Modification de hauteur, d’échelle temporelle et de timbre

Signal audiofréquence et parole
SI 340 - Module Audio

Roland Badeau

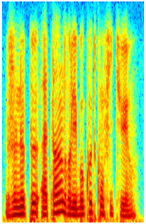
TELECOM ParisTech

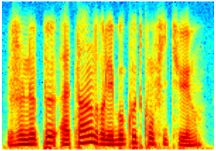
Partie I

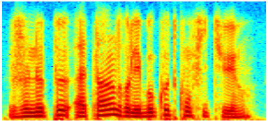
Introduction

Introduction

Changement de la vitesse de lecture d'un son



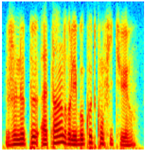


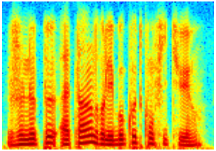


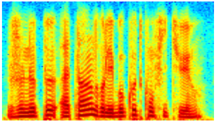
Origine du problème : $y(t) = x(\alpha t) \Leftrightarrow Y(f) = \frac{1}{|\alpha|} X(\frac{f}{\alpha})$

Introduction

But : contrôler séparément les échelles temps et fréquence







Trois catégories de méthodes

- méthodes temporelles : SOLA, PSOLA
- méthodes fréquentielles : vocodeur de phase (TFCT)
- méthodes paramétriques : LPC, sinus+bruit

Partie II

Définitions

Modèle de production vocale

- Modèle source / filtre linéaire variant dans le temps :
  $x(t) = \int_{-\infty}^{+\infty} g(t,\tau)\, e(t-\tau)\, \mathrm{d}\tau$
- Réponse en fréquence du filtre :
  $G(t,f) = \int_{-\infty}^{+\infty} g(t,\tau)\, e^{-j2\pi f\tau} \mathrm{d}\tau = M(t,f)\, e^{j\varphi(t,f)}$
- Source harmonique : $e(t) = \sum\limits_{k=1}^{L} e^{j\xi_k(t)}$, où $\frac{\mathrm{d}\xi_k}{\mathrm{d}t} = 2\pi f_k(t)$
- Hypothèse de quasi-stationnarité : $\xi_k(t-\tau) \simeq \xi_k(t) - 2\pi f_k(t)\tau$
- Signal filtré : $x(t) = \sum\limits_{k=1}^{L} M(t, f_k(t))\, e^{j(\xi_k(t) + \varphi(t, f_k(t)))}$

Modèles de signaux

Modèle de McAulay et Quatieri (codage de la parole)
$x(t) = \sum_{k=1}^{L} A_k(t)\, e^{j\Psi_k(t)}$ où $\frac{\mathrm{d}\Psi_k}{\mathrm{d}t} = 2\pi f_k(t)$
et $A_k(t)$ et $f_k(t)$ sont à variation lente devant $e^{j\Psi_k(t)}$

Modèle de Serra et Smith (synthèse de signaux de musique)
$x(t) = \sum_{k=1}^{L} A_k(t)\, e^{j\Psi_k(t)} + b(t)$
où $b(t)$ est un bruit blanc filtré par un filtre variant dans le temps

Système complet d'analyse / modification / synthèse

- estimation des composantes déterministes
- interpolation linéaire des amplitudes et cubique des phases
- soustraction de la partie déterministe pour obtenir $b(t)$
- transformation de chacune des deux composantes
- resynthèse

Modifications d'échelles

Distortion temporelle

- Fonction de distortion temporelle : $\tau = T(t)$
- Signal modifié : $y(\tau) = \sum_{k=1}^{L} A_k(T^{-1}(\tau))\, e^{j\phi_k(\tau)}$
- Conservation des fréquences : $\phi_k(\tau) = 2\pi \int_0^{\tau} f_k(T^{-1}(u))du$

Modification de hauteur

- Taux de compression fréquentiel : $\alpha(t)$
- Signal modifié : $y(t) = \sum_{k=1}^{L} A_k(t)\, e^{j\Phi_k(t)}$
- Altération des fréquences : $\Phi_k(t) = 2\pi \int_0^{t} \alpha(u) f_k(u)du$

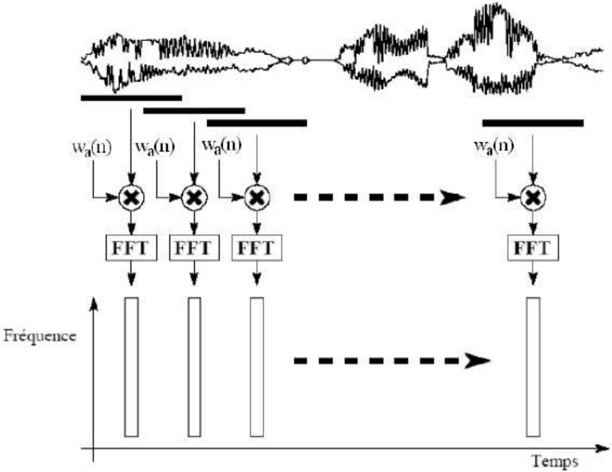
Réciprocité

- distortion temporelle T plus ré-échelonnement T^{-1}
  $\Leftrightarrow$ modification de hauteur de taux $\alpha(t) = T'(t)$

Partie III

Transformée de Fourier à Court Terme

Schéma de principe



Rappels théoriques

Définition : $\widetilde{X}(t_a, \nu) = \sum\limits_{n \in \mathbb{Z}} x(n + t_a)\, w_a(n)\, e^{-j2\pi\nu n}$, où

- la fenêtre d'analyse $w_a(n)$ est finie, réelle et symétrique
- les instants d'analyse t_a sont indexés par un entier u

Interprétation : convention passe-bande

- $\widetilde{X}(t_a, \nu_p) = [x \star h](t_a)$ où $h(n) = w_a(-n)\, e^{j2\pi\nu_p n}$
- la TF de $h(n)$ est $H(e^{j2\pi\nu}) = W_a\left(e^{j2\pi(\nu_p - \nu)}\right)$

Version discrète de la TFCT : on pose $\nu_p = \frac{p}{N}$

- $\widetilde{X}(t_a, \nu_p) = \sum\limits_{n=0}^{N-1} x(n + t_a)\, w_a(n)\, e^{-j2\pi\frac{pn}{N}}$
- la longueur des fenêtres d'analyse doit être $\leq N$

Filtrage passe-bande équivalent

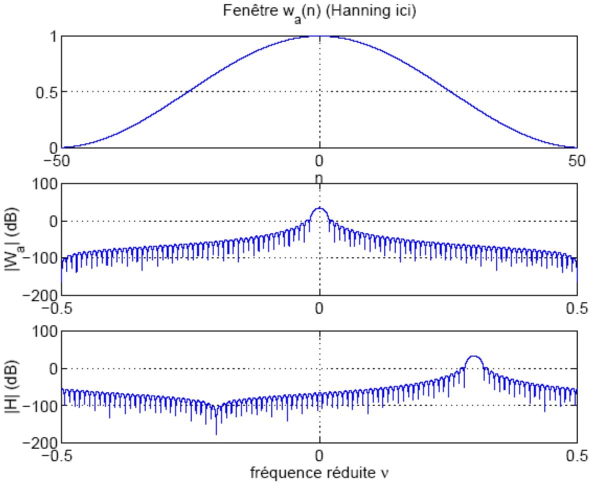
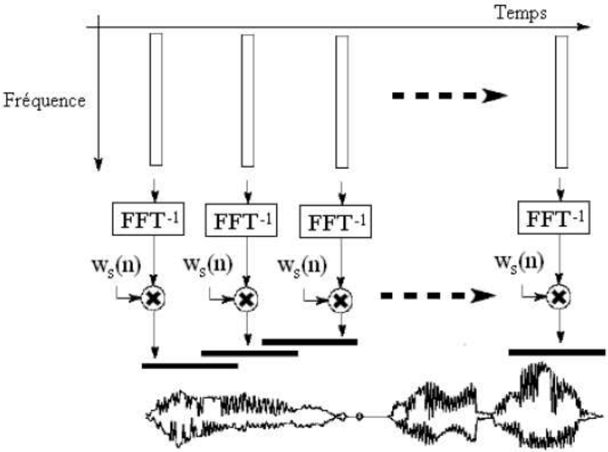


Schéma de synthèse



Reconstruction du signal

Modifications et problèmes posés :

- $t_a \longrightarrow t_s$, $\widetilde{X}(t_a(u), \nu_p) \longrightarrow Y(t_s(u), \nu_p)$
- Difficulté : Y n'est généralement pas la TFCT d'un signal

Condition de reconstruction parfaite ($t_s = t_a$ et $Y = \widetilde{X}$)

- synthèse par addition-recouvrement (overlapp-add ou OLA)

$$y(n) = \sum_u w_s(n - t_s(u))\, y_w\, (n - t_s(u), t_s(u))$$

où $\mathrm{supp}(w_s) \subset [0, N-1]$ et $y_w(n, t_s(u)) = \frac{1}{N} \sum_{p=0}^{N-1} Y(t_s(u), \nu_p)\, e^{j2\pi\nu_p n}$

- condition suffisante : $\sum_u w_a(n - t_a(u))\, w_s(n - t_a(u)) \equiv 1$

Partie IV

Vocodeur de phase

Fréquence instantanée

- Modèle de McAulay et Quatieri : $x(t) = \sum_{k=1}^{L} A_k(t)\, e^{j\Psi_k(t)}$
- **Hypothèses de quasi-stationnarité** : $\forall n \in \{0 \dots N-1\}$
  $$\begin{cases} A_k(n+t_a) & \simeq & A_k(t_a) \\ \Psi_k(n+t_a) & \simeq & \Psi_k(t_a) + 2\pi f_k(t_a) n \end{cases}$$
- Alors $\widetilde{X}(t_a(u), \nu_p) = \sum_{k=1}^{L} A_k(t_a)\, e^{j\Psi_k(t_a)}\, W_a\left(e^{j2\pi(\nu_p - f_k(t_a))}\right)$
- Soit f_c la fréquence de coupure du filtre passe-bas $w_a(n)$
- **Condition de bande étroite** : $\exists!\ l$ tel que $|\nu_p - f_l(t_a)| \leq f_c$
  Interprétation (spectre harmonique) : $N \geq \frac{4}{f_0}$
- Alors $\widetilde{X}(t_a(u), \nu_p) = A_l(t_a)\, e^{j\Psi_l(t_a)}\, W_a\left(e^{j2\pi(\nu_p - f_l(t_a))}\right)$
  $\Rightarrow$ la TFCT donne accès aux phases $\Psi_l(t_a)$ modulo 2π

Condition de recouvrement

Levée de l'indétermination de la phase modulo 2π :

- Différence de phases entre deux instants successifs :
  $\Delta\Phi_p = 2\pi(f_l(t_a) - \nu_p)\Delta t_a(u) + 2\pi\nu_p\Delta t_a(u) + 2n\pi$
- Condition de recouvrement minimal : $f_c\,\Delta t_a(u) < \frac{1}{2}$
  Interprétation (fenêtre de Hanning) : $f_c = \frac{2}{N} \Rightarrow \Delta t_a < \frac{N}{4}$
- $\exists!\, n$ tel que $|\Delta\Phi_p - 2\pi\nu_p\Delta t_a(u) - 2n\pi| < \pi$

Estimation de la fréquence instantanée $\forall p \in \{0 \ldots N-1\}$

1. calcul de la TFCT à deux instants successifs $\longrightarrow \Delta\Phi_p$
2. calcul de $Q(n_0) = \Delta\Phi_p - 2\pi\nu_p\Delta t_a - 2n_0\pi$ tel que $|Q(n_0)| < \pi$
3. calcul de la fréquence instantanée $f_l(t_a) = \nu_p + \frac{Q(n_0)}{2\pi\Delta t_a}$

Distortion temporelle

Déroulement des phases instantanées pour une distortion $T(t)$

Algorithme de modification :

1. calcul de la TFCT et de $f_p(t_a(u))$ dans chaque canal
2. calcul du nouvel instant de synthèse $t_s(u) = T(t_a(u))$
3. calcul de la phase instantanée de synthèse
   $\Phi_s(t_s(u+1), \nu_p) = \Phi_s(t_s(u), \nu_p) + 2\pi f_p(t_a(u))\,(t_s(u+1) - t_s(u))$
4. calcul de la TFCT de synthèse en $u+1$
   $\widetilde{Y}(t_s(u+1), \nu_p) = A_p(t_a(u+1))\, e^{j\Phi_s(t_s(u+1),\nu_p)}$

Modification de hauteur

Méthode de rééchantillonage temporel

1. étirement temporel de distortion $T(t) = \int_0^t \alpha(u)du$
2. ré-échelonnement de taux $T^{-1}(\tau)$

Méthode de rééchantillonage spectral

1. interpolation linéaire de la TFCT d'analyse
   - $\alpha(t_a) > 1$: perte d'information en hautes fréquences
   - $\alpha(t_a) < 1$: complétion du spectre en hautes fréquences
2. resynchronisation des phases en synthèse

Problème en traitement de la parole : effet "Donald Duck"

- estimation de l'enveloppe du spectre (LPC) et "blanchiment"
- modification de hauteur, puis filtrage inverse

Partie V

Méthodes paramétriques

Prédiction linéaire

Modèle source / filtre : $x(t) = \int_{-\infty}^{+\infty} g(t,\tau)\, e(t-\tau)\, \mathrm{d}\tau$

Paramétrisation au voisinage d'une marque d'analyse $t_a(u)$

- g^u est un filtre autorégressif (obtenu par analyse LPC)
- partie voisée : e^u est une série d'impulsions de fréquence f_1^u
- partie bruitée : e^u est un bruit blanc de variance ${\sigma^u}^2$

Synthèse sans modification : par addition / recouvrement

Synthèse avec modification de hauteur

- partie voisée : on modifie la fréquence des impulsions
- partie bruitée : inchangée

Synthèse avec modification de durée

- les marques peuvent être dupliquées / supprimées
- les paramètres (g^u, f_1^u, ${\sigma^u}^2$) peuvent être interpolés

Modèle Harmoniques plus Bruit (HNM)

Modèle de Serra et Smith : $x(t) = \sum_{k=1}^{L} A_k(t)\, e^{j\Psi_k(t)} + b(t)$

Paramétrisation du signal autour d'une marque $t_a(u)$:

- partie voisée : on estime A_k^u, Ψ_k^u, f_k^u pour chaque partiel
  - par extraction des pics de la TFCT
  - par détection de hauteur (f_k^u) et moindres carrés (A_k^u, Ψ_k^u)
- partie bruitée : modèle auto-régressif (analyse LPC)

Identification avec le modèle source-filtre :

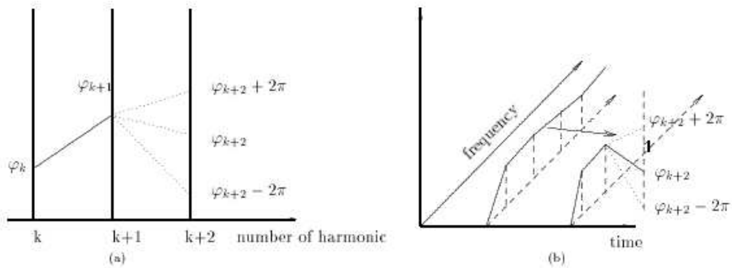
- amplitude du filtre : $M(t_a(u), f_k^u) = A_k^u$
- phase du filtre : $\varphi(t_a(u), f_k^u) = \Psi_k^u - \xi_k(t_a(u))$

Modification de la partie voisée : interpolation temporelle et fréquentielle de l'amplitude et de la phase du filtre

Estimation de la phase du filtre

Modèle source filtre : $\varphi\left(t_a(u), f_k^u\right) = \Psi_k^u - \xi_k(t_a(u))$
Déroulement de la phase modulo 2π :

$$\Delta\varphi_k^u = \left\langle \Delta\Psi_k^u - 2\pi\Delta t^u \bar{f}_k^u \right\rangle = (\Delta\Psi_k^u + 2\pi M_k^u) - 2\pi\Delta t^u \bar{f}_k^u$$



Reconstruction de la phase par interpolation linéaire

Estimation de l'amplitude du filtre

Modèle cepstral : $\log\left(M(t_a(u), f)\right) = \gamma_0 + 2\sum_{m=1}^{p} \gamma_m \cos(2\pi m f)$

Approximation directe : $\varepsilon = \sum_{k=1}^{L} \left(\log(A_k^u) - \log\left(M(t_a(u), f_k^u)\right)\right)^2$

Solution : $\gamma = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\boldsymbol{a}$, où $\boldsymbol{a} = [\log(A_1^u)\ldots\log(A_L^u)]^T$ et

$$\mathbf{C} = \begin{bmatrix} 1 & 2\cos(2\pi f_1^u) & \ldots & 2\cos(2\pi p f_1^u) \\ \vdots & \vdots & \ldots & \vdots \\ 1 & 2\cos(2\pi f_L^u) & \ldots & 2\cos(2\pi p f_L^u) \end{bmatrix}$$

Régul. : $\varepsilon = \sum_{k=1}^{L} \left(\log(A_k^u) - \log\left(M(t_a(u), f_k^u)\right)\right)^2 + \lambda\mathcal{R}(M^u(f))$

où $\mathcal{R}(M^u(f)) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \left(\frac{d}{df}\log\left(M(t_a(u), f)\right)\right)^2 df$

Solution : $\gamma = \left(\mathbf{C}^T\mathbf{C} + \lambda\mathbf{R}\right)^{-1}\mathbf{C}^T\boldsymbol{a}$, $\mathbf{R} = 8\pi^2\,\mathrm{diag}(0, 1^2, \ldots, p^2)$.

Synthèse du signal

Synthèse sans modification : interpolation temporelle

- Interpolation linéaire de l'amplitude
  $$A_k(t) = A_k^u \frac{t_s(u+1)-t}{t_s(u+1)-t_s(u)} + A_k^{u+1} \frac{t-t_s(u)}{t_s(u+1)-t_s(u)}$$
- Interpolation cubique de la phase $\Psi_k(t) =$
  $$\left(2\pi f_k^u (t - t_s(u)) + \Psi_k^u \left(1 + 2\frac{t-t_s(u)}{t_s(u+1)-t_s(u)}\right)\right)\left(\frac{t_s(u+1)-t}{t_s(u+1)-t_s(u)}\right)^2 +$$
  $$\left(2\pi f_k^{u+1} (t - t_s(u+1)) + \Psi_k^{u+1} \left(1 + 2\frac{t_s(u+1)-t}{t_s(u+1)-t_s(u)}\right)\right)\left(\frac{t-t_s(u)}{t_s(u+1)-t_s(u)}\right)^2$$
  où les Ψ_k^u sont les phases déroulées

Synthèse avec modification : interpolation temporelle et fréquentielle de l'amplitude et de la phase du filtre