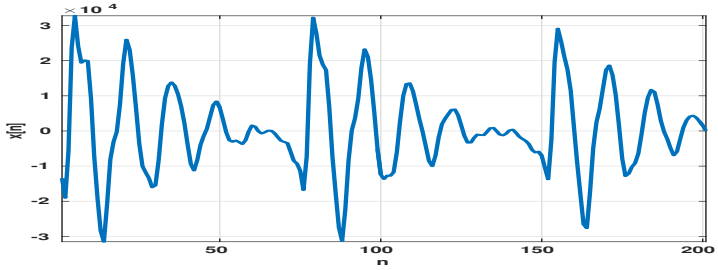
Semestrální zkouška ZRE, 1. op. termín, 29.5.2024, skupina Jelinek

Login: ..... Příjmení a jméno: ..... Podpis: .....

1. Určete frekvenci základního tónu F_0 v Hertzích pro zadaný rámec řeči. Vzorkovací frekvence $F_s = 8000$ Hz. Stačí přibližný výsledek.



---

2. Popište princip kódování řeči pomocí diferenční pulsně-kódové modulace (DPCM), především čeho diference se počítají. O kvantování, vestavbě dekodéru do kodéru, atd, psát nemusíte.

---

3. O co se v kodérech řeči typu Codebook-Excited Linear Prediction (CELP) stará perceptuální filtr $W(z)$?

---

4. Proč je implementace rozpoznávání řeči (ASR) token-passingem v logaritmické oblasti komplikovanější, pokud potřebujeme vyhodnocovat Baum-Welchův likelihood $\sum_X P(\mathbf{O}, X|M)$ na rozdíl od vyhodnocení Viterbiho likelihoodu $\max_X P(\mathbf{O}, X|M)$?

---

5. Co znamená v ASR kontextová závislost fonémů ?

6. Při běhu DTW byla naplněna matice **B** inromacemi o tom, odkud se do kterého bodu cesty přišlo. Její prvky jsou $b[m.n]$, kde m indexuje testovací promluvu o délce T a n indexuje referenční promluvu o délce R. Pokud se cesta do $[m, n]$ "pohnula" jen v referenci, je $b[m, n] = 1$, pokud jen v testu $b[m, n] = -1$, pokud v obou směrech, je $b[m, n] = 0$. Napište algoritmus (C / Python / pseudo) pro back-tracing - určení optimální srovnávací cesty DTW. Můžete ji reprezentovat jako $r[k]$ a $t[k]$, ukazující na jednotlivé vektory reference a testu, kde k je index kroku cesty.

---

7. Jak se liší váhované konečné stavové akceptory a převodníky (wFSA a wFST) od těch, co jste viděli v TINu ?

---

8. Dekódovací síť pro rozpoznávání řeči s velkým slovníkem LVCSR se staví jako komposice wFSA a wFST: $HCLG = H \circ C \circ L \circ G$. Vysvětlete význam, vstup a výstup alespoň dvou ze symbolů H, C, L, G.

---

9. Co označuje v ASR pojem lattice a proč někdy chceme, aby ASR systém takové lattice generoval ?

---

10. Co je v klasickém jazykovém modelování (LM) založeném na n-gramech "back-off" a proč je nutný ?

11. Jaký je princip jazykového modelování pomocí rekurentních neuronových sítí (RNN-LM) ?

---

12. V ASR založeném na neuronových sítích je několik přístupů, jak využít dlouhého časového kontextu. Uveďte alespoň jeden.

---

13. Proč se v moderních ASR systémech využívá jednotek typu Byte-Pair nebo Word-Piece ?

---

14. Jak se trénují velké Semi-Supervised Learning (SSL) modely typu wav2vec nebo Hubert, které "vidí" poue velká množství řečového signálu, ale ne transkripce ?

---

15. Jak se v klasickém GMM přístupu k rozpoznávání mluvčího tvoří model pozadí (universal background model, UBM) a daného mluvčího ?

16. Jak byste využili velké SSL modely pro rozpoznávání mluvčího ?

---

17. Na detection trade-off error (DET) křivce je na vodorovné ose pravděpodobnost chybného přijetí (false alarm) P_{FA} a na svislé ose pravděpodobnost chybného odmítnutí (false rejection) P_{FR}. Nakreslete příklad DET křivky a ukažte na ní (a) region zajímavý pro přístupový systém, (b) region zajímavý pro hledání teroristy ve velkém množství nahrávek. Velmi krátce vysvětlete.

---

18. Uveďte, co je v syntéze z textu (TTS) potřeba dělat s prozodií, pokud je v textu tečka ukončující větu.

---

19. V TTS založené na výběru jednotek (unit selection) se hovoří o "target cost" a "concatentation (join) cost". Co to znamená ?

---

20. Jaký je rozdíl mezi klasickou a end-to-end TTS ?