



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

25. Compresia vorbirii

Tehnici de compresie a vorbirii in domeniul timp

- Motivarea principala pentru compresia vorbirii in timp este pentru reducerea timpului necesar unui utilizator pentru a asculta un mesaj, prin cresterea capacitatii de comunicare a urechii
 - A doua motivatie este legata de reducerea cantitatii de date in vederea minimizarii spatiului de memorare si a benzii de transmisie a mesajelor vorbite
- Compresia in timp a vorbirii poate fi utilizata intr-o varietate de aplicatii incluzand instruirea, ajutorarea celor cu deficiente auditive si pentru interfetele om-calculator
 - Studiile au aratat ca ascultarea de doua ori a materialelor redade cu o viteza dubla este mult mai eficienta decat ascultarea mesajelor la viteza normala

Tehnici de compresie a vorbirii in domeniul timp

- Compresia temporala a fost utilizata in prezentarea mesajelor in sistemele de *mail* vocal si in ajutorarea persoanelor cu deficiente majore de vedere
 - Mai mult, vorbirea poate fi incetinita pentru invatarea limbajelor, de exemplu
- Tehnicile de compresie in timp se utilizeaza si in sistemele de recunoastere a vorbirii cand formele de intrare sunt normalizate la o lungime standard
- Vorbirea comprimata in timp este referita ca accelerata, comprimata, cu scara de timp modificata, cu viteza marita, sau cu timp modificat

Vorbirea rapida

- Rata de vorbire normala in Engleza este intre 130 si 200 cuvinte pe minut (wpm)
- Cand se vorbeste repede, vorbitorul schimba ne-intentionat attributele relative ale vorbirii relativ la durata pauzelor, durata vocalelor consonante etc.
- In general, vorbitorii pot comprima vorbirea lor pana la 70% , limitarea fiind din cauza limitarilor fiziologice

Schimbarea vitezei

- Schimbarea vitezei este similara cu redarea unei casete la o viteza mai mare sau mai mica
- Metoda poate fi reprodusa digital prin schimbarea ratei de esantionare in timpul intoarcerii casetei
- Metodele nu sunt prea folosite intrucat produc un efect nedorit al modificarii frecventei proportional cu schimbarea vitezei de redare, cauzand o scadere a inteligibilitatii

Sinteza vorbirii

- Cu o voce pur sintetica este posibil sa se genereze vorbire la diverse rate ale cuvintelor
- In mod curent sintetizoarele text-vorbire pot produce rate de pana la 550 wpm
- Acest lucru se obtine prin reducerea selectiva a *fonemelor* (phoneme) si a duratei pauzelor
- Tehnica este puternica, in particular pentru ajutorarea persoanelor cu dizabilitati, dar nu este relevanta in raport vorbirea inregistrata

Eliminarea pauzelor (*Silence Removal*)

- Cea mai simpla metoda consta in utilizarea energiei sau mediei amplitudinii, combinata cu pragurile de timp
- Alte metode utilizeaza masuratori de trecere prin zero sau parametrii LPC
 - De exemplu tehnica TASI (*Time Assigned Speech Interpolation*) este utilizata pentru a dubla capacitatea cablurilor telefonice transoceanice
 - Tehnica DSI (*Digital Speech Interpolation*) este similara dar se lucreaza in totalitate in domeniul digital
 - DSI/TASI doresc sa conserve banda sistemului

Esantionare neuniforma

- Principiul metodei este prezentat in urmatoarea figura
- Daca intreruperile se fac la intervale regulate, largi portiuni din semnal se pot sterge fara a se afecta inteligibilitatea

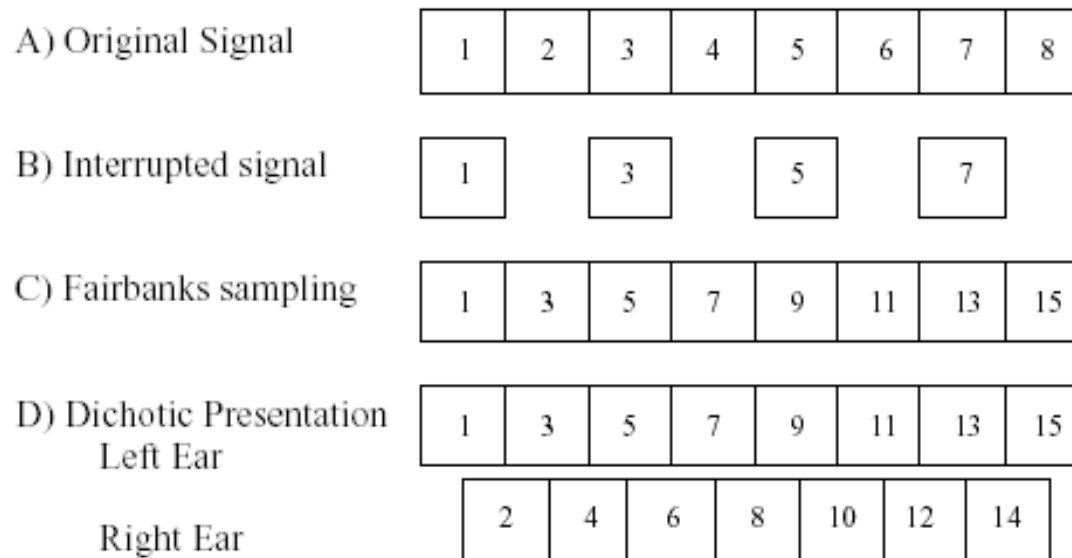
Terminologia esantionarii



sampling interval	I_s
discard interval	I_d
time compression ratio	$R_c = I_d / (I_d + I_s)$

Esantionare neuniforma

- O alta tehnica prezentata in figura urmatoare, segmentele semnalului vorbire sunt alternativ indepartate si retinute
- Durata fiecarui interval de esantionare trebuie sa fie cel putin egala cu a unui formant (*pitch*) (15 ms) dar trebuie sa fie – de asemenea – mai mica decat lungimea unui fonem



Esantionarea cu prezentare dihotonica

(*Sampling with Dichotic Presentation*)

- O varianta interesanta a metodei de esantionare consta (figura 3D) este obtinuta prin redarea semnalului esantionat standard unei urechi si a materialului „indepartat” celeilalte urechi
- Sub aceasta conditie „*dichotic*” inteligibilitatea si propunsierea cresc.
- Desi initial exista o slaba dificultate in intelegerea mesajului, adaptarea este imediata
- Raportul de compresie creste pana la 50% prin pastrarea intregii informatii din semnalul original

Esantionarea cu prezentare dihotonica

- Cuantizarea numerica pe un numar mare de biti conserva formele de unda, in sensul ca semnalul de la iesire este apropiat de semnalul de la intrare, ca forma de unda
 - Diferenta formelor de unda, de la iesire si de la intrare constituie criteriul eroare pentru proiectarea circuitului de cuantizare
- Cand cuantizarea se realizeaza cu numar mic de biti (numar mic de simboluri binare), in proiectarea codorului trebuie sa se includa informatie despre producerea si perceptia semnalelor audio, astfel incat semnalul de iesire poate sa difere mult de semnalul de intrare, insa efectul acestora asupra urechii umane sa fie acelasi

Esantionarea cu prezentare dihotonica

- In principiu, un algoritm de codare eficienta a sursei va:
 - indeparta componentele redundante ale sursei de semnal prin exploatarea corelatiilor dintre esantioane. (*codare entropica*)
 - Indeparteaza componentele care nu sunt percepute de urechea umana
 - Irelevanta se manifesta ca o amplitudine ne-necesara sau ca o rezolutie in frecventa
 - Portiunile de semnal care sunt mascate nu trebuie transmise. (*codare perceptuala*)