

1. Lexikální úroveň zpracování řeči se projeví v této aplikační oblasti:

A: kódování

B: analýza a zobrazování spektra

C: rozpoznávání řeči s velkým slovníkem

D: rozpoznávání mluvího založené na statistickém modelování vektorů akustických parametrů (např. cepstrálních koeficientů).

2. Proč nelze z řeči prošlé telefonním kanálem ($F_s = 8000$ Hz) rozpoznat, zda mluví řekl izolované 's' nebo 'f' ?

A: Říkat tyto hlásky izolovaně do telefonu je zakázáno, pokud je přesto někdo řekne, běžné telefonní ústředny je nahradí tichem.

B: Hlásky se odlišují jemnými vibracemi dolního rtu na nízkých frekvencích, které nejsou tlf. kanálem přenášeny.

C: Hlásky se odlišují ve spektru nad 4000 Hz, tyto frekvence nejsou tlf. kanálem přenášeny.

D: Hlásky trvají většinou méně než 1 vzorkovací periodu $\frac{1}{8000}$ s.

3. U verifikace mluvího je úkolem

A: Říci, zda dvě nahrávky patří témuž mluvímu nebo ne.

B: Rozhodnout, který z N mluvích je na dané nahrávce.

C: Určit, zda na nahrávce je celebrita (od které máme pouze interview v časopise, nikoliv nahrávku)

D: Určit, zda jsou dvě nahrávky jednoho mluvího stejné (např. pro ochranu copyrightu).

4. Bloky z^{-1} jsou v implementaci číslicových filtrů v C realizovány:

A: funkcemi

B: Paměťovými buňkami

C: inkrementací pomocí operátoru ++

D: dekrementací pomocí operátoru --

5. Máme bílý šum (signál, kde jednotlivé vzorky nejsou korelovány) a 3 číslicové filtry s těmito přenosovými funkcemi:

$$H_1(z) = \frac{1 + 0.5z^{-1}}{1 - 0.5z^{-1}}$$

$$H_2(z) = 1 + 0.5z^{-1}$$

$$H_3(z) = 1 - 0.5z^{-1}$$

Které z nich způsobí, že signál na výstupu již nebude bílý (neboli jednotlivé vzorky **budou** korelovány) ?

A: pouze H_1

B: H_2 a H_3

C: ani jeden

D: všechny

6. Provádíme DFT s $N = 1024$ vzorky signálu, který má vzorkovací frekvenci $F_s = 8000$ Hz. Které frekvenci odpovídá poslední bod v intervalu: $N - 1$?

A: 7.81 Hz

B: polovině vzorkovací frekvence: 4000 Hz.

C: 7992.2 Hz

D: vzorkovací frekvenci: 8000 Hz.

7. Čistý řečový signál je $s[n]$. Stejný signál, ale s aditivně přidaným šumem, je $ss[n]$. Určete, jaký bude vztah mezi průběhem střední krátkodobé energie pro oba signály:

A: Energie $ss[n]$ bude vždy vyšší než $s[n]$.

B: Energie $ss[n]$ bude v některých rámcích vyšší než $s[n]$, v některých nižší.

C: Energie $ss[n]$ bude vyšší než $s[n]$ jen pro úseky ticha.

D: Energie $ss[n]$ bude vyšší než $s[n]$ jen pro vysokoenergetické úseky řeči (samohlásky).

8. Obyvatelé planety $\tau\xi\lambda\nu$ mají na rozdíl od lidí lepší frekvenční rozlišení na vysokých frekvencích. Jak byste pro rozpoznávání jejich řeči upravili výpočet Mel-frekvenčních cepstrálních koeficientů MFCC ?

A: Zvýšili bychom vzorkovací frekvenci.

B: Změnili bychom funkci pro frekvenční "warpování", která přepočítává Hz na Mely.

C: Na stávající Melovské stupnici pro lidi bychom umístili lineárně méně filtrů.

D: Vypustili bychom DCT na konci výpočtu MFCC.

9. Jak by u metody LPC **nešel** určit použitelný řád LPC filtru $\frac{1}{A(z)}$.

A: podle průběhu energie chyby predikce $E(i)$ v závislosti na řádu prediktoru.

B: odhadem pomocí počtu rezonančních frekvencí hlasového ústrojí.

C: poslechovými pokusy, kdy by se řeč generovala filtrováním bílého šumu filtrem $\frac{1}{A(z)}$ a vyhodnocovala by se srozumitelnost.

D: počítáním průměrného počtu period řečového signálu za jeden rámeček.

10. Při výpočtu koeficientů LPC se používají nenormalizované autokorelační koeficienty

$$R[k] = \sum_{n=0}^{N-1-k} x[n]x[n+k].$$

Jak by se na vypočtených koeficientech a_i projevilo, kdybychom místo nich použili standardní vychýlený odhad autokorelačních koeficientů:

$$R[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]x[n+k],$$

který se od nenormalizovaných liší o konstantu?

A: nijak

B: všechny koeficienty by byly vynásobeny hodnotou N

C: všechny koeficienty by byly vynásobeny hodnotou N^2

D: koeficienty by byly násobeny hodnotami $(N - i)$ podle svého indexu.