

Prepoznavanje Govora

Kompresija audio signala

Kompresija glasa

- U praksi se koriste tri opšte kategorije algoritama za kompresiju govornog signala:
 - Kodiranje bazirano zvučnom talasu (waveform based): DPCM, μ -law, a-law, ADPCM
 - Parametarsko kodiranje (parametric compression coding): LPC
 - Hibritni pristup: CELP

DPCM

- **Differential Pulse Code Modulation**
- Zapisuje se razlika između susednih semplova. Dodatno, moguće je kodovati razliku entropijski.
 - Delta modulacija – posebna vrsta DPCM
 - Perceptualno kodiranje – a-law ili μ -law algoritmi koriste osobine slušnog aparata za kodiranje

DPCM - entropijsko kodovanje

- Kod algoritama entropijskog kodovanja, koriste se reči promenljivih dužina, koje se konstruišu tako da reči koje će se češće javljati imaju kraći kod.
- Kod DPCM sa kodovanjem razlika između susednih signala, možemo da primenimo entropijsko kodovanje koje favorizuje male brojeve.

DPCM - delta modulacija

- Kod delta modulacije, izlazni signal je 1 bit koji naznačava da li je signal u prethodnom semplu rastao ili padao.
- Osnovni problem je što je neophodna velika frekvencija semplovanja da bi DM imao smisla.
- Prednosti su:
 - Jednostavna implementacija
 - Radi dovoljno dobro da može da se primenjuje kod prenosa telefonskog signala

ADPCM

- Adaptive DPCM.
- Radi slično kao DPCM, samo se prilagođava promeni signala. Ono što se prilagođava je širina izlazne reči.
 - Ako dođe do clipping-a (nije moguće zapisati reč), širina izlazne reči se duplira.
 - Ako je izlazni moguće zapisati duplo kraćom reči, širina se prepolovljava.

Perceptualno kodiranje

- Zasnovano na oduzimanju onog dela informacije koji nije značajan za percepciju zvuka.
- Koriste se osobine slušnog aparata.

μ -law principi

- Oslanja se na činjenicu da je čovekov slušni aparat po prirodi logaritamski.
- Zvukovi u višim frekvencijama nam se čine manje različitim od onih u nižim frekvencijama.
=> moguće je koristiti manju rezoluciju za predstavljanje viših frekvencija.

μ -law - osobine

- Tipično se koristi kod PCM, gde smanjuje širinu sa 16 bita na 8. (zapravo sa 14 na 8, gde se kod algoritama koji kao ulaz imaju 16 bita odbacuju najmanje značajna 2 bita)
- Postoji gubljenje informacije, ali ono najčešće nije značajno.

μ -law - kontinualno

- $F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)}, -1 \leq x \leq 1$
- Gde je $\mu = 255$

μ -law - digitalne implementacije

- Kodovanje

- Kompletna matematička operacija
 - Zbog skupoće operacije, retko se koristi u praksi
- Linearna aproksimacija – de facto standard
- Tabela – za kompresiju sa 16 bita (tj. sa 14), tabela je veličine 8K bajtova; ako prostor nije problem, ovo je preferiran metod.

- Dekodovanje

- Obično se radi primenom tabele, pošto mogućih vrednosti ima samo 255.

μ -law – linearna aproksimacija

- Na 16-bit kod dodajemo “bias” vrednost tako da se pobrinemo da imamo neki od sledećih oblika u levoj koloni, i onda kodiramo tako da dobijemo oblik u desnoj koloni:

z00000001abcdxxx - z000abcd

z0000001abcdxxxx - z001abcd

z000001abcdxxxxx - z010abcd

z00001abcdxxxxxx - z011abcd

z0001abcdxxxxxxx - z100abcd

z001abcdxxxxxxxx - z101abcd

z01abcdxxxxxxxxx - z110abcd

z1abcdxxxxxxxxxx - z111abcd

- Napomena: najznačajniji bit je znak koji se prosto kopira

μ -law - linearna aproksimacija - kodovanje

- Čuva se znak sempla (prvi bit 16-bit vrednosti snage signala).
- Klipujemo vrednost signala na 32635 i dodajemo fiksnu “bias” vrednost, kako bismo pojednostavili proces enkodiranja. Bias se tako bira da osigura pojavljivanje bita “1” negde u najviših 8 bitova (ne računajući znak). Tipična vrednost za bias je 132.
- Eksponent je region od najviših 8 bitova nakon znaka. Neka je k pozicija najviše jedinice u eksponentu. One se kodira kao trobitna niska p prema prethodnoj tabeli.
- Neka je mantisa region od 4 bita nakon k.
- Dobijamo kod od 8 bita na sledeći način:
 - Znak + p + mantisa. (gde je + konkatencija)

μ -law - linearna aproksimacija - dekodovanje

- Sačuvamo znak
- Neka je e sačinjeno od naredna tri bita posle znaka.
- Neka je m sačinjeno od naredna četiri bita posle e .

$$A = e + 3 \text{ (jer 16-bit, a ne 14-bit)}$$

$$B = (bias \ll e) - bias$$

$$Kod = (m \ll A) + B$$

- Primeniti znak na kod

Vežba

- Ručno izvršiti μ -law enkodiranje i dekodiranje za vrednosti: 10520, 5014, 1100.

Vežba

- Implementirati μ -law enkoder i dekoder.

Reference

- <http://www.threejacks.com/?q=node/176>
- Sun, L., Mkwawa, I.-H., Jammeh, E., Ifeachor, E. - Guide to Voice and Video over IP; Springer 2013
- https://instruct1.cit.cornell.edu/Courses/ee476/FinalProjects/s2006/wc227_rks33/wc227_rks33/index.htm