

ALGORITME DE CODARE A VOCII ÎN SISTEMELE DE TELECOMUNICAȚII

Autori: Nicolai BOGACIUC, Ion AVRAM

Universitatea Tehnică a Moldovei

Abstract: *This document contains a review of the modern algorithms for speech coding in contemporary telecommunication systems. In the analyzing algorithms is indicated his advantages and disadvantages. It highlights the most effective algorithms.*

Cuvinte cheie: *algoritme de codare a vocii, sisteme de telecomunicații, codor, decodor, bandă de frecvențe.*

1. Introducere

Un codor de voce, de obicei este specificat ca un algoritm, care este definit ca o procedură de calcul, care are ca operanzi de intrare niște valori, iar la ieșire conține niște valori diferite. Un algoritm poate fi implementat ca fiind un program (o instrucțiune de procesor) sau ca un dispozitiv (are loc execuția directă prin circuite digitale). Datorită gradului înalt de accesibilitate și prețului redus astăzi multe sarcini de procesare a semnalelor sunt executate în format digital. El își are avantajele sale: programabilitate, fiabilitate și abilitate de a face față unor operații complicate care contribuie la codificarea vocii, pe care nici un aparat analogic nu le poate executa.

Orice algoritm pe larg folosit funcționează în baza unui standard, scris în formă de text. Precum că se folosesc tehnologiile digitale, algoritmele se nasc dintr-un cod, elaborat foarte minuțios într-un anumit limbaj de programare, spre exemplu în C++. În general este nevoie de îndeplinit 2 funcții pentru a coda vocea: codarea și decodarea acesteia. Codul de referință este foarte general, pentru ca algoritmul să lucreze este nevoie ca acesta să fie implementat într-un anumit dispozitiv. Pe baza domeniului de activitate și al echipamentului hardware, codul suferă niște schimbări pentru a îndeplini condițiile cerute. Acesta este deja un proces ingineresc și presupune optimizarea codului pentru a mări viteza efectuării calculelor și a micșora cerințele de memorie/consum de putere.

Exemplu: *Supercalculatoarele au destule resurse de operare, deci, procesul modificării codului este mult mai ușor. Stațiile mobile (cum ar fi un smartphone), deja sunt limitate în resurse și este necesar de a adapta codul pentru un raport calitate/consum rezonabil de energie.*

Pentru a înțelege mai bine algoritmele descrise în articolul dat, este nevoie de însușit următoarele noțiuni:

- a. Limitele sistemului auditiv la om – reprezintă banda de frecvențe pe care le poate recepționa urechea omului, el variază de la 20 Hz până la 20 kHz;
- b. Codec – un program sau o bibliotecă de software, eventual chiar un aparat hardware, care asigură codarea și decodarea unei informații (în cazul dat semnalelor voce);
- c. Banda vocală – este banda folosită în telefonie pentru reproducerea vocii umane, în limitele căreia poate fi efectiv transmisă și deasemenea înțeleasă, ea se conține între frecvențele 300 Hz – 3400 Hz.

2. Codarea vocală

Codurile potrivite pentru domeniul audio-vocal, exploatează limitele sistemului auditiv uman pentru a oferi o compresie cât mai transparentă pe baza unei analize a semnalului. Deasemenea, codorul trebuie să codeze cu fidelitate toate particularitățile vocii umane. Toată informația dată trebuie să fie transmisă în ambele direcții dacă suntem în cazul unui sistem radio mobil. Pentru ca acesta să nu transmită semnale nedorite, există modelele vocale care sunt construite pe bază de gen, vârstă, atribute fizice, diferite limbi și accente regionale, care servesc drept reper pentru codec. Deasemenea el trebuie să fie destul de performant ca să facă față unui posibil zgomot al mediului ambiant în care se află vorbitorul.

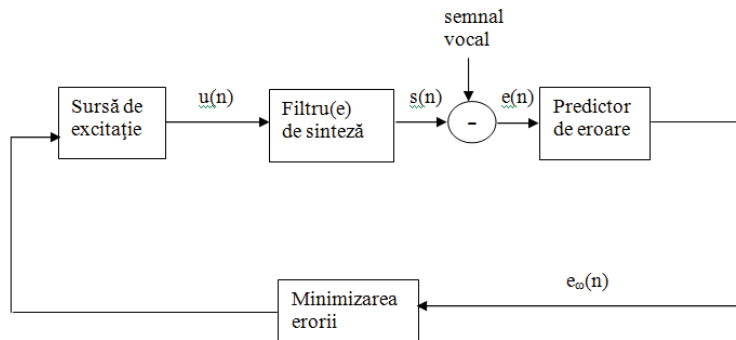
3. Codul de Predicție Liniară

Acest tip de codare a vocii este bazat pe un model foarte simplu de producere a semnalului vocal. Acest model tinde să reproducă vocea umană din observații asupra aspectelor fizice ale acesteia. Structura de bază pentru modelul general de coduri cu analiza prin sinteză este indicat în figura 1.

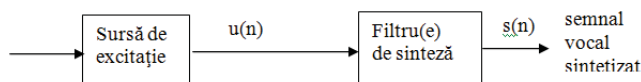
Modelul este compus din trei părți principale:

Sinteza filtrelor

Filtre cu zerouri în toată banda, care modelează anvelopa semnalului pentru forma de undă a semnalului vocal. Este numit filtru cu corelare pe termen scurt, datorită eficienței sale dată de calculul cu predicție pentru eșantionare (8-16 eșantioane, ceea ce înseamnă termen scurt). Sinteza de filtre poate de asemeni include și filtre cu corelație pe termen lung, care se obțin prin cascada filtrelor pe termen scurt. Predictoarele pe termen lung modelează structuri apropiate de spectrul vocal.



(a) Codor



(a) Decodor

Figura 1 Modelul general de analiză prin sinteza LPC

Generarea excitației

Aceasta produce secvența de excitație care va determina filtrul de sinteză să realizeze reconstituirea semnalului vocal la receptor. Excitația este optimizată prin minimizarea importanței erorii dintre semnalul original și cel sintetizat. După cum se indică în Figura 1 în codor se află un decodor local, iar metoda de analiză pentru optimizarea excitației folosește diferența dintre semnalul original și cel sintetizat ca un criteriu de eroare și alege secvența de excitație care minimizează rolul erorii. Eficiența acestei metode de analiză provine din procesul de optimizare prin bucla închisă. Acest proces de optimizare permite reprezentarea predicției restante folosind o viteză de bit mică, în timp ce se păstrează o calitate înaltă a semnalului vocal. Aceasta explică superioritatea codului cu analiză prin sinteză cu predicție asupra codurilor cu predicție, care au o structură de inel deschis, cum ar fi codurile RELP cu predicție liniară cu excitație liniară. Punctul cheie în structura de inel închis este acela că restul de predicție este cuantizat prin minimizarea importanței erorii dintre semnalul original și cel refăcut, decât minimizarea erorii dintre semnalul reziduu și forma sa cuantizată în structura de buclă deschisă.

Criteriul de minimizare a erorilor

Cel mai utilizat criteriu de minimizare a erorilor este eroarea medie pătratică. Eroarea este trecută printr-un filtru de pondere perceptuală (sesizează înălțimea semnalului), care taie spectrul zgomotului pentru a concentra puterea la frecvențele din spectrul vorbirii astfel încât zgomotul este mascat de semnalul vocal.

4. Codorul LPC FS1015

Semnalul vocal eșantionat la 8 kHz este cuantizat într-un PCM uniform, cu 12 biți/eșantion.

Analiza cu predicție liniară

Codorul FS1015 încorporează o sincronizare pas cu pas, prin care 130 eșantioane sunt scoase din cardu, cu poziția acestuia ajustată după un punct de referință de o formă de undă, ca un început de ciclu tonal. Prin aceasta la vocea sintetizată se adaugă omogenitate.

Detectarea vocii

Un clasificator linear este folosit pentru a determina vocea. Vectorii șablonului dat sunt caracterizați de:

- Energie de bandă joasă;
- Rată de la vârf la bază a funcției diferențiale a magnitudinii. Acest parametru indică periodicitatea unui cadru. Pentru cadrele vocale, rata este în general mult mai înaltă decât pentru cadrele "fără voce";
- Rata de trecere prin zero.

Situația în care un cadru voce este localizat între două cadre "fără voce" poate cauza artefacte nedorite la sinteza semnalului final. Pentru a preveni acest fapt, decizia cadrelor învecinate se ia în faza finală de decizie.

Alocarea biților

Schema alocării biților pentru codorul FS1015. Perioada de 60 de pași valoare a tonului vocal este codată folosind 7 biți. Puterea este codată folosind 5 biți. Sincronizarea este o alternanță dintre 0/1. Protecția împotriva erorilor este doar pentru cadrele "fără voce", unde se transmit 21 de biți. Schema utilizează 54 de biți pe cadru, cu lungimea unui cadru de 22,5 ms, având o rată de 2400 biți pe secundă.

Tabelul 1 Alocarea biților pentru codorul FS1015

| Parametrul | Rezoluția | |
|--------------------------|-----------|-----------|
| | Voce | Fără Voce |
| Perioada pasului / tonal | 7 | 7 |
| Puterea | 5 | 5 |
| LPC | 41 | 20 |
| Sincronizarea | 1 | 1 |
| Protecția de erori | - | 21 |
| Total | 54 | 54 |

5. Limitele modelului LPC

Modelul dat, deși este destul de practic din cauza ratei mici de biți, este lipsit de acuratețe în mai multe circumstanțe, creând artefacte nedorite în vocea sintetizată. Limitele date sunt analizate, pentru ca următoarea generație de codori de voce să fie mai eficientă.

Dezavantajul 1:

În multe cazuri, cadrul nu poate fi determinat cu claritate ca "voce" sau "fără voce". Astfel apar sunetele nedorite și zgomotele tonale, pentru că unele cadre nu pot fi cu acuratețe sortate: "fără voce" tranzitoriu spre "voce" sau "voce" tranziție spre "fără voce".

Dezavantajul 2:

Folosind un zgomot aleator sau un impuls periodic ca exitant, nu se poate compara cu observațiile folosind semnale vocale reale.

Dezavantajul 3:

Nici o informație despre faza semnalului original nu este păstrată. De obicei urechea umană nu este prea capricioasă la faza semnalelor auzite, însă aceasta poate construi un semnal sintetizat foarte calitativ adăugându-i naturalețe.

Concluzie

În general modelul LPC era cândva activ folosit în domeniul telecomunicațiilor, din cauza ratei joase de biți pe secundă. În zilele noastre, la necesitățile consumatorilor acesta deja nu poate fi activ implementat, ci doar în unele cazuri. Un domeniu în care se va folosi acest tip de codare vocală este comunicațiile militare, unde este nevoie de viteză de transmisiune mult mai mare.

Bibliografie

Wai Chu, Speech coding algorithms. Foundation and evolution of standardized coders, Copyright, 2003, John Wiley & Sons Inc., 585 pp.