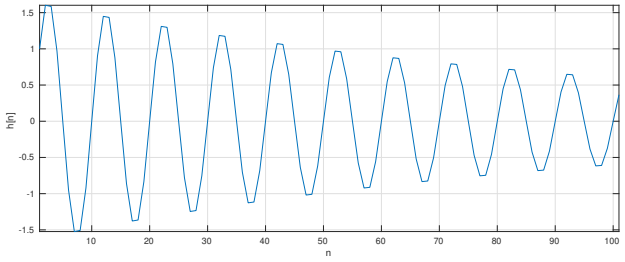
Semestrální zkouška ZRE, řádný termín, 9.6.2020, skupina CORONA

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

1. Máme řečový signál $x[n]$ o délce N vzorků. Napište slovně, matematicky, pseudokódem nebo v jakémkoliv progamovacím jazyce, jak ho převést na signál s nulovou střední hodnotou a jednotkovou směrodatnou odchylkou (mean and variance normalization).

---

2. Impulsní odezva filtru IIR druhého řádu má tvar jen velmi pomalu se zeslabující cosinusovky o periodě $N = 10$ vzorků. Nakreslete polohy pólů přenosové funkce tohoto filtru v rovině z.



---

3. Nakreslete modulové spektrum znělého úseku řeči - hlásky "a". Frekvence základního tónu je F_0 =100 Hz, frekvence dvou hlavních formantů jsou F_1 =850 Hz, F_2 =1610 Hz. Vzorkovací frekvence je F_s =8000 Hz, spektrum kreslete jen od nuly do poloviny F_s.

---

4. Proč se prediktoru LPC říká "krátkodobý" ? Kvantifikujte tu "krátkou dobu" ve vzorcích nebo v milisekundách.

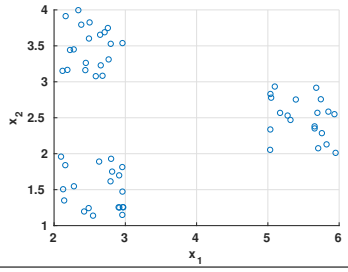
---

5. Při detekci základního tónu F_0 mají velmi nepříznivý vliv rezonance hlasového ústrojí — formanty. Popište alespoň dvě metody, kterými se jejich vliv dá potlačit nebo odstranit.

6. Co je v kódování řeči A-law / μ-law ? Nemusít psát přesné vztahy, stačí napsat, co to je a proč se to používá.

---

7. Na obrázku je 60 trénovacích vektorů $\mathbf{x}[n]$. Abyste je nemuseli počítat, v každém clusteru je jich 20. Nakreslete pozice kódových vektorů $\mathbf{y}_i$ kódové knihy vektorového kvantování (VQ) s $L = 6$ kódovými vektory. **Přibližně** určete totální vzdálenost této kódové knihy při kódování trénovacích dat: $D_{VQ} = \frac{1}{N} \sum_{n=1}^{N} d\left(\mathbf{x}[n], \mathbf{y}_i[n]\right)$, kde $\mathbf{y}_i[n]$ je nejbliší kódový vektor k trénovacímu vektoru $\mathbf{x}[n]$. Jako vzdálenost $d(\cdot, \cdot)$ použijte běžnou Euklidovu vzdálenost.



---

8. Vysvětlete ve zkratce CELP (codebook excited linear prediction) všechna čtyři slova v názvu.

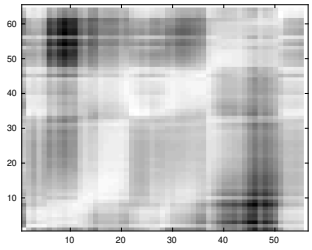
---

9. Jaký je v kódování řeči pomocí CELP vztah mezi dlouhodobým prediktorem a adaptivní kódovou knihou ?

---

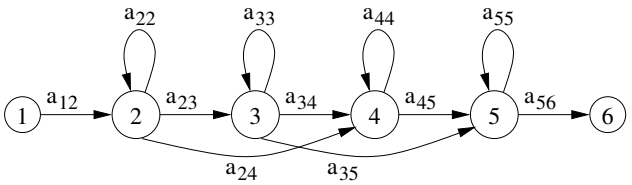
10. Proč je velikost normalizačního faktoru v klasické metodě dynamického borcení času (DTW) $N = R + T$, kde R je počet vektorů referenční promluvy a T je počet vektorů testovací promluvy ?

11. Na obrázku je matice lokálních vzdáleností vektorů ("každý s každým") pro výpočet DTW. Menší vzdálenosti jsou zobrazeny jako světlejší. Nakreslete do obrázku průběh optimální srovnávací cesty.



---

12. Srytý Markovův model (HMM) na obrázku má reprezentovat promluvu o délce $T = 4$ vektory. Napište všechny možné stavové sekvence X. Uvědomte si, že v každé sekvenci musí být stav č. 1 na začátku a stav č. 6 na konci. Tyto dva stavy nereprezentují žádný vektor.



---

13. Je definován levo-pravý HMM se čtyřmi stavy, z toho 2 vysílací, přechodové log. 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 likelihoodů je:

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

Provádíme Viterbiho algoritmus pomocí "token passing". Hodnota tokenu ve stavu 2 v čase 46 je $\Psi_2(46) = -21$. Určete hodnotu tokenu ve stavu 2 v čase 48.

$\Psi_2(48) =$ .....

---

14. Při rozpoznávání řeči pomocí váhovaných konečných stavových převodníků (wFST) je výsledná rozpoznávací síť dána jako $HCLG = H \circ C \circ L \circ G$. Napište význam nejméně dvou ze čtyř symbolů $H,\ C,\ L,\ G$.

---

15. Jedním z možných výstupů rozpoznávače řeči s velkým slovníkem může být tzv. lattice (mřížka). Popište, co v ní najdeme a/nebo malou lattici nakreslete.

16. V rozpoznávání řeči pomocí neuronových sítí se v poslední vrstvě používá speciální nelinearita, která zajišťuje, že výstupy budou dobře reprezentovat **pravděpodobnosti** jednotlivých jednotek (např. fonémů). Napište, jak se nelinearita jmenuje a napište její rovnici. Pokud si ji nepamatujete, vymyslete ji !

---

17. Co je při vyhodnocování systémů pro verifikaci řečníka (a mimochodem jakýchkoliv jiných detektorů) **equal error rate** (EER) ?

---

18. Máte k disposici modul, který ze zvukového souboru vypočítá vektorovou reprezentaci s nízkou fixní dimenzionalitou (embedding), např. extraktor i-vektorů nebo x-vektorů. Jak pro dva embeddingy dostanete skóre udávající podobnost mluvčích ?

---

19. Co v systémech pro syntézu řeči z textu (TTS) znamená "normalizace textu" ?

---

20. Popište existující nebo navrhněte vlastní systém pro syntézu řeči z textu (TTS) postavený na neuronových sítích.