

# Semestrální zkouška ZRE, řádný termín, 21.5.2010, skupina A

Login: .....

Podpis: .....

---

1. Na spektrogramu s délkou analyzačního okna 40 vzorků nelze rozlišit jednotlivé násobky základního tónu. Co je nutné udělat, aby to bylo možné ?

A: doplnit okno před výpočtem FFT nulami (zero padding)

B: zvětšit analyzační okno.

C: zmenšit analyzační okno.

D: zvětšit překryv (overlap) mezi analyzačními okny.

---

2. Co určuje nejvíce charakter souhlásek ?

A: místo artikulace (rty, patro, atd.)

B: síla nádechu.

C: potraviny zkonsumované za posledních 24 hodin.

D: pozice hlasivek vůči artikulačnímu ústrojí.

---

3. Je dán rámeček signálu, pro  $n = 0, 1, 2, 3$ :  $x[n] = 3 \ 4 \ 5 \ 8$ . Vypočtěte nenormovaný autokorelační koeficient  $R[2]$ :

A:  $R[2] = 24$

B:  $R[2] = 47$

C:  $R[2] = 72$

D:  $R[2] = 1$

---

4. Rámeček signálu  $x[n]$  je tvořen harmonickým signálem (cosinusovkou) se středním výkonem 1. Rámeček signálu  $y[n]$  je tvořen bílým šumem se střední hodnotou nula a středním výkonem 1. Na každém z těchto rámečků je proveden odhad koeficientů LPC filtru řádu 10 a je provedena filtrace filtrem  $A(z)$ . Výsledkem je chyba lineární predikce  $e[n]$ . Jaký bude střední výkon této chyby pro sinusovku a pro náhodný signál ?

A:  $P_{cos} = 0.99$ ,  $P_{nah} = 0.99$

B:  $P_{cos} = 0.01$ ,  $P_{nah} = 0.99$

C:  $P_{cos} = 0.01$ ,  $P_{nah} = 0.01$

D:  $P_{cos} = 0.99$ ,  $P_{nah} = 0.01$

---

5. Základní tón má frekvenci  $F_0 = 120$  Hz. Vzorkovací frekvence je  $F_s = 8000$  Hz. Při hledání základního tónu hledáme maximální autokorelační koeficient  $R[k]$  pro  $k = 20 \dots 160$ . Určete, kde bude:

A:  $k = 57$

B:  $k = 67$

C:  $k = 77$

D:  $k = 119$

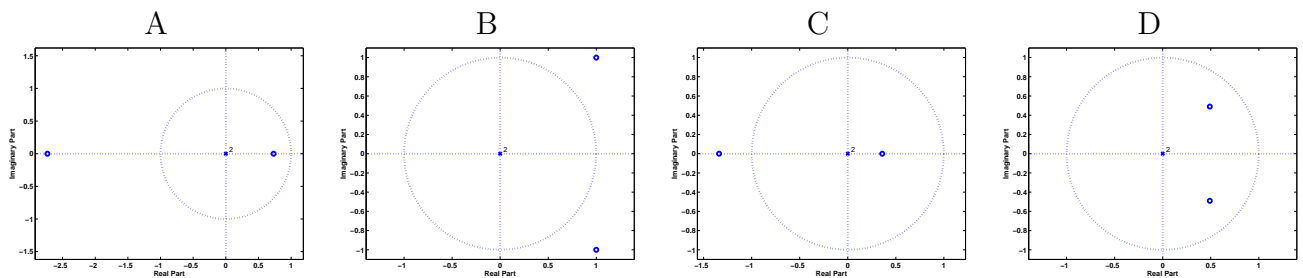
6. Autokorelační koeficienty jsou  $R[0] = 5$ ,  $R[1] = 4$ ,  $R[2] = 3$ ,  $R[3] = 2$ . Prediktor 2. řádu má koeficienty:  $a_2^{(2)} = 0.1111$ ,  $a_1^{(2)} = -0.8889$ . Spočítejte pomocí algoritmu Levinsona-Durbina hodnotu koeficientu  $a_3^{(3)}$  prediktoru 3. řádu.

- A:  $a_3^{(3)} = -0.8889$   
 B:  $a_3^{(3)} = -0.1111$   
 C:  $a_3^{(3)} = 0.1111$   
 D:  $a_3^{(3)} = 0.1250$

7. V klasickém LP dekodéru (FS 1015, ZRE laboratoře) je přepínač mezi buzením sledem impulsů a šumem. Bylo by možné jej nahradit součtem ?

- A: ano, přímo.  
 B: ano, s patřičným nastavením gainů.  
 C: ano, s úpravou  $F_0$  podle podílu šumové složky.  
 D: ne.

8. Filtr  $A(z)$  druhého řádu má dva nulové body:  $n_1 = e^{j\frac{\pi}{4}}$ ,  $n_2 = e^{-j\frac{\pi}{4}}$ . Určete, na kterém obrázku jsou nulové body filtru  $A(z/\gamma)$ , kde  $\gamma = 0.7$ .



9. Co znamená CELP (codebook excited linear prediction) ?

- A: Hodnoty parametrů LPC filtru jsou kódovány pomocí VQ s kódovou knihou, buzení určuje index do této kódové knihy.  
 B: Minulost signálu je uložena v adaptivní kódové knize, další kódová kniha obsahuje tzv. “inovace”. Součtem se budí LPC filtr.  
 C: Náhodné signály odpovídající buzení jsou uloženy v adaptivní kódové knize, další kódová kniha obsahuje tzv. “inovace”. Součtem se budí LPC filtr.  
 D: Je přenášen plný budící signál, výběr jeho vzorků pro buzení je přenášen jako index v kódové knize.

10. V CELP kódování je  $g$  gain z adaptivní kódové knihy,  $b$  je gain ze stochastické kódové knihy. Pro znělý stabilní úsek je poměr gainů  $g_z/b_z$ . Určete, jak bude vypadat hodnota pro neznělý úsek řeči.

- A:  $g_n/b_n > g_z/b_z$   
 B:  $g_n/b_n = g_z/b_z$   
 C:  $g_n/b_n < g_z/b_z$   
 D: Nelze určit, protože  $g_n = 0$ .

---

11. Co zřejmě způsobuje tonální artefakty (“zpívání”) mobilu v případě špatného signálu ?

A: je aktivován generátor tónu o základní frekvenci.

B: Dekodér se snaží nahradit rámce s chybějící informací vzorky hudby z paměti mobilu.

C: Dekodér se snaží najít v 30 sekundách v minulosti rámce s největším podílem adaptivní kódové knihy, ty použije.

D: Dekodér opakuje syntézu z parametrů posledního rámce, který má k dispozici.

---

12. Mřížka lokálních vzdáleností u DTW je následující (reference svisle, test vodorovně):

9	16
6	8
1	2

Určete průběhy indexovacích funkcí  $r(k)$  a  $t(k)$ :

A:  $r(k) = [1\ 2\ 3\ 3]$ ,  $t(k) = [1\ 1\ 1\ 2]$

B:  $r(k) = [1\ 2\ 3]$ ,  $t(k) = [1\ 2\ 2]$

C:  $r(k) = [1\ 2\ 3]$ ,  $t(k) = [1\ 1\ 2]$

D:  $r(k) = [1\ 1\ 2\ 3]$ ,  $t(k) = [1\ 2\ 2\ 2]$

---

13. Cepstral mean normalization (CMN) znamená, že před prací s HMM odečteme střední hodnoty všech parametrů, pro každou promluvu. Máme k dispozici pouze jednu trénovací promluvu a odhadujeme parametry HMM. Jak budou vypadat střední hodnoty Gaussovek ?

A: všechny budou nulové pro libovolný počet stavů.

B: střední hodnota bude nulová pouze pokud bude mít HMM pouze jeden vysílací stav, pro více stavů mohou být nenulové.

C: budou nenulové, i pokud bude mít HMM pouze 1 vysílací stav.

D: trénování na takové promluvě selže, nedostaneme uspokojivé hodnoty state occupation probabilities  $L_j(t)$ .

---

14. HMM, které má plně propojené všechny vysílací stavy, se říká ergodický HMM (EHMM). Určete, kolik různých stavových sekvencí  $X$  existuje pro EHMM se čtyřmi vysílacími stavy pro 2 vstupní vektory.

A: 16

B: 12

C: 4

D: žádná, EHMM tuto sekvenci vektorů není schopen přijmout.

---

15. State occupation probability má pro 2. stav HMM a pro časy  $t = 1 \dots 7$  hodnoty:

$$L_2(t) = [1\ 0.9\ 0.9\ 0.7\ 0.1\ 0\ 0]$$

Trénovací vektory jsou následující:  $\mathbf{o}(1) = [1, -1]^T$ ,  $\mathbf{o}(2) = [2, -2]^T$ ,  $\mathbf{o}(3) = [3, -3]^T$ ,  $\mathbf{o}(4) = [4, -4]^T$ ,  $\mathbf{o}(5) = [5, -5]^T$ ,  $\mathbf{o}(6) = [6, -6]^T$ ,  $\mathbf{o}(7) = [7, -7]^T$  Určete střední hodnotu pro 2. stav:

A:  $\boldsymbol{\mu}_2 = [1.25, -1.25]^T$

B:  $\boldsymbol{\mu}_2 = [2.44, -2.44]^T$

C:  $\boldsymbol{\mu}_2 = [7, -7]^T$

D:  $\boldsymbol{\mu}_2 = [8.80, -8.80]^T$

---

16. Proč se běžně používají HMM s diagonálními a ne plnými kovariančními maticemi ?

- A: diagonální matice lépe modelují dekorelované parametry.
  - B: diagonální matice lépe modelují obecné parametry, i pokud mají korelaci.
  - C: diagonální matice reprezentují směry s největší variabilitou dat v prostoru parametrů.
  - D: diagonální matice mají méně parametrů na odhad, zrychlí se všechny výpočty.
- 

17. Je definován levo-pravý HMM se čtyřmi stavy, z toho 2 vysílací, log. přechodové pravděpodobnosti jsou:

$$\log a_{12} = 0, \quad \log a_{22} = -0.51, \quad \log a_{23} = -0.92, \quad \log a_{33} = -0.36, \quad \log a_{34} = -1.2.$$

Tabulka logaritmů hodnot funkcí hustoty vysílacích pravděpodobností je:

t	...	46	47	48	...
$\log b_2(\mathbf{o}(t))$	...	-1	-2	-3	...
$\log b_3(\mathbf{o}(t))$	...	-4	-5	-6	...

Hodnota tokenu ve stavu 2 v čase 46 je  $\Psi_2(46) = -10$ . Určete hodnotu tokenu ve stavu 2 v čase 48.

- A:  $\Psi_2(48) = -13.02$
  - B:  $\Psi_2(48) = -14.02$
  - C:  $\Psi_2(48) = -15.02$
  - D:  $\Psi_2(48) = -16.02$
- 

18. Na co jsou při rozpoznávání spojených slov stavy s labely slov (word-end nodes) ?

- A: pokud token projde tímto stavem, je jeho log-likelihood přímo vypsána jako výstup rozpoznávače.
  - B: pokud token projde tímto stavem, je do tokenu zaznamenána identita slova, případně čas.
  - C: pokud token projde tímto stavem, je do tokenu zaznamenán čas.
  - D: tyto stavy provádí odlogaritmování log-likelihood.
- 

19. Pravděpodobnost neviděného  $n$ -gramu (např. 'a b c') se při rozpoznávání s velkým slovníkem určuje:

- A: nastavuje se na nulu.
  - B: jako součin pravděpodobností  $P(a)P(b)P(c)$
  - C: z pravděpodobnosti kratšího  $n$ -gramu, např. 'b c', nebo 'c'.
  - D: jako součin pravděpodobností  $P(a)P(c)$
- 

20. Co je principem syntézy metodou PSOLA (pitch synchronous overlap and add):

- A: signál je korelován s pilovým signálem o stejné frekvenci, jako má základní tón, výsledek je překryt a sečten s originálem.
- B: ze signálu jsou vybrány úseky podle period základního tónu, s těmi se pak dál pracuje (překrytí, součet).
- C: Základní tón je odhadován pomocí překrytí a součtu prozodických parametrů
- D: Základní tón je modifikován pomocí překrytí a součtu prozodických parametrů