



TELECOM
ParisTech



Institut
Mines-Telecom

Codage de la parole et de la musique

Marco Cagnazzo,
cagnazzo@telecom-paristech.fr

TELECOM-ParisTech – TVNum

A horizontal bar at the bottom of the slide, divided into three segments of equal width: red, black, and brown.

Plan

Introduction

- Signal de parole
- Signal de musique

Perception de l'audio

- L'oreille
- Seuil d'audition
- Le masquage

Compression de la parole

- Codeurs simples
- Codeurs CELP
- Codeur 3GPP AMR-WB

Compression de la musique

- MP3
- AAC

Plan

Introduction

Signal de parole

Signal de musique

Perception de l'audio

Compression de la parole

Compression de la musique



Signal de parole

- ▶ Signal non-stationnaire, mais localement stationnaire (20 ms)
- ▶ Types de son : voisés (voyelles, consonnes sonores, vibrantes, nasales) non-voisés (consonnes sourdes), autre (transition entre phonèmes)
- ▶ Modèles de prédictions simples et efficaces :
 - ▶ Filtrage linéaire (autorégressif) d'une suite d'impulsions pour les sons voisés
 - ▶ Même filtre appliqué sur du bruit blanc pour les sons non-voisés
 - ▶ Ce système reproduit les caractéristiques du trait vocal

Signal de parole

Numérisation

- ▶ PCM (Pulse Code Modulation) / MIC (Modulation par Impulsions Codées)
 - ▶ Bande 200 Hz ÷ 3400 Hz
 - ▶ Suffisante pour l'intelligibilité de la parole
 - ▶ Échantillonnage à 8000 Hz : $F_s > 2F_{\max}$
 - ▶ Représentation de chaque échantillon sur 8 bits
 - ▶ Cela donne 64 kbps
- ▶ Parole en bande élargie
 - ▶ Introduction 50-200 Hz : voix plus naturelle, amélioration de l'effet de présence
 - ▶ Extension 3.4-7 kHz : plus grande intelligibilité



Signal de parole

Normes de compression : Réseau téléphonique commuté

- G.711** (1972) PCM (pas de compression), 8 échantillons par ms codés sur 8 bits : 64 kbps
- G.721** (1984) Codage ADPCM (prédiction par filtre linéaire) : 32 kbps
- G.728** (1991) Codeur de type CELP (Code Excited Linear Predictive), avec faible délai de décodage : 16 kbps
- G.729** (1995) Codeur de type CELP, sans contrainte sur le retard : 8kbps
- G.723.1** (1995) Codeur à 6.3 kbps, pour visiophone

Signal de parole

Normes de compression : Communication mobile

GSM 06.10 (1988) RPE-LTP : Regular Pulse Excitation Long Term Predictor. 13 (22.8) kbps

GSM 06.20 (1994) "Half-Rate". Débit de 5.6 (11.4) kbps

GSM 06.60 (1996) "ACELP". Débit de 12.2 (22.8) kbit/s

GSM 06.90 (1999) Codage source/canal à débit variable $4.75 \div 12.2$ ($11.4 \div 22.8$) kbps (ACELP-AMR : Adaptive Multi Rate)

G.722 Codeur de parole en bande élargie, débit 6.6, 8.85, 12.65 kbps (AMR-WB) ; débits ultérieurs 15.85 et 23.85 kbps

- ▶ Orange et SFR utilisent ce codeur sur le réseau 3G+ et pour le VoIP (dépuis 2010)
- ▶ Utilisé aussi en UK, Espagne, Belgique, Canada, Taïwan, US, Ukraine

Signal de musique

- ▶ Variations de puissance importantes (dynamique de 90dB)
- ▶ Signal localement stationnaire
- ▶ Pas de modèles simples



Signal de musique

Normes de compression

Format CD : échantillonnage à 44.1 kHz, quantification à 16 bits par échantillon : 705 kbps en mono

MP3 : Il s'agit de la partie audio du standard MPEG-1. Trois couches, de qualité équivalente et de complexité croissante, à 192, 128 et 96 kbps

AAC : Partie audio de MPEG-2. Il est réputé le codeur plus performant à l'heure actuelle, avec une qualité "transparente" à 64 kbps

MPEG-4 : Représentation de sons d'origine quelconque (naturelle et synthétique), représentation des objet sonores.

Signal de musique

Normes de compression

Dans MPEG-4 on a plusieurs codeurs :

- ▶ Harmonic Vector eXcitation Coding (HVXC), signal de parole en bande téléphonique, débits entre 2 et 4 kbps
- ▶ CELP, signal de parole en bande téléphonique ou élargie
- ▶ Pour le signal de musique, une nouvelle version du AAC, pas très différentes de celui-ci
- ▶ Un codeur sans pertes (compression ≈ 2)
- ▶ Un algorithme de synthèse de la parole, un langage pour engendrer de la musique, un langage pour la description d'une scène audio

Approche de codage *par objet* : une scène audio peut être décomposée en plusieurs objets audio, chacun codé avec le codeur le plus adapté

Évaluation de la qualité

- ▶ Les critères objectives (fonctions mathématiques comme l'EQM) ne sont pas satisfaisants : pour un même niveau de bruit le résultat peut être très différent (forme spectrale du bruit, masquage)
- ▶ Tests subjectifs :
 - ▶ Codeurs de parole : tests d'intelligibilité
 - ▶ Codeurs de musique : critère de "transparence". Méthode doublement aveugle à triple stimulus et référence dissimulée (Norme UIT-T BS.1116)
 - ▶ Codeurs de musique à qualité intermédiaire : autres tests subjectifs (Norme UIT-T BS.1534-1)
 - ▶ Quelques test objectif donne des résultats significatifs (Norme UIT-T BS.1387-1)



Plan

Introduction

Perception de l'audio

L'oreille

Seuil d'audition

Le masquage

Compression de la parole

Compression de la musique

Audition

L'oreille

- ▶ Oreille externe (pavillon, conduit auditif)
- ▶ Oreille moyenne (chaîne ossiculaire, tympan)
- ▶ Oreille interne (cochlée: 3,5cm; membrane basilaire)

Oreille externe et oreille moyenne : filtre passe-bande
(20Hz ÷ 20kHz)

La membrane basilaire est densément innervée

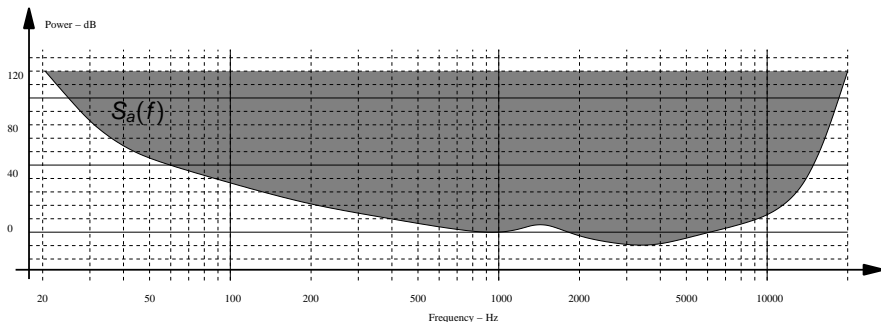
Perception d'un son pur

- ▶ Son pur : $x(t) = a \sin(2\pi f_1 t)$ sinusoïde de puissance $\sigma^2 = \frac{a^2}{2}$
- ▶ Ce son excite plusieurs fibres nerveuses (étalement de la puissance)
- ▶ Modèle : banc de M filtres
 - ▶ Le k filtre correspond à la k -ème fibre nerveuse
 - ▶ La réponse en fréquence du k -ième filtre est
 $H_k(f) = A_k(f) \exp(j\phi_k(f))$
 - ▶ La réponse à la sinusoïde à fréquence f_1 est :

$$y_k(t) = aA_k(f_1) \sin[2\pi f_1 t + \phi_k(f_1)]$$

- ▶ Le rapport entre les puissances est la fonction d'étalement :
 $S_E(k) = A_k^2(f_1)$

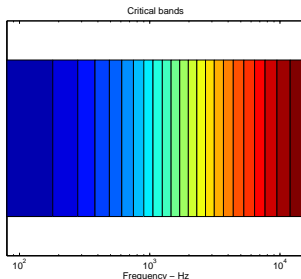
Seuil d'audition



- ▶ La gamme de fréquence audible est comprise entre 20Hz et 20kHz
- ▶ La puissance minimale nécessaire pour que le son soit audible est $S_a(f)$
- ▶ $S_a(f)$ varie avec la fréquence et a un minimum entre 1 et 4kHz (parole)

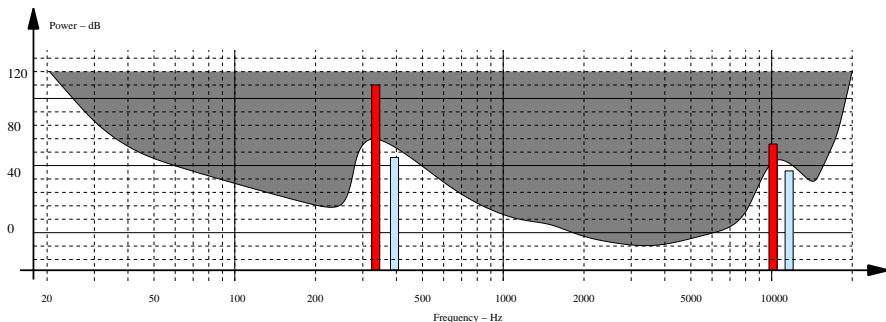
Bande critique (BC)

- ▶ Une sinusoïde de fréquence f_1 doit avoir puissance $\sigma_1^2 > S_a(f_1)$ pour être audible
- ▶ Pour N sinusoïdes de fréquence *proche* à f_1 il suffit que $\sum_i \sigma_i^2 > S_a(f_1)$
- ▶ Les sinusoïdes sont *proches* si sont dans la *bande critique*
- ▶ L'amplitude de la BC varie avec f_1
- ▶ La BC donc est la largeur de bande des filtres de la cochlée



Courbes de masquage

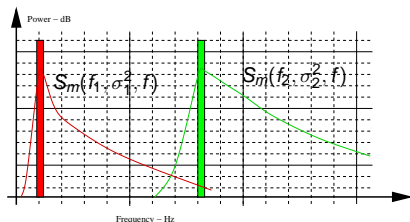
Masquage fréquentiel



- ▶ Le son masquant (rouge) réduit la sensibilité à un deuxième son
- ▶ On définit $S_m(f_0, \sigma^2, f)$ la puissance minimale pour un son pur à fréquence f pour ne pas être masqué par un son pur à f_0 et de puissance σ^2 , avec $\sigma^2 > S_a(f_0)$
- ▶ La même courbe est valable pour du bruit à bande étroite

Fonction de masquage fréquentiel

$$S_m(f_0, \sigma^2, f)$$

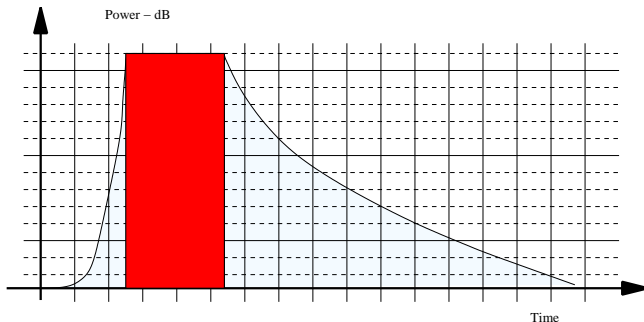


- ▶ Pour f_0 et σ^2 donnés, $S_m(f)$ a une allure triangulaire
- ▶ Le maximum est pour $f = f_0$
- ▶ Indice de masquage:
 $S_m(f, \sigma^2, f) - \sigma^2$

- ▶ On observe que $S_m(f, \sigma^2, f) < \sigma^2$ (le deuxième son ne doit pas forcément être plus puissant du premier)
- ▶ Le décroissance est moins rapide quand f_1 augmente
- ▶ La pente de décroissance est proportionnelle à la BC
- ▶ La pente vers le fréquence supérieures est fonction (décroissante) de σ^2

Courbes de masquage

Masquage temporel



- ▶ Pré-masquage : $2 \div 5$ ms
- ▶ Post-masquage : $100 \div 200$ ms

Applicabilité du modèle

- ▶ Le modèle psychoacoustique permet de déterminer certains parties du signal non-audibles
- ▶ On permet au bruit de quantification de monter en puissance à condition de rester non-audible
- ▶ Tout de même, le modèle est loin d'être parfait :
 - ▶ Seul les sons pur ou à bande étroite sont considérés
 - ▶ On est capable de évaluer l'influence réciproque de pas plus que 3 sons à la fois
 - ▶ Les signaux réels sont composés de très nombreuses contributions : comment interagissent-elles ?
- ▶ En pratique, les paramètres des algorithmes de compression de son sont déterminés de façon expérimentale, après un grand nombre de tests



Plan

Introduction

Perception de l'audio

Compression de la parole

Codeurs simples

Codeurs CELP

Codeur 3GPP AMR-WB

Compression de la musique

Codeurs MIC et MICDA

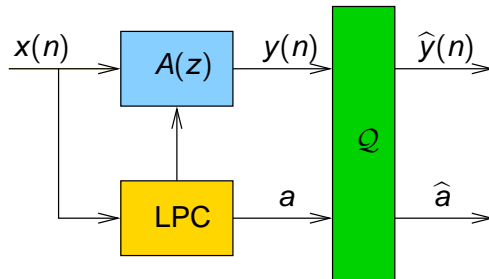
Modulation par Impulsions Codées (Différentielle Adaptive)

- ▶ Termes anglophones : PCM (Pulse Coded Modulation) et ADPCM (Adaptive Differential PCM)
- ▶ Débit de référence : $f_e = 8\text{kHz}$, quantification sur 12 bits, $\Rightarrow 96\text{ kbps}$
- ▶ MIC à 64 kbps : quantification scalaire non uniforme (Max-Lloyd)
- ▶ MICDA à 32 kbps : quantification scalaire prédictive en boucle fermée

Codeur LPC10 à 2.4 kbps

Linear Prediction Coding avec 10 échantillons

- ▶ Intérêt pédagogique, pas utilisé en pratique
- ▶ Fenêtre d'analyse de $N = 160$ échantillons
- ▶ Prédiction de l'échantillon basé sur les statistiques de la fenêtre et sur les $P = 10$ échantillons précédents
- ▶ Schéma de principe



Codeur LPC10 à 2.4 kbps

Filtre

$$y(n) = x(n) - x_P(n) \qquad x_P(n) = \sum_{k=1}^P h_k x(n-k)$$

$$y(n) = \sum_{k=0}^P a_k x(n-k) \qquad a_0 = 1 \qquad a_k = -h_k$$

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_P z^{-P} \qquad Y(z) = A(z)X(z)$$

Calcul de A : minimisation de la puissance du résidu dans la fenêtre courante

Codeur LPC10 à 2.4 kbps

Filtre

$$\mathbf{c} = [r_X(1) \ r_X(2) \ \dots \ r_X(P)]^T$$
$$\mathbf{R} = \begin{bmatrix} r_X(0) & r_X(1) & \dots & r_X(P-1) \\ r_X(1) & r_X(0) & \dots & r_X(P-2) \\ \dots & \dots & \dots & \dots \\ r_X(P-1) & r_X(P-2) & \dots & r_X(0) \end{bmatrix}$$
$$\mathbf{a} = -\mathbf{R}^{-1}\mathbf{c}$$

Estimation de r_X : sur les N échantillon de la fenêtre courante

$$\hat{r}_X(k) = \frac{1}{N-k} \sum_{n=k}^{N-1} x(n)x(n-k) \quad k \in \{0, 1, \dots, P\}$$

Codeur LPC10 à 2.4 kbps

Codage des sons non-voisés

- ▶ Si on avait éliminé toute corrélation de X , Y ne serait que du bruit blanc
- ▶ Il n'est pas donc nécessaire d'envoyer Y : en synthèse on produit du bruit blanc et on filtre avec $1/A(z)$
- ▶ Résultat : distorsion de phase, pas audible
- ▶ Le résidu y est blanc avec bonne approximation seulement pour les sons non-voisés
- ▶ Pour les sons voisés il reste une périodicité (vibration des cordes vocales)

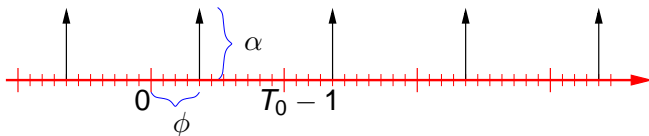
Codeur LPC10 à 2.4 kbps

Représentation des sons voisés/non-voisés

Sons non-voisés : Filtrage par le filtre $1/A(z)$ d'un bruit blanc

Sons voisés : Filtrage par le filtre $1/A(z)$ du peigne de Dirac :

$$\hat{y}(n) = \alpha \sum_{m \in \mathbb{Z}} \delta(n - mT_0 + \phi)$$



Codeur LPC10 à 2.4 kbps

Détermination des sons voisés/non voisés et des paramètres de modèle

- ▶ Estimation de la fonction de auto-corrélation $\hat{r}_x(k)$
- ▶ $\hat{r}_x(0)$ permet d'estimer la puissance σ_Y^2 ou le paramètre α
- ▶ Si \hat{r}_x décroît rapidement vers zéro, c'est un son non voisé
- ▶ Si \hat{r}_x est périodique, c'est un son voisé, et la période nous donne T_0

Codeur LPC10 à 2.4 kbps

Contrainte de débit

Tout le 20 ms le codeur doit envoyer :

- ▶ Les P coefficients du filtre A . $P = 10$ et les coefficients sont représentés sur 3 ou 4 bits. En moyenne $b_P = 36$ bits
- ▶ Un bit pour la distinction voisé/non-voisé : $b_v = 1$ bit
- ▶ Pour les sons voisés :
 - ▶ La puissance des impulsions α sur 6 bits (dynamique de 50dB) : $b_{\alpha} = 6$
 - ▶ La période fondamentale T_0 , comprise dans une dynamique de 7 bits : $b_{T_0} = 7$
- ▶ Pour les sons non-voisés :
 - ▶ La puissance du bruit sur 6 bits (dynamique de 50dB) : $b_{\sigma^2} = 6$ bits

Codeur LPC10 à 2.4 kbps

Contrainte de débit

En total, pour les sons voisés :

$$\begin{aligned} R &= (b_P + b_V + b_{T_0} + b_\alpha) / (20\text{ms}) \\ &= (36 + 1 + 6 + 7) / 0.02 \\ &= 2500\text{bps} \end{aligned}$$

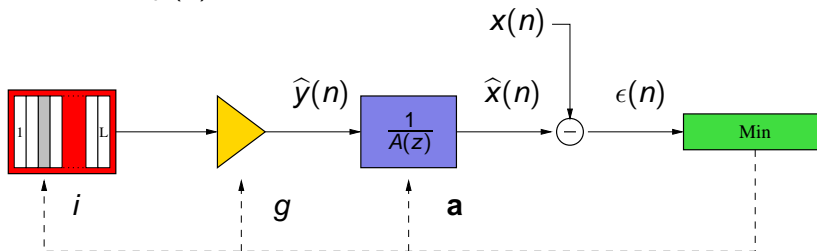
Pour les non voisés :

$$\begin{aligned} R &= (b_P + b_V + b_{\sigma^2}) / (20\text{ms}) \\ &= (36 + 1 + 6) / 0.02 = 2150\text{bps} \end{aligned}$$

En moyenne, $R = 2.4 \text{ kbps}$

Codeur CELP

- ▶ Déterminer un filtre et un signal d'erreur de prédiction
- ▶ Filtre $A(z)$: principes du LPC
- ▶ Erreur $y(n)$: choisi dans un dictionnaire "GS-VQ"



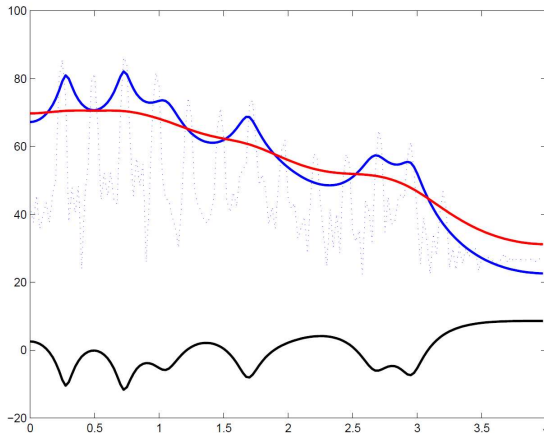
Codeur CELP

- ▶ Codage des coefficients du filtre
 - ▶ *Line Spectrum Pairs* : Représentation du polynôme $A(z)$ par des point sur le cercle unitaire, dont la phase est codée en différentiel
- ▶ Fonction de pondération perceptuelle
 - ▶ Le bruit est moins important ou le signal est fort
 - ▶ Fonction de poids qui dépende de $A(z)$

Ponderation perceptuelle

- ▶ On utilise le filtre $W(z) = \frac{A(z)}{A(z/\gamma)}$ pour ponderer le signal avant de minimiser l'EQM
- ▶ Le filtre $W(z)$ permet d'avoir un bruit plus important là où le signal de parole est plus fort (zones formantiques)
- ▶ Le filtre $1/A(z)$ a des pics en correspondance des formatiques
- ▶ Le filtre $1/A(z/\gamma)$, avec $\gamma \in (0, 1)$ a des pics aux mêmes fréquences, mais moins prononcés
 - ▶ Car, si les poles de $1/A(z)$ sont p_i , ceux de $1/A(z/\gamma)$ sont γp_i , donc sont ramenés vers le centre du cercle unitaire.
- ▶ Le rapport entre les deux donne donc la ponderation souhaitée

Ponderation perceptuelle



Bleu : $1/A(z)$; Rouge : $1/A(z/\gamma)$; Noir : $W(z) = A(z)/1/A(z/\gamma)$

Codeur CELP

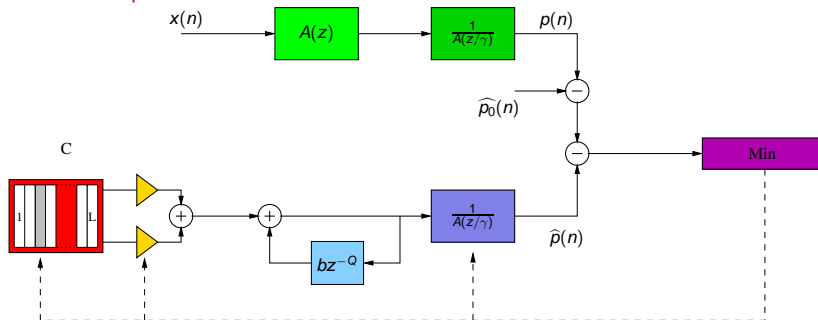
- ▶ Choix du modèle d'excitation
 - ▶ On peut choisir une somme de K vecteurs pris de dictionnaires différents
 - ▶ Normalement K est égal à 2 ou à 3
- ▶ Choix des vecteurs et du gain
 - ▶ Complexité très élevée
 - ▶ Algorithmes sous optimaux
 - ▶ Algorithme itérative standard : d'abord on optimise par rapport au premier des K vecteurs, en suite on optimise le résidu par rapport au deuxième, et ainsi de suite

Codeur CELP

- ▶ Construction du dictionnaire
 - ▶ Impulsions uniformément étalées
 - ▶ Tirage de v.a. Gaussiennes centrées
 - ▶ Codes avec structure algébrique (G.729)
- ▶ Dictionnaire adaptative
 - ▶ On essaye de réduire la périodicité résiduelle
 - ▶ On estime la période et la constant de temps

Codeur CELP

Schéma complet



Codeur CELP

Débit de codage

- ▶ Coefficients du filtre, actualisé tout le 10 ms :
 - ▶ $P = 10$. Avec le codage *Line Spectrum Pairs*, $b_P = 18$ bits
- ▶ Prédicteur à long terme, actualisé tout le 5 ms :
 - ▶ Période ("pitch") codé sur 7 bits
 - ▶ Puissance, codée sur 3 bits
- ▶ Résidu, codé par GS-VQ, actualisé tout le 5 ms :
 - ▶ *Shape* : dictionnaire à 17 bits
 - ▶ *Gain* : codé sur 4 bits

$$\begin{aligned} R &= 18/0.010 + (7 + 3 + 17 + 4)/0.005 \\ &= 8 \text{ kbps} \end{aligned}$$

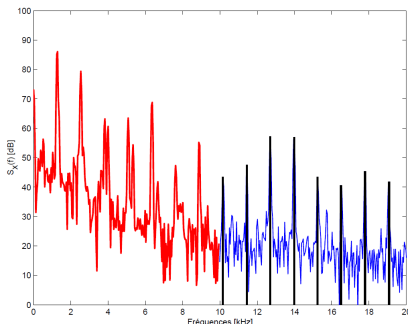
Codeur 3GPP AMR-WB

UIT-T G.722.2

- ▶ État de l'art en codage de parole
- ▶ Introduction 50-200 Hz : voix plus naturelle, amélioration de l'effet de présence
- ▶ Extension 3.4-7 kHz : plus grande intelligibilité
- ▶ Premier codeur adopté pour les réseaux fixes ou mobiles (suppression des transcodages)
- ▶ Codeur de type ACELP très comparable au G.729 mais :
 - ▶ modification du filtrage perceptuel (extension à la bande élargie)
 - ▶ modification de l'exploitation de l'information de pitch (pas de structure harmonique sur toute la bande)
 - ▶ introduction d'un très grand dictionnaire d'excitation ($\log_2 L = 88 \text{ bits}$)
- ▶ De 6.6 à 23.85 kbps

Codeur 3GPP AMR-WB

- ▶ Génération d'énergie dans la zone fréquentielle considérée (exploitation des harmoniques)
- ▶ Modulation en amplitude de cette énergie (utilisation de l'enveloppe spectrale)



Plan

Introduction

Perception de l'audio

Compression de la parole

Compression de la musique

MP3

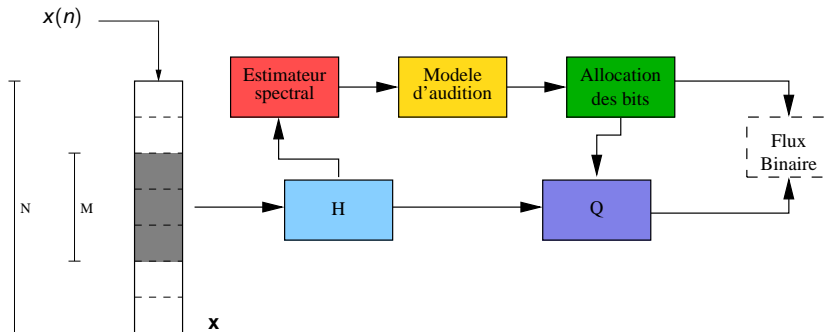
AAC

Codeurs perceptuels

- ▶ Codage par fenêtres à recouvrement
- ▶ Buffer de N échantillons ; M nouveaux entrent dans les buffer et sont codés
- ▶ Exemples : $M = 32$ et $N = 512$ (MP3) ; $M = 1024$ et $N = 2048$ (AAC)
- ▶ Trois modules qui sont actives pour chaque fenêtre d'analyse
 - ▶ Transformation temps-fréquence
 - ▶ Allocation de bits (sous le contrôle d'un modèle d'audition)
 - ▶ Quantification (scalaire ou vectorielle)/codage sans pertes

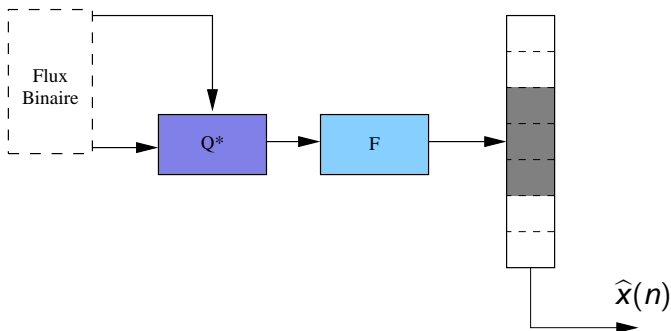
Codeurs perceptuels

Codeur



Codeurs perceptuels

Décodeur



Codeurs perceptuels

- ▶ Transformée temps-fréquence : analyse spectrale (M-DCT) pour chaque fenêtre temporelle
- ▶ Principe : faire en sorte que le bruit de quantification reste inaudible
- ▶ Basé sur un modèle d'audition

Codeur de musique MPEG-1/MP3

- ▶ Codeur de musique *transparent* : qualité subjective parfaite
- ▶ Trois “couches” de complexité, correspondants à des débits de plus en plus faibles
 - ▶ MP3 : codeur MPEG-1, 3-ème couche
- ▶ Trois fréquences d'échantillonnage sont possibles, mais on fera référence au cas $f_e = 44.1\text{kHz}$

Codeur de musique MPEG-1/MP3

Transformée temps-fréquence

- ▶ Banque de 32 filtres
- ▶ Répartition uniforme des fréquences entre 0 et 22 kHz
 - ▶ À peu près 700 Hz par sous-bande
- ▶ Sous-échantillonnage critique et reconstruction quasi-parfaite ($\text{SNR} > 90\text{dB}$ en absence de quantification)
- ▶ Dans chaque sous-bande on regroupe 12 échantillon, en on les code conjointement
- ▶ Le vecteur de 12 coefficients de la sous-bande k est indiqué comme \mathbf{y}_k
- ▶ Chaque vecteur correspond à environ une dizaine de ms

Codeur de musique MPEG-1/MP3

Représentation des sous-bandes

- ▶ Normalisation des vecteurs de sous-bande :

$$\mathbf{y}_k = g_k \mathbf{a}_k$$

- ▶ g_k : facteur d'échelle, correspondants à la valeur absolue la plus élevée entre les composantes, et quantifiée sur 6 bit
- ▶ \mathbf{a}_k : vecteur normalisé (valeurs entre -1 et +1)
- ▶ Les échantillons de x de la fenêtre courante sont aussi utilisé pour calculer $\hat{S}_X(f)$ et un seuil de masquage $\Phi(f)$ basé sur un modèle psychoacoustique

Codeur de musique MPEG-1/MP3

Allocation de débit et quantification

- ▶ Pour chaque sous-bande on connaît le rapport signal sur masque
- ▶ On alloue les bits disponible en donnant d'abord au sous-bandes avec le plus grand rapport signal sur masque et en suite aux autres (algorithme *greedy*)
- ▶ On choisit donc le nombre de bits utilisé pour coder la sous-bande k , pour tout k
- ▶ La sous-bande est codé avec un codeur scalaire uniforme
 - ▶ On choisit entre 16 quantificateurs (c'est-à-dire, débits) pour les premières 11 sous-bandes, 8 pour les 12 suivantes et 4 pour les 4 dernières. Les sous-bandes 27 à 32 ne sont jamais codées

Codeur de musique MPEG-2 AAC

- ▶ C'est aussi un codeur perceptuel
- ▶ Transformée M-DCT avec $N=2048$ et $M=1024$
- ▶ Le vecteur des coefficients de la transformé est

$$\mathbf{X} = [X(0) X(1) \dots X(M-1)]$$

- ▶ Il est représenté par un couple de vecteurs :
 - ▶ Facteurs d'échelle :

$$\mathbf{g} = [g(0) g(1) \dots g(M-1)]$$

- ▶ Valeurs normalisées :

$$\mathbf{i} = \left[\frac{X(0)}{g(0)} \frac{X(1)}{g(1)} \dots \frac{X(M-1)}{g(M-1)} \right]$$

Codeur de musique MPEG-2 AAC

Vecteur normalisé

- ▶ Les éléments du vecteur normalisé sont quantifiés sur 3 bits
- ▶ En suite ils sont groupé par demi-bandes critiques (51 groupes)
- ▶ Dans chaque groupe, on utilise un codeur de Huffman qui dépend de la valeur de la composante plus importante dans le groupe

Codeur de musique MPEG-2 AAC

Facteurs d'échelle

- ▶ Problème d'optimisation : déterminer \mathbf{g} qui minimise l'erreur de reconstruction
 - ▶ Contrainte : débit maximale donné
- ▶ Dans ce cas on obtient des taux de compression faibles (≈ 2)
- ▶ On comprime plus si on ajoute la contrainte perceptuelle : la puissance du bruit est inférieure au seuil de masquage

$$S_Q(F) < \Phi(f)$$

- ▶ Algorithme du gradient pour trouver les \mathbf{g} optimaux ; ils sont groupés comme les \mathbf{i} et codés en différentiel avec Huffman

Codeur MPEG-4 HE AAC

- ▶ High Efficiency Advanced Audio Coder
- ▶ État de l'art dans le codage de musique
- ▶ Développé par ISO, 3GPP, ETSI
- ▶ Basé sur MPEG-2 AAC, plus des outils :
 - ▶ Spectral Band Replication, pour élargir la largeur de bande représentée
 - ▶ Parametric Stereo, pour optimiser les cas de canaux multiples (du 2.0 au 5.1)

MPEG Unified Speech and Audio Codec

- ▶ ISO/IEC (JTC1/SC29/WG11) : Call for Proposals en 2007
- ▶ Codeur hybride : 2 modes distincts + un classifieur
 - ▶ Utiliser 3GPP AMR-WB pour la parole et MPEG-4 HE AAC pour la musique
- ▶ Difficultés :
 - ▶ assurer des transitions rapides et douces entre parole, musique et signaux mixtes
 - ▶ exploiter différents types de fenêtres, de différentes dimensions suivant le type des signaux

Conclusions

Codage de parole

- ▶ Modèle psychoacoustique
- ▶ Modèle de source et de filtre
- ▶ Quantification vectorielle

Codage de musique

- ▶ Modèle psychoacoustique
- ▶ Allocation du débit
- ▶ Optimisation avec contrainte en boucle fermée