LINKÖPINGS UNIVERSITET

Lösningsförslag till tentamen i
Kompression av ljud och bild
TSBK35

19:e augusti 2022

1 Se kurslitteraturen

2

a) Den psykoakustiska modellen används för att hitta de ljud som kan tas bort eller kvantiseras hårdare utan att en mänsklig lyssnare hör någon skillnad. På så vis kan man få en lägre datatakt utan att den subjektiva kvaliteten ändras. På avkodarsidan behövs ingen psykoakustisk modell, för de ljud som inte kan uppfattas har redan tagits bort av kodaren.

b) Ofta är höger och vänster kanal väldigt lika och man kan därför utnyttja det beroende som finns mellan dem. Ett enkelt sätt att göra detta är att koda summan respektive skillnaden mellan kanalerna, istället för att koda kanalerna direkt. Skillnadskanalen kommer då att innehålla väldigt lite information som inte kräver så hög datatakt.

c) Ljudsignalen går först genom en filterbank som delar signalen i 32 lika breda frekvensband. En MDCT används sen för att ge ytterligare en finare frekvensupplösning. Komponenterna delas in i band som ungefär motsvarar kritiska band enligt Barkskalan. Bitar delas ut till de olika banden (med hjälp av den psykoakustiska modellen) och komponenterna kvantiseras och kodas med fixa trädkoder.

3 Se kurslitteraturen.

4 Se kurslitteraturen

5 Se kurslitteraturen.

6 a) Den teoretiskt lägsta takten ges av källans entropitakt. Eftersom källan är minnesfri blir entropitakten

$$H = -0.7 \cdot \log 0.7 - 0.2 \cdot \log 0.2 - 0.1 \cdot \log 0.1 \approx 1.1568$$

b) Om man låter a ligga närmast 0 blir det resulterande intervallet $[0.3087 \;\; 0.33271)$, av storlek 0.02401.

Antalet bitar i kodordet måste vara minst

$$\lceil -\log_2 0.02401 \rceil = 6$$

men eventuellt krävs 7 bitar.

Skriv de två gränserna som binära tal:

$$\begin{aligned} 0.3087 &= 0.0100111100\ldots \\ 0.33271 &= 0.0101010100\ldots \end{aligned}$$

Det minsta talet med 6 bitar i intervallet är 0.010100. Även alla andra tal som börjar på detta sätt som ligger i intervallet (dvs de är mindre än den övre gränsen.

Alltså räcker det med 6 bitar i kodordet.

Kodordet blir **010100**.

7 a) Distorsionen minimeras om beslutsgränserna läggs mitt emellan rekonstruktionspunkterna (dvs om man alltid kvantiserar till närmaste rekonstruktionspunkt):

$$b_0 = -1, \quad b_1 = -0.25, \quad b_2 = 0.25, \quad b_3 = 1$$

b) Distorsionen minimeras om rekonstruktionspunkterna ligger i tyngdpunkterna av sina motsvarande områden. Det ger oss

$$y_3 = \frac{\int_{b_2}^{b_3} x \cdot f_X(x)dx}{\int_{b_2}^{b_3} f_X(x)dx} = \frac{\int_{0.25}^{1} x \cdot (1-x)dx}{\int_{0.25}^{1} (1-x)dx} = \frac{27/192}{9/32} = 0.5$$

$$y_1 = -y_3 = -0.5$$

$$y_2 = 0$$

Notera att detta är samma rekonstruktionspunkter som vi hade innan, alltså är den kvantiserare vi har nu en trepunkters Lloyd-Max-kvantiserare för den givna fördelningen.

8 Antag att kvantiseringen är så fin att vi kan göra approximationen att prediktionen görs från originalsignalen. Eftersom vi kan välja typ av kvantisering själva, så väljer vi förstås likformig kvantisering följt av källkodning. Återigen kan vi använda approximationen för fin kvantisering och anta att prediktionsfelet också är gaussiskt. Då blir distorsionen

$$D \approx \sigma_d^2 \cdot \frac{\pi e}{6} \cdot 2^{-2R}$$

där σ_d^2 är prediktionsfelets varians, och $R = 5$.

$$p_n = a_1 \cdot \hat{Y}_{n-1} + a_2 \cdot \hat{Y}_{n-2} \approx a_1 \cdot Y_{n-1} + a_2 \cdot Y_{n-2}$$

Hitta a_1 och a_2 som minimerar prediktionsfelets varians σ_d^2.

$$\begin{aligned}
\sigma_d^2 &= E\{(Y_n - p_n)^2\} \approx \\
&\approx E\{(Y_n - a_1 \cdot Y_{n-1} - a_2 \cdot Y_{n-2})^2\} = \\
&= (1 + a_1^2 + a_2^2)R_{YY}(0) - 2a_1 \cdot R_{YY}(1) - 2a_2 \cdot R_{YY}(2) + 2a_1a_2 \cdot R_{YY}(1)
\end{aligned}$$

Derivera med avseende på a_1 och a_2 och sätt lika med 0, vilket ger oss lösningen

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} R_{YY}(0) & R_{YY}(1) \\ R_{YY}(1) & R_{YY}(0) \end{pmatrix}^{-1} \begin{pmatrix} R_{YY}(1) \\ R_{YY}(2) \end{pmatrix} \approx \begin{pmatrix} 1.4993 \\ -0.7987 \end{pmatrix}$$

$$\sigma_d^2 \approx 1.0023$$

Det resulterande signal-brus-förhållandet blir

$$SNR \approx 10 \cdot \log_{10} \frac{\sigma_Y^2}{\sigma_d^2 \frac{\pi e}{6} 2^{-2R}} \approx 38.1 \text{ [dB]}$$

som är bättre än de önskade 35 dB, därmed har vi löst problemet.

(Man kan även välja att lägga sig precis på 35 dB och istället ha en datatakt som är lägre än 5 bitar/sampel.)

9 Eftersom X_n är gaussisk så blir även transformkomponenterna gaussiska. Man kan beräkna transformkomponenternas varianser som diagonalelementen i $\mathbf{A} \cdot \mathbf{R}_X \cdot \mathbf{A}^T$, där

$$\mathbf{R}_X = 4 \cdot \begin{pmatrix} 1 & 0.97 & 0.9409 \\ 0.97 & 1 & 0.97 \\ 0.9409 & 0.97 & 1 \end{pmatrix}$$

Varianserna blir

$$\begin{aligned}
\sigma_0^2 &\approx 10.7036 \\
\sigma_1^2 &\approx 1.0930 \\
\sigma_2^2 &\approx 0.2034
\end{aligned}$$

(Kontrollera att medelvärdet av varianserna är lika med signalens varians: $\frac{10.7036+1.0930+0.2034}{3} = 4 = R_{xx}(0) = \sigma_X^2$.)

Bittilldelningen som minimerar distorsionen ges av $R_0 = 4, R_1 = 2, R_2 = 0$ som ger distorsionen

$$D \approx \frac{0.009497 \cdot \sigma_0^2 + 0.1175 \cdot \sigma_1^2 + \sigma_2^2}{3} \approx 0.1445$$

Det resulterande signal-brus-förhållandet blir

$$\mathrm{SNR} \approx 10 \cdot \log_{10} \frac{4}{0.1445} \approx 14.4 \text{ [dB]}$$