

Semestrální zkouška ZRE, řádný termín, 18.5.2009, skupina A

Login:

Podpis:

-
1. Nenormalizované autokorelační koeficienty mají hodnoty: $R[0] = 5$, $R[1] = 4$ a $R[2] = 3$. Určete koeficienty prediktoru 2. řádu a_1 a a_2 .

A: $a_1 = 0.889$, $a_2 = 0.111$

B: $a_1 = -0.889$, $a_2 = 0.111$

C: $a_1 = 0.889$, $a_2 = -0.111$

D: $a_1 = -0.889$, $a_2 = -0.111$

-
2. Pro výpočet koeficientů odrazu (PARCOR) k_i je potřeba:

A: FFT spektrum rámce.

B: FFT spektrum rámce, krátkodobá energie a počet průchodů nulou

C: LPC koeficienty a_i a autokorelační koeficienty $R[k]$

D: pouze LPC koeficienty a_i

-
3. V případě běžného počtu $P = 10$ LPC koeficientů, bude možné z LPC-cepstra určit frekvenci základního tónu ?

A: Ano, technika bude spolehlivější než odhad pomocí autokorelační funkce ACF.

B: Ano, technika bude zhruba stejně spolehlivá než odhad pomocí autokorelační funkce ACF.

C: Ano, ale technika bude mnohem méně spolehlivá než odhad pomocí autokorelační funkce ACF.

D: Ne, LPC-cepstrum tuto informaci neobsahuje.

-
4. Velikost lagu L (perioda základního tónu vyjádřená ve vzorcích) se pro řeč se vzorkovací frekvencí $F_s = 44100$ Hz může pohybovat v rozsahu

A: 110 — 882

B: 50 — 400

C: 20 — 160

D: 0 — 4000

-
5. Proč se u metody AMDF pro určení lagu nehledá maximum, ale minimum funkce

$$R_D(m) = \sum_{n=0}^{N-1-m} |s[n] - s[n+m]| \quad ?$$

A: Protože posunutí mezi signály $s[n]$ a $s[n+m]$ jsou podstatně menší než u autokorelační metody.

B: Protože funkce $R_D(m)$ nemá žádná (ani lokální) maxima.

C: Protože u funkce $R_D(m)$ odpovídají maxima sudým násobkům správného lagu.

D: Protože při posunutí signálu o správný lag jsou rozdíly mezi vzorky $s[n]$ a $s[n+m]$ minimální.

6. Proč je při určování lagu metodou NCCF nutná normalizace energie originálního a posunutého rámce?

A: Abychom zabránili nesmyslným hodnotám NCCF koeficientů způsobeným rozdílnou dynamikou původního a posunutého rámce.

B: Abychom normalizovali různé průměrné hodnoty lagu u mužských a ženských mluvčích.

C: Abychom normalizovali různé délky hlasového traktu (vocal tract length) u mužských a ženských mluvčích.

D: Abychom vyloučili vliv formantů na odhad lagu.

7. Při převzorkování jsou hodnoty původních vzorků $x[5] = 5$ a $x[6] = 5$. Jaké mohou být hodnoty vzorků, které se po převzorkování objeví mezi nimi ?

A: všechny menší než 5

B: všechny větší než 5

C: pokud jsou $x[5] = 5$ a $x[6] = 5$, budou mít i všechny vzorky mezi nimi hodnotu 5.

D: menší nebo větší než 5, záleží to i na hodnotách vzorků $x[n]$, kde $n < 5$ a $n > 6$.

8. Zázrakem se stalo, že po dekódování byl získán signál zcela totožný s kódovaným: $\hat{s}[n] = s[n]$. Jaký je v tomto případě poměr signálu k šumu ?

A: $SNR = -\infty$ dB

B: $SNR = -96$ dB

C: $SNR = 96$ dB

D: $SNR = \infty$ dB

9. Co je kódováno v diferenční pulsní kódové modulaci (DPCM) ?

A: Rozdíl mezi dvěma minulými hodnotami vzorků.

B: Rozdíl mezi současným vzorkem v současném rámci a současným vzorkem v minulém rámci.

C: Rozdíl mezi současným vzorkem a odhadem současného vzorku získaným lineární kombinací několika předcházejících vzorků.

D: Rozdíl mezi minulým vzorkem a odhadem minulého vzorku získaným lineární kombinací současného vzorku a několika budoucích vzorků.

10. Kódová kniha VQ má dva kódové vektory $\mathbf{y}_1 = [-1, 2]^T$ a $\mathbf{y}_2 = [3, 2]^T$. Určete totální vzdálenost (normalizovanou na jeden vektor) při kódování 4 vektorů:

$$\mathbf{x}_1 = [-1.1, 2.1]^T$$

$$\mathbf{x}_2 = [-0.9, 2.2]^T$$

$$\mathbf{x}_3 = [2, 2]^T$$

$$\mathbf{x}_4 = [2.9, 2]^T$$

Použitá vzdálenost je Euklidova.

A: 0.1270

B: 0.3663

C: 0.8147

D: 0.9058

11. U metody trénování kódové knihy Linde Buzo Gray (LBG) se dva nové vektory ze stávajícího vytvoří “štípáním” takto: $\mathbf{y}_i \rightarrow \mathbf{y}_i + \Delta, \mathbf{y}_i - \Delta$. Jak je nejhodnější zvolit směr vektoru Δ ?

A: $\Delta = [1 \ 1 \ 1 \ \dots \ 1]$

B: Jako směr největší variability dat, která náleží k danému kódovému vektoru \mathbf{y}_i

C: Jako směr kolmý na směr největší variability dat, která náleží k danému kódovému vektoru \mathbf{y}_i

D: Jako směr největší variability všech dat.

12. Velikosti vzorků signálu z krátkodobého prediktoru jsou pro $n = 0, 1, 2$ následující: $x[n] = [80 \ 100 \ 120 \ 0]$ a pro $n = 100, 101, 102$ následující: $x[n] = [40 \ 50 \ 60 \ 0]$. Jako lag byla určena hodnota $L = 100$. Jak bude vypadat přenosová funkce filtru pro výpočet chyby optimálního dlouhodobého prediktoru ?

A: $B(z) = 1 - 0.5z^{-1}$

B: $B(z) = 1 - 0.5z^{-100}$

C: $B(z) = 1 - 2z^{-1}$

D: $B(z) = 1 - 2z^{-100}$

13. Proč je v kodérech řeči CELP kódování buzení prováděno na kratších rámcích (obvykle 1/4 standardního rámce, tedy 40 vzorků?)

A: Vektorové kvantování 160-prvkových vektorů by bylo příliš náročné.

B: 40 je bližší hodnotě $2^6 = 64$

C: 40 je bližší počtu koeficientů obvykle používaných pro LPC filtr $P = 10$.

D: Nalezení adresy 40-prvkového vektoru v paměti je jednodušší než 160-prvkového

14. Jaká je maximální délka cesty K_{max} u standardní varianty DTW? R je počet vektorů v referenční matici, T je počet vektorů v testovací matici.

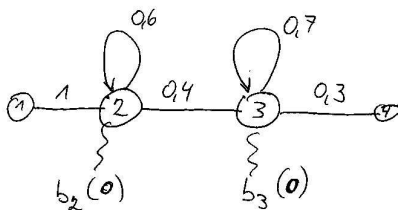
A: $K_{max} = 2(R + T)$

B: $K_{max} = 2R + T$

C: $K_{max} = R + 2T$

D: $K_{max} = R + T$

15. Model používaný na numerickém cvičení:



má pro sekvenci 4 vstupních vektorů následující hodnoty vysílacích likelihoodů:

	\mathbf{o}_1	\mathbf{o}_2	\mathbf{o}_3	\mathbf{o}_4
b_2	0.3	0.4	0.02	0.05
b_3	0.08	0.09	0.12	0.4

Určete likelihood $P(\mathbf{O}, X|M)$, pro stavovou sekvenci $X = [1 \ 2 \ 2 \ 2 \ 3 \ 4]$.

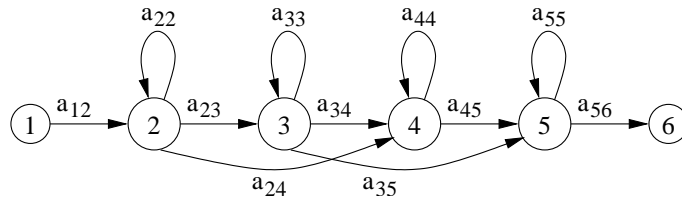
A: 2.785e-05

B: 4.147e-05

C: 5.469e-05

D: 9.575e-05

16. V modelu



probíhá algoritmus token passing. Hodnotu tokenu ve stavu j v čase t označíme $\Psi_j(t)$. V čase $t = 67$ budou ve stavu 4 soutěžit o přežití tokeny s těmito hodnotami:

- A: $\Psi_3(66) + \log a_{34} + \log b_4(\mathbf{o}_{67})$, $\Psi_4(66) + \log a_{44} + \log b_4(\mathbf{o}_{67})$
 B: $\Psi_2(66) + \log a_{24} + \log b_4(\mathbf{o}_{65})$, $\Psi_3(66) + \log a_{34} + \log b_4(\mathbf{o}_{66})$, $\Psi_4(66) + \log a_{44} + \log b_4(\mathbf{o}_{67})$
 C: $\Psi_2(66) + \log a_{24} + \log b_4(\mathbf{o}_{67})$, $\Psi_3(66) + \log a_{34} + \log b_4(\mathbf{o}_{67})$, $\Psi_4(66) + \log a_{44} + \log b_4(\mathbf{o}_{67})$
 D: $\Psi_2(64) + \log a_{24} + \log b_4(\mathbf{o}_{67})$, $\Psi_3(65) + \log a_{34} + \log b_4(\mathbf{o}_{67})$, $\Psi_4(66) + \log a_{44} + \log b_4(\mathbf{o}_{67})$

17. Používáme-li HMM modely fonémů, jsou potřeba k natrénování modelu fonému 'a' následující data:

- A: Nahrávky řeči s přepisy, které po převodu na fonémy budou obsahovat dostatek výskytů 'a'. Polohy 'a' musí být označeny.
 B: Nahrávky řeči s přepisy, které po převodu na fonémy budou obsahovat dostatek výskytů 'a'. Polohy 'a' si trénovací algoritmus najde sám.
 C: Nahrávky řeči s přepisy, které po převodu na fonémy budou obsahovat dostatek výskytů 'a'. Polohy 'a' si trénovací algoritmus najde sám, ale data nesmí následovat dva fonémy 'a' těsně za sebou.
 D: Nahrávky řeči s přepisy, které po převodu na fonémy budou obsahovat dostatek výskytů 'a'. Polohy 'a' musí být označeny včetně dělení na stavy.

18. V anglickém systému pro syntézu řeči z textu desambiguace homografů (česky "zjednoznačňování stejně napsaných slov") pro výskyty slova "read" určí:

- A: Jaká bude prozódie jednotlivých slov.
 B: Na kterém místě se bude "stříhat" pro výběr akustických jednotek
 C: Které bude přečteno syntezátorem jako "red" a které jako "ríd".
 D: Zda bude věta "klesat do tečky" (oznamovací) nebo "stoupat do otazníku" (tázací).

19. V korpusově orientovaném systému pro syntézu řeči je cena cíle C^t

- A: Zkreslení, ke kterému dojde při napojení jednotky s dalšími jednotkami.
 B: Zkreslení mezi jednotkou, kterou chceme syntetizovat, a jednotkou, která je dostupná v korpusu.
 C: Zkreslení, ke kterému dojde při modifikaci hlasitosti (gainu) při přehrávání nekvalitním reproduktorem.
 D: Zkreslení, ke kterému dojde při změně souhlásky na samohlásku.

20. Proč je při vyhledávání klíčových slov lepší použít pro krátká slova systém založený na rozpoznávací s velkým slovníkem (LVCSR) ?

- A: Díky jazykovému modelu jsou krátká slova lépe detekována a méně se pletou s fragmenty jiných slov.
 B: Akustické modely v rozpoznávací s velkým slovníkem (LVCSR) jsou vždy podstatně lepší.
 C: Rozpoznávač s velkým slovníkem (LVCSR) kombinuje likelihoody HMM s údaji o řečníkovi a je schopen efektivněji odfiltrovat falešné záchyty klíčových slov.
 D: Akustický detektor klíčových slov je obvykle přetrénován pro jednoho konkrétního řečníka.