

1. Proč je rozpoznávání spojitě spontánní řeči s velkým slovníkem složitější než volba telefonních čísel hlasem ?

A: mluvčí může být ovlivněn nemocí nebo stresem.

B: Je nutné řešit změnu slov v kontextu (koartikulace) a jazykový model.

C: Pro rozpoznávání s velkým slovníkem je nutné volit podstatně kratší délku rámce, parametrizace bude tedy výpočetně náročnější.

D: Před rozpoznáváním spojitě řeči je nutné ručně označit polohy jednotlivých slov - tato úloha se nedá zautomatizovat.

2. Dva FIR filtry s impulsními odezvami (pro $n = [0 \ 1 \ 2]$):

$$h_1 = [1 \ 2 \ 3]$$

$$h_2 = [1 \ -2 \ 3]$$

jsou zapojeny paralelně (vedle sebe) a jejich výstupy jsou sečteny. Jaká je impulsní odezva celého systému ?

A: $h = [0 \ 4 \ 0]$

B: $h = [2 \ 0 \ 6]$

C: $h = [0 \ 0 \ 0]$, takže filtr nic nepropouští.

D: $h = [1 \ 0 \ 2 \ 0 \ 9]$.

3. Filtr, který signál nijak nemění, ale zpožďuje jej o 53 vzorků, bude mít přenosovou funkci:

A: $H(z) = \sum_{k=0}^{53} b_k z^{-k}$ se všemi koeficienty b_k nenulovými.

B: $H(z) = \frac{\sum_{k=0}^{53} b_k z^{-k}}{\sum_{k=1}^{53} a_k z^{-k}}$ se všemi koeficienty b_k a a_k nenulovými.

C: $H(z) = \frac{1}{\sum_{k=1}^{53} a_k z^{-k}}$ se všemi koeficienty a_k nenulovými.

D: $H(z) = z^{-53}$

4. Při určování frekvenční charakteristiky filtru **neodpovídá** jeden "oběh" bodu $e^{j2\pi f}$ (F_s je vzorkovací frekvence):

A: intervalu $0 \dots 2\pi$ pro normované kruhové frekvence v rad.

B: intervalu $0 \dots 2$ pro normované frekvence.

C: intervalu $0 \dots 2\pi F_s$ pro kruhové frekvence v rad/s.

D: intervalu $0 \dots F_s$ pro obyčejné frekvence v Hz.

5. Obvyklá délka rámce řeči pro rozpoznávání je 25 ms, s překrytím 15 ms. Určete, kolik takových rámců je obsaženo v 1 s řeči (uvažujte 1 s řeči uprostřed např. minutové promluvy).

A: 40

B: přibližně 67

C: 100

D: 25.

6. Detektor řečové aktivity založený na střední krátkodobé energii bude správně pracovat:

- A: pro řečový signál bez šumu.
 - B: v autě.
 - C: pro řeč nahranou v silně větrném prostředí.
 - D: pro řeč z prostředí bojiště (palné zbraně, křik, atd.)
-

7. Před odhadem počtu průchodů nulou je nutné provést odečtení stejnosměrné složky řečového signálu. Pokud bychom to neudělali:

- A: v rámcích s velkou energií (samohlásky – 'a', 'e', 'i', atd.) by mohl být odhad v extrémním případě i nulový (žádný průchod nulou).
 - B: v rámcích s malou energií (frikativy – 's', 'f', 'ch', atd.) by mohl být odhad v extrémním případě i nulový (žádný průchod nulou).
 - C: nenulová stejnosměrná složka způsobí, že v každém rámci je napočítán počet průchodů nulou $\frac{l_{ram}}{2}$, kde l_{ram} je délka rámce.
 - D: nenulová stejnosměrná složka způsobí, že odhad počtu průchodů nulou bude ve všech rámcích stejný.
-

8. Polohy formantů jsou určeny:

- A: Tím, jak moc se člověk nadechne (tlakem v plicích).
 - B: Frekvencí kmitání hlasivek.
 - C: Postavením artikulačního traktu, především jazyka.
 - D: Polohou mluvího (stojící, sedící, ležící, ...).
-

9. Spektrum jednoho rámce řeči, na kterém bude vidět pouze hrubá struktura daná rezonancemi artikulačního traktu (ne buzení) **nelze** dostat takto:

- A: Odhad koeficientů filtru, který modeluje artikulační trakt a zobrazení jeho frekvenční charakteristiky.
 - B: Průchod signálu horní propustí s mezní frekvencí 400 Hz, výpočet jeho spektra, zobrazení modulu.
 - C: Výpočet cepstra, vynulování koeficientů pro $n > 30$, zpětný převod na spektrum.
 - D: Odhadem koeficientů filtru, který modeluje artikulační trakt, výpočet jeho impulsní odezvy, a zobrazení modulového spektra této impulsní odezvy.
-

10. Jak by se dal jednoduše popsat výpočet cepstrálních koeficientů z jednoho rámce řeči ?

- A: FFT, filtrace IIR filtrem, zpětná FFT.
- B: FFT, filtrace IIR filtrem, moduly na druhou, logaritmus, zpětná FFT.
- C: FFT, moduly na druhou, logaritmus, zpětná FFT.
- D: FFT, zpětná FFT.