

3. Detectie Pitch (F0) prin metoda cepstrala (Documentatie V.01)

3.1. Prezentare

Dintre numeroasele metode pentru detectia frecventei fundamentale (F0) întâlnite în literatura, s-au implementat detectia pitch-ului prin metoda ceptrala si cu ajutorul autocorelatiei. Rezultatele furnizate de cele doua metode sunt comparate si în cazul unor diferente prea mari între valorile obtinute se aplica un algoritm de selectie, ce va fi prezentat ulterior.

Aplicarea metoda cepstrale pentru detectia pitch-ului este justificata de faptul ca spectrul unei secvente rostite este produsul a doua componente spectrale – una care tine de modul în care este „generat” sunetul (de frecventa de rezonanta a corzilor vocale, dimensiunea tubului generator – laringelui) si o alta care tine de modul în care este „filtrat” semnalul generat.

Data fiind formula de calcul a cepstrului operatia de înmultire dintre spectrele semnalului excitator si cel al functiei de transfer este transformata prin aplicarea logaritmului într-o operatie de aditie.

$$\begin{aligned}H(\mathbf{w}) &= H_g(\mathbf{w}) \cdot H_f(\mathbf{w}) \\ \text{cepstrum} &= \text{IFFT}(\log|\text{FFT}(s)|) \\ \text{cepstrum} &= \mathfrak{S}^{-1}(\log|H_g(\mathbf{w}) \cdot H_f(\mathbf{w})|) = \mathfrak{S}^{-1}(\log|H_g(\mathbf{w})|) + \mathfrak{S}^{-1}(\log|H_f(\mathbf{w})|)\end{aligned}$$

Cele doua spectre fiind clar separabile este justificabila cautarea maximului cepstral corespunzator frecventei fundamentale, între valorile corespunzatoare benzii de frecvente [70Hz-500Hz]. Pentru voci provenite de la subiecti de sex masculin, valorile uzuale ale F0 sunt în jurul a 100-150Hz; pentru femei, respectiv copii acestea urcând la 250-350Hz.

La recomandarea de a folosi o fereastră de ponderare (alta decât cea dreptunghiulara) pentru a nu falsifica rezultatele spectrale generate si implicit rezultatele cepstrale, mai adaugam si recomandarea de a folosi ferestre de analiza de minim 1024 esantioane.

De exemplu pentru o frecventa de esantionare de $F_s=22050$ pentru banda de frecvente [70Hz-500Hz] avem nevoie ca vectorul cepstral (corespunzator frecventelor pozitive) sa contina minim $F_s/f_{low}=22050/70=315$ esantioane. Deoarece jumatate din spectrul unui semnal este asociat frecventelor pozitive, iar cealalta jumatate frecventelor negative dimensiunea minima a ferestrei de analiza este de $2 \cdot F_s/f_{low} = 650$ esantioane. Cum algoritmul transformatei Fourier rapide necesita ferestre de analiza de dimensiuni egale cu puteri ale lui 2 se vor analiza pentru fiecare fereastră 1024 esantioane.

Pentru eliminarea zgomotelor introduse de frecventa de retea, semnalul este filtrat cu un filtru „Butter” trece sus la frecventa de 70Hz.

Daca pentru o fereastră de analiza energia semnalului în timp nu depaseste de 10 ori energia minima sau a 20 parte din energia maxima calculate pentru toate ferestrele de analiza de dimensiune N, sau daca energia spectrala în banda [80Hz 2500 Hz] nu depaseste 70% din energia spectrala atunci se considera ca informatia continuta nu este „vocalica” si ca în acea zona nu este indicat sa se determine frecventa fundamentala F0.

3.2. Mod de lucru

Instrumentul este realizat sub forma unui executabil, numit "FO_cepst.exe". Acesta trebuie sa fie plasat în acelasi director cu fisierele se sunet (wav) ce vor fi analizate.

Utilizatorul selecteaza urmatoorii parametri:

- Numele fisierului (se poate alege din listing-ul fisierelor wav gasite în director). Dupa introducerea numelui fisierului de sunet sunt verificate informatiile din header si sunt afisate informatii cum ar fi frecventa de esantionare, numarul de canale, numarul de biti per esantion si numarul total de esantioane. Sunt acceptate doar fisierele de sunet mono-canal.
- Numarul de esantioane prelucrate – dimensiunea ferestrei de analiza (de ex. "1024");
- Pasul de deplasare (un întreg pozitiv nenul, conditionat ca valoare de numarul de esantioane din fisierul de sunet)
- Se selecteaza tipul de fereastră (implicit = 0, fereastră dreptunghiulara).

Datele – frecventele corespunzatoare pitch-ului sunt salvate sub forma unei coloane, dimensiunea vectorului de date generat fiind determinata pe baza pasului de deplasare-*step* si a numarului de esantioane prelucrate pe o fereastră de analiza *N*:

$$(nr_total_esantioane - N)/step + 1$$

Citare si Copyright

Programul a fost scris de Marius Zbancioc cu colaborarea lui Horia-Nicolai Teodorescu.

Programul poate fi utilizat gratuit în aplicatii educationale si de cercetare, cu citarea de mai jos:

Marius Zbancioc, Horia-Nicolai Teodorescu: Aplicatia "Detectie Pitch (F0) prin metoda cepstrala". Instrumente pentru Arhiva Limbii Române Vorbite – Sunetele Limbii Române.

http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/ro/instrumente.htm.

În lipsa unei citari corespunzatoare, ne rezervam dreptul de a solicita aplicarea legii.

Pentru utilizare în aplicatii comerciale, aveti nevoie de un acord scris, contractual, din partea noastra, a autorilor. Contactati-ne la adresa de mai jos, cu o descriere a aplicatiei si a utilizarii pe care doriti sa o dati Dvs. În lipsa unei citari corespunzatoare, ne rezervam dreptul de a solicita aplicarea legii.

Copyright © 2006 Marius Zbancioc si H.N. Teodorescu. Pentru acord scris, contactati-ne la adresa zmarius@etc.tuiasi.ro, sau la adresa hnteo2004@yahoo.com.

Autori: H.N. Teodorescu si M.Zbancioc