

Semestrální zkouška ZRE, řádný termín, 18.5.2023, skupina 8re0

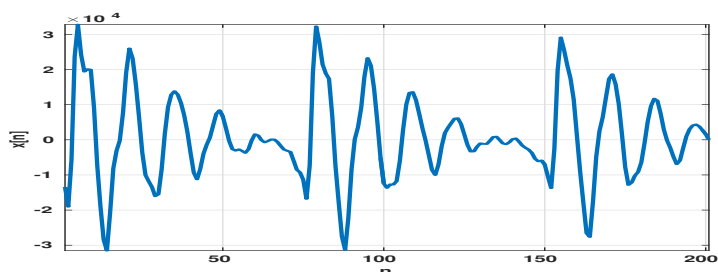
Login: Příjmení a jméno: Podpis:

1. Vysvětlete, jaký bude rozdíl v DFT spektru signálu $x[n]$ a ustředněného signálu $x'[n] = x[n] - \bar{x}[n]$.

-
2. Řámec řeči se vzorkovací frekvencí $F_s = 8000$ Hz má první formant $F_1 = 200$ Hz a druhý formant $F_2 = 2000$ Hz. Nakreslete póly filtru $H(z) = \frac{1}{A(z)}$ 4. řádu, který ho reprezentuje.

-
3. Je dán “zubatý” signál o délce 10 vzorků: $[1, -1, 1, -1, 1, -1, 1, -1, 1, -1]$. Výpočtem nebo odhadem určete hodnotu jediného koeficientu a_1 prediktoru 1. řádu: $1 - A(z) = -a_1 z^{-1}$.

-
4. Nakreslete velmi přibližně průběh autokorelačních koeficientů rámce řeči na obrázku. Cílem je detekce základního tónu.



-
5. Proč jsou v CELP kodecích dvě kódové knihy (a je jedna z nich vůbec kódová kniha) ?

6. Co je u filtrů v kódování CELP (a obecně u jakýchkoliv číslicových filtrů) zero-input response a jak je možné, že může být nenulová ?

7. Popište nebo nakreslete, jak byste implementovali detektor řečové aktivity (voice activity detector, VAD) pro automatické rozpoznávání řeči (ASR).

8. Proč byly až do nástupu neuronových sítí Mel-frekvenční cepstrální koeficienty (MFCC) nejoblíbenější parametry pro ASR ?

9. Referenční promluva je slovo “kvák” o délce 0.5 s. Testovací promluva je také “kvák”, ale o délce 1 s, kde první půlsekunda je ticho. Nakreslete, jak bude vypadat srovnávací cesta dynamického borcení času (DTW).

10. Pro zadanou matici přechodových pravděpodobností skrytého Markovova modelu (HMM) o pěti stavech nakreslete model a napište možné stavové sekvence pro 5 vstupních vektorů. První a poslední stav HMM nejsou emitující (“nežerou” žádný vstupní vektor).

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

11. Proč je implementace token-passingu komplikovanější, pokud potřebujeme vyhodnocovat Baum-Welchův likelihood $\sum_X P(\mathbf{O}, X|M)$ na rozdíl od vyhodnocení Viterbiho likelihoodu $\max_X P(\mathbf{O}, X|M)$?

12. Popište, jaké má klasický statistický jazykový model (LM) parametry a jak se určují.

13. Jaké jsou možnosti reprezentace výstupu, pokud má ASR produkovat něco více než jen nejlepší sekvenci slov (1-best) ?

14. K čemu jsou v ASR dobré váhované konečné stavové transducery (WFST) ?

15. ASR systémy založené na neuronových sítích používají v poslední vrstvě nelinearitu typu softmax. Proč ?

16. Vyberte si příklad **jednoho** moderního ASR systému založeného na NN (RNN-T, sequence to sequence, Listen&spell, transducer, ...) a vysvětlete, jak pracuje.

17. V současné době jsou pro reprezentaci řeči populární velké neurální modely předtrénované bez přepisů (wav2vec, WavLM, ...). Popište nebo vymyslete, jak byste takový model natrénovali.

18. U rozpoznávání mluvího bojujeme s nechtěnou variabilitou. Uveďte, jakou variabilitu generuje mluvčí sám (intrinsic variability - sem nepatří mikrofon, přenosový kanál, šum, reverberace, atd).

19. Moderní systémy pro verifikaci mluvího nejprve reprezentují nahrávku pomocí jednoho vektoru (embedding), pak srovnávají dva embeddingy dvou promluv a produkují skóre “same/different”. Uveďte nebo vymyslete, jak z takových dvou embeddingů skóre vygenerovat.

20. Proč je ve verifikaci mluvího (a obecně v jakékoliv detekční úloze) důležitá kalibrace skóre ?