

Název veřejné zakázky:

Dodávka řečových databází pro projekt NTIS

Odůvodnění vymezení technických podmínek podle § 156 odst. 1 písm. c) zákona č. 137/2006 Sb., o veřejných zakázkách

Technická podmínka:

Odůvodnění

Frekvence vzorkování 8 kHz pro databáze pro telefonní aplikace, 16 kHz pro databáze pro diktovací aplikace

Frekvence vzorkování velice úzce souvisí s přenášeným pásmem v daném kanálu. Z Nyquistova (též Shannonova či Kotělnikovova) vzorkovacího teorému vyplývá, že má-li být spojitý signál zpětně rekonstruovatelný ze své digitalizované verze, musí vzorkování spojitého signálu probíhat s alespoň dvojnásobnou frekvencí, než je nejvyšší frekvence obsažená v signálu. Jestliže rozsah přenášeného pásma u telefonního kanálu je tedy přibližně 0 až 4 kHz, musí být vzorkovací frekvence pro signál přenášený telefonním kanálem alespoň 8 kHz. Mikrofon umožňuje přenést frekvenční pásmo, jehož horní mez se v závislosti na typu mikrofonu pohybuje od 8 kHz výše. Minimální hodnota vzorkovací frekvence u mikrofonních nahrávek je tedy 16 kHz. Pokud by byla zvolena vzorkovací frekvence nižší, než jaká vyplývá z Nyquistova vzorkovacího teorému, docházelo by ke zkreslení frekvenčních charakteristik řečového signálu a získaný navzorkovaný signál by byl nepoužitelný pro další zpracování.

Rozlišení 8 bitů pro databáze pro telefonní aplikace, 16 bitů pro databáze pro diktovací aplikace

Rozlišení souvisí s přenášeným pásmem, a to z hlediska odstupu signálu od kvantizačního šumu. Pro telefonní kanál, kde je dynamický rozsah řečového signálu zhruba 60 dB, je třeba minimálně 8 bitové rozlišení. Pro mikrofonní kanál, kde se doporučuje odstup signálu od kvantizačního šumu alespoň 90 dB, je třeba minimálně 16 bitové rozlišení. Pokud by bylo zvoleno rozlišení menší, docházelo by k příliš velkému kvantizačnímu zkreslení a nahraný řečový signál by byl nesrozumitelný.

řečová data musí být ve formátu PCM, A-law nebo u-law

Formát řečových dat je určen způsobem zakódování spojitého signálu do digitální verze. V současné době je ve světě standardně využíváno kódování PCM, u-law a A-law. Tento způsob kódování podporují všechny dostupné systémy zpracování řečového signálu. Jakýkoli jiný nestandardní způsob kódování by znamenal nemožnost využití stávajícího programového vybavení a vyvolal by potřebu získání či vývoje speciálního zařízení umožňujícího s takovým kódováním pracovat.

Rozsah databází pro telefonní a diktovací aplikace, tj. dle typu databáze počet řečníků 500 či 1000 a doba promluv 200 hodin

Rozsah databáze (tj. celková doba promluv a počet řečníků) ovlivňuje chybovost rozpoznávače. Pokud bude rozpoznávač natrénován pomocí nahrávek od jediného řečníka, bude správně rozpoznávat promluvy tohoto řečníka, u řečníků jiných bude vykazovat nepřijatelnou chybovost. Pokud má být tedy systém využíván různými řečníky, je třeba k trénování využít data získaná od různých řečníků. Ze zákona velkých čísel z matematické statistiky přitom vyplývá, že čím více řečníků bude k dispozici, tím lépe rozpoznávač namodeluje libovolného (průměrného) řečníka. Rovněž čím více promluv bude využito k natrénování rozpoznávače, tím menší chybovost bude rozpoznávač vykazovat. Odhaduje se, že pro dosažení nulové chybovosti by bylo zapotřebí alespoň 600 tisíc hodin řeči. Pokud množství řeči klesne pod zhruba 200 hodin, je chybovost větší než 20%, což znamená, že minimálně 1 slovo z 5 je rozpoznáno špatně a výsledný text generovaný rozpoznávačem je pak nesrozumitelný.

Dostupnost transkripce

Dostupnost transkripce je nezbytná pro nastavení parametrů rozpoznávače souvisejících s fonetikou daného jazyka. Bez informace, co konkrétně zaznělo v dané promluvě, se rozpoznávač nenaučí převádět jednotlivé zvuky na jejich textovou reprezentaci.

Dostupnost informace o tom, který řečník vyslovil kterou promluvu

Dostupnost informace o tom, který řečník vyslovil kterou promluvu, je nezbytná k vyváženému namodelování modelu obecného (průměrného) řečníka. Jestliže by tato informace nebyla k dispozici, hrozí nebezpečí, že od nějakého řečníka bude pro trénování využito více dat než od jiných řečníků a tím dojde k nesprávnému nastavení modelu obecného řečníka a tudíž k nesprávné funkčnosti celého systému rozpoznávání řeči.