionné précédemment, le codage Golomb–Rice code le quotient q d'un nombre x avec le codage unaire, et le reste r avec le codage binaire tronqué.
- ▶ Construction du code:
 1. On choisit une valeur M (habituellement, la médiane de la distribution)
 2. Pour chaque symbole x , on trouve:
 - Quotient: $q = \text{int}[x/M]$
 - Reste: $r = x \bmod M$

Codage Golomb–Rice

- ▶ Construction du code:
 - 3. Génération du code <code de q><code de r>
 - a) Quotient
 - Écrit q bit 1 + un bit 0
 - b) Reste
 - Codage binaire tronqué selon M

Codage Golomb–Rice

- ▶ Codage Golomb–Rice
 - ▶ Soit le signal suivant à coder (0 à 22):
 - 5, 3, 4, 0, 3, 9, 5, 1, 1, 1, 1
 - Trouver le code Golomb–Rice avec $M=6$.

Codage Golomb–Rice

- ▶ Exemple MATLAB:
 - *CodageGolombRice.m*

Codage et compression de l'audio

- ▶ Codage prédictif linéaire (Linear Predictive Coding, LPC)
 - Le codage prédictif que nous avons vu pour l'image est appliqué aussi pour la voix.
 - RAPPEL: L'erreur e_n pour le $n^{\text{ième}}$ symbole est:

$$e_n = x_n - \hat{x}_n$$

- où x_n est un symbole à coder et \hat{x}_n est le symbole prédit obtenu par

$$\hat{x}_n = \sum_{i=1}^m \alpha_i x_{n-i}$$

Codage prédictif linéaire

- ▶ Pour obtenir la meilleure compression possible, il s'agit d'estimer le mieux possible les α_i .
- ▶ Avec une dizaine de coefficients bien choisis. Il est possible de reproduire la voix humaine.
- ▶ Un signal de voix peut être divisé en deux parties:
 - La fonction de transfère (cordes vocales) définie par les coefficients -> commun à tous
 - Fréquences et amplitudes-> propre à chaque personne.

Codage prédictif linéaire

- ▶ Ainsi pour coder un signal de voix, la prédiction est faite à partir d'un modèle des cordes vocales, et l'erreur à encoder sera donc essentiellement les caractéristiques propres à chaque personne et selon les mots prononcés.
- ▶ Utilisation: dans le standard GSM en téléphonie cellulaire, MPEG-4 ALS, FLAC

Codage et compression de l'audio

- ▶ Format FLAC (Free Lossless Audio Codec)
 - Lancer en 2001
 - Compression 50 à 60%
 - Utilise des algorithmes de compression spécifiquement conçus pour l'audio (exemple: par rapport à DEFLATE, compression de 10–20% en audio).
 - Utilise les codages:
 - DPCM ou LPC
 - Golomb–Rice sur le signal DPCM ou LPC
 - Codage par plage (RLE) pour les blocs de valeur identique

Codage et compression de l'audio

- ▶ La loi μ exploite les caractéristiques de l'oreille au niveau de la perception du bruit.
- ▶ On peut aller plus loin dans cette direction par l'utilisation des modèles psychoacoustiques.
- ▶ Exemple: Codage de sous–bandes (Sub–Band Coding, SBC)

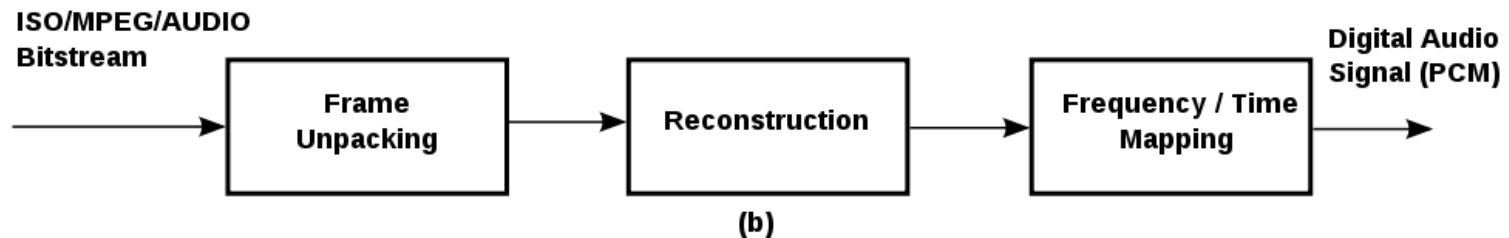
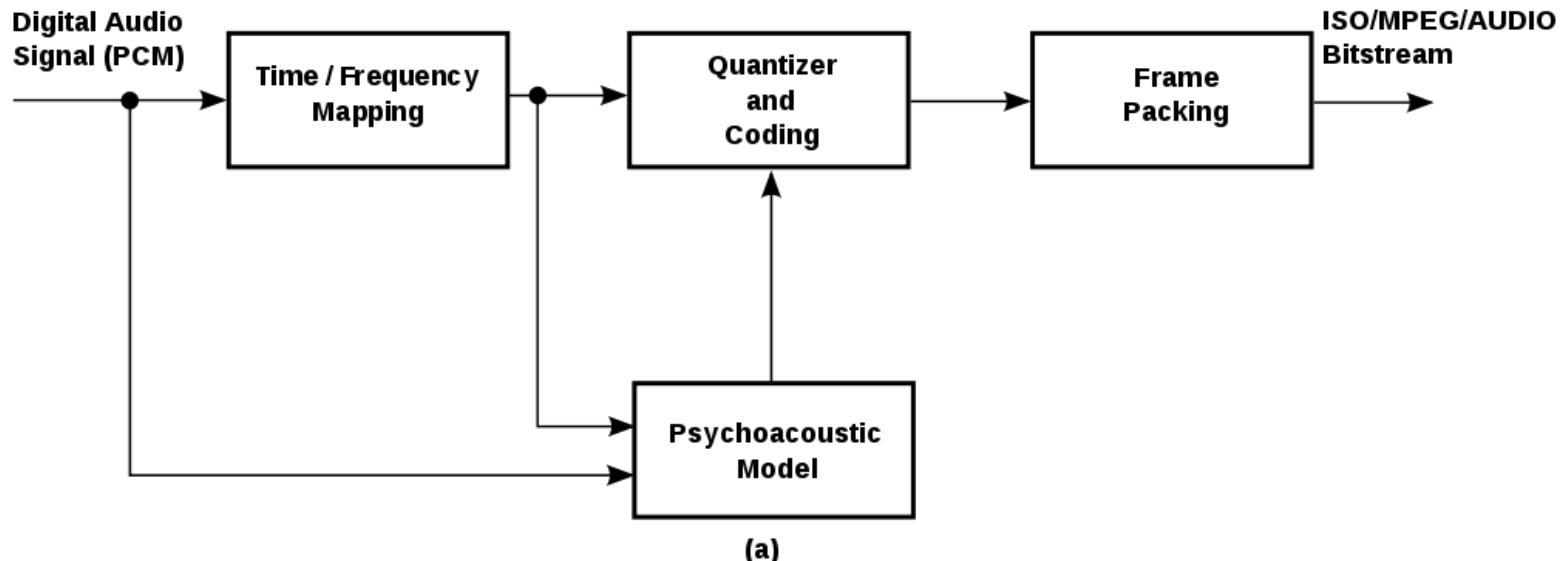
Codage de sous-bandes

- ▶ Pour une bonne qualité avec PCM, il faut suffisamment de bits pour que les erreurs ne soient pas perçues. Donc, pour certaines parties du signal, la précision est plus grande que nécessaire.
- ▶ L'oreille humaine est sensible à un grand intervalle de fréquences, mais si un signal fort est présent à une fréquence, l'oreille n'entendra pas les signaux plus faibles à des fréquences proches. C'est un effet de masquage fréquentiel.

Codage de sous-bandes

- ▶ Ainsi, lors de l'encodage d'un signal audio, certaines fréquences peuvent être éliminées, car elles sont pas perceptibles par l'oreille.
- ▶ C'est ce que permet de faire le codage de sous-bandes.
 - Le signal original est divisé en sous-bandes.
 - Pour chaque sous-bande, à l'aide des informations psychoacoustiques, on évalue si un signal est masquant, et un seuil de masquage est défini en conséquence.
 - Le codage est ensuite fait en tenant compte du seuil de masquage. Les erreurs doivent être plus petites que le seuil de masquage (elles ne seront pas perceptibles).

Codage de sous-bandes



Codage de sous-bandes

- ▶ Le codage de sous-bandes est utilisé dans le format mp3 (MPEG-1 Audio Layer 3).

Codage et compression de l'audio

- ▶ La compression par transformée est aussi appliquée en audio.
- ▶ Exemple: La transformée en cosinus discrète modifiée (Modified Discrete Cosine Transform, MDCT).
- ▶ La MDCT est utilisée dans MP3, AC-3, Vorbis, WMA, AAC, etc.

MDCT

- ▶ La MDCT est comme la transformée en cosinus discrète que nous avons vue précédemment, sauf que:
 - Les blocs transformés par la MDCT sont en recouvrement pour éviter les artéfacts liés aux effets de blocs.
 - Donc, le bloc transformé par la MDCT est calculé à partir du bloc original + des parties du bloc suivant et du bloc précédent.
- ▶ Proposée en 1987.

MDCT

▶ Calcul:

$$X_k = \sum_{n=0}^{2N-1} x_n \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right)$$

x_n est le signal d'entrée de taille $2N$, et X_k est le signal de sortie de taille N . La transformée inverse est

$$y_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cos\left(\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right)$$