



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale
2007-2013



Platformă de e-learning și curriculum e-content pentru învățământul superior tehnic

Transmisia datelor multimedia in rețele de calculatoare

27. Criterii de performanta in codarea vorbirii

Criteria de performanta in codarea vorbirii

- Pentru evaluarea unui codor particular trebuie considerate o serie de criterii, cel mai importante fiind:
 - *calitatea vorbirii*
 - Codoarele vorbirii urmaresc producerea unei distorsiuni audibile minime la o rata data
 - Naturalitatea si inteligibilitatea sunetului produs sunt criterii – de asemena – importante
 - Calitatea vorbirii poate fi determinata prin teste de ascultare ce calculeaza opinia media a unor ascultatori
 - Calitatea vorbirii poate fi apreciata si prin masuri cantitative cum sunt castigul predictiei, distorsiunea spectrala

Criteria de performanta in codarea vorbirii

- *Rata de bit*
 - Inseamna numarul de biti pe secunda ce trebuie transmisi de codor
 - Obiectivul algoritmului de codare este de a reduce rata de bit prin mentinerea unei calitati inalte a vorbirii
- *Necesarul de memorie pentru implementare; toti algoritmi de codare a vorbirii sunt executati in circuite integrate DSP dedicate sau de uz general*
 - Acestea au capacitate limitata de memorie RAM precum si o viteaza finita
 - Rezulta ca algoritmi de codare-decodare nu trebuie sa fie prea complecsi, deci sa nu depaseaca capacitatea circuitelor DSP

Criteria de performanta in codarea vorbirii

- Adesea, algoritmi de codare prelucreaza impreuna grupuri de esantioane
 - Daca numarul de esantioane este prea mare, se introduce o *intarziere* intre vorbirea originala si cea codata
 - Aceasta nu este de dorit in transmisiile de timp real, dar poate fi tolerata in cazul memorarii vorbirii sau a inregistrarii acesteia
- *Banda de frecventa* a semnalului vorbire ce trebuie codat este o alta problema
 - In transmisiunile telefonice se utilizeaza o banda de 200-3400 Hz
 - Exista insa transmisii audio, cum sunt cele din teleconferinte sau tele-invatare, ce necesita benzi de 7-20 KHz

Criteria de performanta in codarea vorbirii

- Algoritmii de codare trebuie sa fie *robusti la erorile* ce pot apare pe canal
 - Erorile de pe canal sunt cauzate de zgomotul de pe canal, interferenta inter-simbol, fading etc.
- In timpul transmisiilor reale, semnalele vorbire sunt afectate de diferite zgomote acustice ce provin din mediu cum sunt:
 - zgomotul strazii
 - zgomotul masinii
 - al biroului, etc.
- Algoritmii de codare trebuie sa mentina o buna *calitate* a vorbirii in prezenta unor astfel de surse de zgomot

Aprecierea distorsiunilor

- **Inteligibilitate si Intelegere**

- Inteligibilitatea se refera uzual la abilitatea de a identifica cuvinte izolate
- Intelegerea se refera la intelegerea continutului unui material
 - Aceasta este uzual testata prin intrebari despre materialul inregistrat

Aprecierea distorsiunilor

- **Perceptia si redundanta vorbirii**

- Importanta analizei modului in care se realizeaza perceptia consta in faptul ca aprecierea calitatii semnalului reconstituit din parametrii transmisi pe canal, se face de catre urechea umana
 - De aceea, trebuie explorat modul in care functioneaza urechea umana, prin evidentierea aspectelor importante si a celor mai putin importante.
- Sistemul auditiv uman actioneaza ca un set de filtre si este foarte sensibil in domeniul 200-5600 Hz in ceea ce priveste perceptia.
 - Trasaturile importante ale perceptiei, cum este vocea, sunt determinate din structura armonica care este prezenta in diomeniul frecventelor joase (ce nu depaseste 3 KHz).
- Intrucat sursa si – mai ales – destinatarul semnalului vorbire este un operator uman, pe langa masurile clasice de apreciere a distorsiunilor apar si masuri de natura subiectiva

Masuri de apreciere subiectiva a distorsiunilor

- Testele subiective permit o evaluare comparativa a diverselor codoare
- In astfel de teste, calitatea vorbirii este apreciata uzual prin inteligibilitate, tipic definita ca procentul de cuvinte sau foneme care sunt corect percepute
 - Aspectele semnificative ale perceptiei vorbirii sunt *inteligibilitatea si naturalitatea*
 - Pentru aprecierea acestor calitati se poate folosi masura *Scorul opiniei mediii (MOS = Mean Opinion Score)*
 - Determinarea acesteia reclama un proces lent
 - Decizia este impartita in 5 domenii ca in tabelul de mai jos

Score	Calitatea vorbirii	Nivelul distorsiunii
5	Excelenta	Imperceptibila
4	Buna	Abia perceptibila dar nu deranjanta
3	Acceptabila	Perceptibila si slab deranjanta
2	Saraca	Deranjabila dar nu intolerabila
1	Nesatisfacatoare	Foarte deranjanta si intolerabila

Masuri de apreciere subiectiva a distorsiunilor

- Opinia sau nivelul perceput al distorsiunii este transformat in termeni descriptivi:
 - excelent
 - foarte bun
 - bun
 - sarac
 - nesatisfacator
- Aceste aprecieri se refera la tot ansamblul de distorsiuni, permit compararea directa cu masurile obiective (cantitative), dar nu ajuta imediat la intelegerea cauzelor distorsiunilor
- Testele subiective necesita o multime numeroasa de ascultatori, ce necesita o instruire adecvata in ascultare si calibrare
 - Este nevoie si de un mediu special pentru testare

Masuri de apreciere obiectiva a distorsiunilor

- **In domeniul timp:**

- raportul semnal-zgomot: daca $s(n)$ este semnalul original al vorbirii, este vorbirea codata si fisierul vorbire are N_T esantioane, atunci raportul semnal-zgomot este

$$SNR[\text{dB}] = 10 \lg \frac{\sum_{n=0}^{N_T-1} s^2(n)}{\sum_{n=0}^{N_T-1} (s(n) - \hat{s}(n))^2}$$

- Este o masura globala, definita pe intreg fisierul

Masuri de apreciere obiectiva a distorsiunilor

- raportul semnal-zgomot segmental (pe segment) considera raportul semnal –zgomot calculat pe un segment si face media geometrica a valorilor obtinute.
 - Daca semnalul vorbire are N_F cadre si lungimea fiecarui cadru este N_S , raportul semnal-zgomot segmental este

$$SEGSNR[\text{dB}] = \frac{1}{N_F} \sum_{i=10}^{N_F-1} 10 \lg \frac{\sum_{n=0}^{N_S-1} s^2(N_S i + j)}{\sum_{n=0}^{N_S-1} (s(N_S i + j) - \hat{s}(N_S i + j))^2}$$

- In general, este o masura mai buna decat raportul semnal zgomot global

Masuri de apreciere obiectiva a distorsiunilor

▫ Castigul predictiei

$$P_g [\text{dB}] = 10 \lg \frac{\sum_{n=0}^{N_T-1} s^2(n)}{\sum_{n=0}^{N_T-1} r^2(n)}$$

unde N_T este numarul de esantioane

- Un castig mare al predictiei presupune ca filtrarea LP este aproape de aceea a tractului vocal, astfel incat semnalul eroare $r(n)$ va fi mai aproape de excitatia adevarata
- Exemplu

Fisierul de intrare	P_g [dB]
Fisier 1, vorbitor feminin, 23808 esantioane	16.12
Fisier 2, vorbitor masculin, 30976 esantioane	17.35
Fisier 3, vorbitor masculin, 28416 esantioane	16.01

Masuri de apreciere obiectiva a distorsiunilor

▫ eroarea energiei

$$E_{err} = \sqrt{\frac{\sum_{n=0}^{N_T-1} e^2(n)}{\sum_{n=0}^{N_T-1} s^2(n)}} = \sqrt{\frac{\sum_{n=0}^{N_T-1} (s(n) - \hat{s}(n))^2(n)}{\sum_{n=0}^{N_T-1} s^2(n)}}$$

- Intrucat se folosesc aceiasi coeficienti pentru filtrare si filtrare inversa, in mod teoretic semnalele vorbire de intrare si de iesire ar trebui sa fie identice, deci senalul eroare sa fie zero
- Experimental, ne asteptam ca semnalul eroare sa fie foarte mica din cauza reprezentarii finite a datelor si a limitarilor impuse similarilor

Masuri de apreciere obiectiva a distorsiunilor

- **In domeniul frecventa:**

- distorsiunea spectrala se defineste ca diferenta patratica medie dintre spectrul de putere original si spectrul de putere obtinut prin cuantizarea coeficientilor LPC

- Pentru cadrul i distorsiunea este

$$SD_i = \sqrt{\frac{1}{F_s} \int_0^{F_s} \left(10 \lg \frac{S_i(f)}{\hat{S}_i(f)} \right)^2 df} \quad [\text{dB}]$$

- unde F_s este frecventa de esantionare, $S_i(f)$ si sunt spectrele originale si estimate

Masuri de apreciere obiectiva a distorsiunilor

- $S_i(f)$ si sunt spectrele originale si estimate sunt definite prin

$$S_i(f) = \frac{1}{A_i(e^{j2\pi f / F_s})} \qquad \hat{S}_i(f) = \frac{1}{\hat{A}_i(e^{j2\pi f / F_s})}$$

- $A_i(z)$ si sunt LPC original si- respectiv- cuantizati
- In locul integrarii se poate poate folosi sumarea (de fapt, DFT) dupa relatia

$$SD_i = \sqrt{\frac{1}{n_1 - n_0} \sum_{n=n_0}^{n_1-1} \left(10 \lg \frac{S_i(e^{j2\pi n})}{\hat{S}_i(e^{j2\pi n})} \right)^2} \quad [\text{dB}]$$

Masuri de apreciere obiectiva a distorsiunilor

- Alta masura este distanta euclidiana ponderata
 - Se evalueaza in raport cu reprezentarea LSF, intrucat acestia au o buna legatura cu forma spectrala, adica formantii, si cu vaile
 - Daca f si \hat{f} sunt doi vectori, original si codata LSF, atunci distanta euclidiana este

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \|\mathbf{f} - \hat{\mathbf{f}}\|^2$$

iar in cazul unei analize de ordinul p se obtine

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \sum_{i=1}^p (f_i - \hat{f}_i)^2$$

Masuri de apreciere obiectiva a distorsiunilor

- Daca w_i este ponderea asignata componentei spectrale i a LSF (care are componente necorelate), atunci distanta euclidiană ponderată este

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \sum_{i=1}^P w_i \cdot (f_i - \hat{f}_i)^2$$

- Calculul ponderilor se face după următoarele considerente:
 - ponderea formantilor este mult mai mare decât a vailor;
 - urechea umană distinge mai ușor diferențele frecvențelor joase decât cele înalte; ca atare frecvențele joase trebuie ponderate mai mult