

Semestrální zkouška ZRE, řádný termín, 10.5.2012, skupina A

Login: Příjmení a jméno: Podpis:
(prosím čitelně!)

1. Napište, k čemu mohou být při vývoji komerčních aplikací zpracování řeči dobré znalosti z fonetiky.
-

2. Uveďte, jaké skutečné frekvenci v Hz odpovídá normovaná kruhová frekvence $\omega = 2\pi$ rad.
-

3. Z řeči vzorkované na $F_s = 8000$ Hz je vybrán rámeček o délce 20 ms, tedy $l_{ram} = 160$ vzorků. Máte provést FFT na 256 vzorcích. Kde vezmete chybějících $256 - 160 = 96$ vzorků?
-

4. Napište příklad číslicového filtru ve zpracování řeči, kde je porušena podmínka časové invariance.
-

5. Inverzní filtr prvního řádu je dán jako $A(z) = 1 - 0.9z^{-1}$. Jeho paměť je vynulována. Na jeho vstupu je signál o délce 5-ti vzorků:

| | | | | | |
|------|----|----|----|----|----|
| n | 0 | 1 | 2 | 3 | 4 |
| x[n] | 10 | 10 | 10 | 10 | 10 |

Napište hodnotu nenormalizované energie chyby predikce E_e . Nezapomeňte, že filtr má paměť.

6. Autokorelační koeficienty jsou $R[0] = 5$, $R[1] = -1$, $R[2] = -1$, $R[3] = 1$.

Koeficienty prediktoru 2. řádu jsou: $a_1^{(2)} = \frac{1}{4}$, $a_2^{(2)} = \frac{1}{4}$. Energie chyby predikce získaná s tímto prediktorem je $E^{(2)} = 4.5$. Vypočtete koeficienty prediktoru 3. řádu:

$$a_1^{(3)} = \dots\dots\dots, a_2^{(3)} = \dots\dots\dots, a_3^{(3)} = \dots\dots\dots$$

7. Máme koeficienty filtru modelujícího hlasové ústrojí $H(z) = \frac{1}{A(z)}$. Jak získáme odhad spektrální hustoty výkonu řečového signálu ?

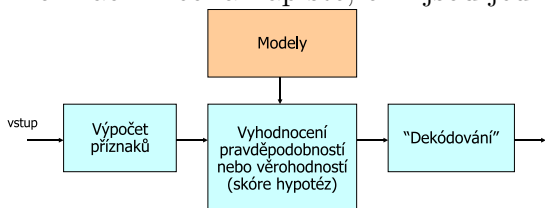
8. Častým problémem při detekci základního tónu je detekce dvojitého lagu (double lag). Krátce popište, proč k ní může dojít.

9. Napište v C nebo Matlabu krátký kód pro výpočet poměru signálu k šumu (SNR) mezi signály $x[n]$ a $e[n]$. Předpokládejte, že mají oba stejně vzorků N a jsou uloženy ve vektorech (v C polích) \mathbf{x} a \mathbf{e} .

10. Navrhněte vlastní protokol pro subjektivní posouzení kvality kódování řeči skupinou posluchačů.

11. Popište, k čemu při kódování řeči pomocí analýzy syntézou slouží perzeptuální filtr $W(z)$.

12. Na obrázku je obecné schéma klasifikátoru/rozpoznávače. Vyberte si libovolnou aplikaci dolování informací z řeči a napište, čím jsou jednotlivé obecné bloky realizovány.



13. Napište, jak spočítat vzdálenost mezi dvěma vektory parametrů: \mathbf{o} a \mathbf{r} při rozpoznávání pomocí DTW.

14. Je dána testovací sekvence vektorů \mathbf{O} o délce $T = 2$. Jsou rány dvě referenční sekvence vektorů \mathbf{R}_1 o délce $R_1 = 2$ a \mathbf{R}_2 o délce $R_2 = 3$ odpovídající slovům w_1 a w_2 . Jsou dány následující tabulky lokálních vzdáleností mezi vektory \mathbf{O} , \mathbf{R}_1 (vlevo) a \mathbf{O} , \mathbf{R}_2 (vpravo).

Proveďte výpočet DTW vzdáleností mezi referencemi a testem (nezapomeňte na normalizaci!) a určete, zda bylo rozpoznáno w_1 nebo w_2 .

| | |
|---|---|
| 1 | 2 |
| 3 | 4 |

| | |
|---|---|
| 1 | 2 |
| 3 | 4 |
| 1 | 2 |

15. 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 | ... |

Při dekódování metodou “token passing” je hodnota tokenu ve stavu 2 v čase 46: $\Psi_2(46) = -10$. Hodnota tokenu ve stavu 3 v čase 46 je $\Psi_3(46) = -11$.

Určete hodnotu přeživšího tokenu ve stavu 3 v čase 47 a určete, ze kterého stavu přišel.

16. Trénujeme 4-stavový HMM se dvěma vysílacími stavy. Odhadnuté hodnoty pro stav $j = 2$ a čas $t = 17$ jsou: částečná dopředná likelihood $\alpha_2(17) = 50$, částečná zpětná likelihood $\beta_2(17) = 40$. Celková Baum-Welchova likelihood je $\alpha_N(T + 1) = \beta_1(0) = 10000$. Jaká je state occupation probability pro stav $j = 3$ (pozor, ptám se opravdu na $j = 3$, ne $j = 2$!) a čas $t = 17$?

$$L_3(17) = \dots\dots\dots$$

-
17. Co je principem syntézy metodou PSOLA (pitch synchronous overlap and add) ?

-
18. Uveďte, k čemu při syntéze řeči z textu slouží blok **normalizace textu**.

-
19. Na co se rozpoznávání řeči používají váhované konečné stavové převodníky (weighted finite state transducers, wFST) ?

-
20. Popište dva základní přístupy k automatické identifikaci jazyka (language identification). Můžete použít velmi stručný textový popis, bloková schémata nebo obojí.
-