

4. Detectie Pitch (F0) prin metoda autocorelatiei (Documentatie V.01)

4.1. 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.

Spre deosebire de metoda ceptrala care lucreaza în domeniul frecventelor, metoda autocorelatiei lucreaza în domeniul timp (direct cu valorile esantioanelor din semnalul vocal). Metoda este des întâlnita în literatura de specialitate, fiind considerata o metoda robusta, care conduce la o buna detectie a frecventei fundamentale.

Funcția de corelatie aplicata pe doua semnale x si y ofera informatii legate de similitudinile dintre acestea. Aceasta metoda de comparatie este utilizata pentru detectia anumitor regularitati (legate de periodicitatea semnalelor, în cazul în care acestea pot fi considerate periodice sau cvasi-periodice). Formula de calcul în continuu este:

$$R_{XY}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) \cdot y(t + \tau) dt; \quad R_{YX}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y(t) \cdot x(t + \tau) dt$$

În discret determinarea coeficientilor de corelatie revine la calculul unor sume:

$$R_{XY}[n] = \sum_{k=0}^{N-n} x[k] \cdot y[k+n]; \quad R_{YX}[n] = \sum_{k=0}^{N-n} y[k] \cdot x[k+n], \quad n = \overline{0, N}$$

Autocorelatia reprezinta corelatia unui semnal cu el însusi (y este înlocuit cu x în formule). Valorile de maxim local prezente pe semnalul de autocorelatie determinat dupa formula de calcul de mai sus, sunt în strânsa legatura cu proprietatile de periodicitate ale semnalului. Valoarea maxima gasita în $R_{XY}[0]$ nu trebuie luata în considerare, reprezinta energia/puterea semnalului într-o fereastră de analiza de dimensiune fixata N . Urmatorul maxim local este asociat cu perioada fundamentala si cautarea lui poate fi limitata într-un interval de valori dat de banda de cautare a frecventei fundamentale [80Hz – 500Hz]. Astfel pentru o frecventa de esantionare de $F_s=16000\text{Hz}$ se va cauta în semnalul furnizat de autocorelatie între valorile:

$$R_{xx}[F_s/f_{\max}, \dots, F_s/f_{\min}], \text{ adica } R_{xx}[16000/500, \dots, 16000/80] \quad R_{xx}[32, \dots, 200]$$

Se recomanda ca fereastră de analiza sa fie de minim 4-5 ori perioada fundamentala ($5 \cdot F_s/f_{\min}$) maxima pentru o determinare buna a pitch-ului. În cazul în care frecventa de esantionare este de 16000Hz, fereastră de analiza ar trebui sa aiba minim 800-1000 de esantioane. Tinând cont de faptul ca se doreste de cele mai multe ori realizarea unor comparatii între rezultatele furnizate de mai multe metode de detectie pitch, trebuie sa avem aceiasi parametri la rulare (dimensiune fereastră de analiza, pas de deplasare). Pentru metoda ceptrala avem restrictia de a rula programul cu o dimensiune a ferestrei de analiza de minim 1024 esantioane (putere a lui 2) si putem accepta aceasta valoare ca parametru de intrare pentru N .

Deoarece se doreste calcularea pitch-ului numai pe segmentele vocalice se verifica daca pentru o fereastră de analiza energia semnalului în timp depaseste de 10 ori energia minima sau a 20 parte din energia maxima (E_{\max} , E_{\min} sunt calculate pentru toate

ferestrele de analiza de dimensiune N), sau daca energia spectrala în banda [80Hz 2500 Hz] depaseste 70% din energia spectrala.

4.2. Mod de lucru

Instrumentul este realizat sub forma unui executabil, numit "FO_corel.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 autocorelatiei". 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