Hlasová komunikácia v inteligentnom automobile (2)

Mikuláš Alexík, Viliam Mateička

Extrakcia akustických vlastností reči

Takto predspracovaný signál možno poslať do extraktora vlastností signálu (FE – feature extractor), ktorý vykoná na oknách nejaký druh krátkodobej spektrálnej analýzy a zredukuje tým počet parametrov pri zachovaní užitočnej informácie. Úlohou metód extrakcie reči je premeniť signál reči $s(i)$ rozdelený v predchádzajúcej fáze do okien na postupnosť vektorov vlastností zvuku (feature vectors) $x(n)$. Tak sa vytvorí kompaktnejšia reprezentácia vstupného rečového signálu, z ktorou sa lepšie pracuje. Taký proces sa nazýva extrakcia vlastností reči (feature extraction). Tieto extrahované vlastnosti možno rozšíriť aj o ich dynamické vlastnosti, často používané sú prvé a druhé derivácie. Takýto komplexný vektor, obvykle ešte transformovaný na menej rozmerný, aby sa zvýšila odolnosť systému, postupuje do ďalšej časti, kde sa štatisticky modelujú rôzne nerovnakosti jednotlivých slov. Extrakcia vlastností sa vykonáva v troch fázach. Prvá, nazývaná analýza reči (speech analysis) alebo aj akustický frontend, vykonáva nejaký druh spektrálnej alebo časovej analýzy signálu a vytvára základné vlastnosti opisujúce do okien zaobalené výkonové spektrum v krátkych časových intervaloch. V aplikácii boli realizované: lineárna predikčná analýza (LPC, LPREFC, LPCEPSTRA), analýza Mellovou frekvenčnou škálou, analýza bankou filtrov (FBANK, MFCC), perceptuálna lineárna predikčná analýza (PLP). Druhá časť rozšíri tento vektor vlastností o jeho dynamické vlastnosti (delta, akceleračné koeficienty). Nakoniec v poslednej fáze, ktorá je nie vždy prítomná, sa transformuje tento rozšírený vektor vlastností do kompaktnejších a odolnejších vlastností a tie už vstupujú do rozpoznávacieho procesu (linear diskrimination analysis – LDA, principial component analysis – PCA).

Na modelovanie podobnosti vektorov vlastností reči bola aplikovaná štatistická metóda (HMM), kde je táto postupnosť použitá na výpočet pravdepodobnosti pozorovania tejto sekvencie, ak bolo vyslovené určité slovo. Pre danú sekvenciu je rozpoznanie určitého slova iba vyhľadávací proces, v ktorom sa hľadá model slova, pre ktoré je pravdepodobnosť pozorovania práve takejto sekvencie najväčšia z modelov jednotlivých slov. V tomto procese treba najskôr zistiť výstupnú pravdepodobnosť pozorovania v jednotlivých modeloch. Ak sa použijú diskrétne HMM, je tento proces rýchlejší ako v spojitých HMM, pretože vektoru vlastností reči stačí iba priradiť index vektora, ku ktorému je tento vektor najbližšie (Euklidovskou vzdialenosťou) vo vektorovom kvantizéri, a potom tento index použiť v tabuľkách jednotlivých modelov na nájdenie pravdepodobnosti. Ak ide o systém rozpoznávania izolovaných slov, je potrebná aj detekcia reči, ktorá ma za úlohu určiť hranice spracúvaného slova, ktoré vstupuje do HMM rozpoznávača.

Realizácia programu a niektoré výsledky

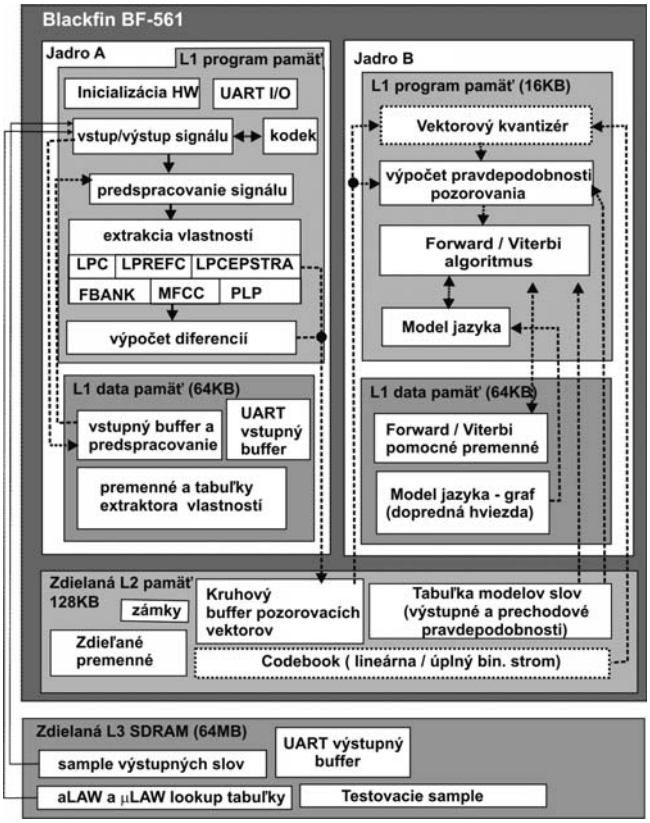
Od začiatku programovej realizácie algoritmov rozpoznávania bolo potrebné myslieť na využitie oboch procesorových jadier, aby sa čo najviac využíval potenciál procesora. Jadro „A“ je v implementácii vyťažené pri najnáročnejšej metóde extrakcii reči (PLP) „iba“ na 38 % (extrakcia jedného okna trvá 3,8 ms, pre porovnanie MFCC trvá asi 2,3 ms, LPC 1,1 ms) svojho procesorového času, takže by bolo možné pridať komplikovanejšie metódy extrakcie alebo implementovať LDA alebo PCA. Keďže však ide o synchrónny proces, treba dbať na to, aby čas spracovania jedného okna nepresiahol určitý čas – periódu okien (10 ms).

Realizovaný program, ktorého bloková schéma je na obr. 3, sa skladá zo štyroch častí:

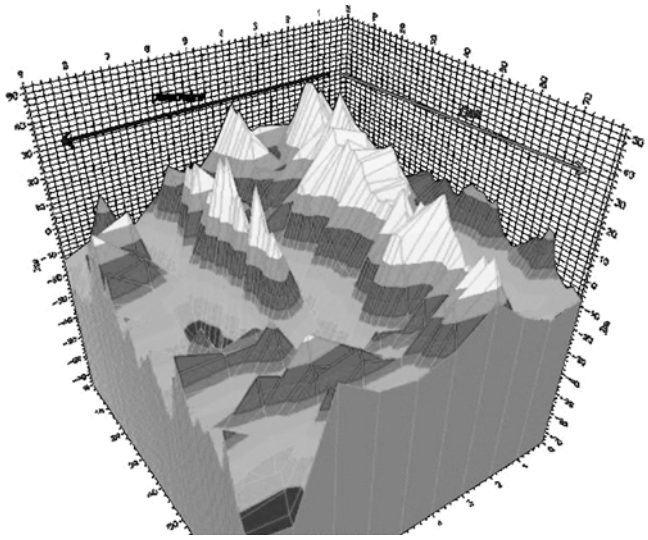
1. Jadro A
- inicializácia procesora (CORECLK, SYSCLK), periférií (kodek, DMA, USART, GPIO), softvéru jadra A a využívanej pamäte,
- spracovanie prerušení (kodek, USART, GPIO), komunikačné rutiny,
- predspracovanie vstupu z A/D, generovanie výstupu do D/A,
- metódy extrakcie reči, parametrizácia okien, extrakcia vlastností reči, nastavenie konštánt a metód extraktora.

2. Jadro B
- inicializácia softvéru jadra B, určenie výstupnej pravdepodobnosti,
- definícia štruktúry modelu,
- inicializácia modelov (prepočet prechodových pravdepodobností do logaritmických), logaritmická aritmetika, komunikačné rutiny,
- forward a Viterbiho algoritmus, vektorový kvantizér, codebook.

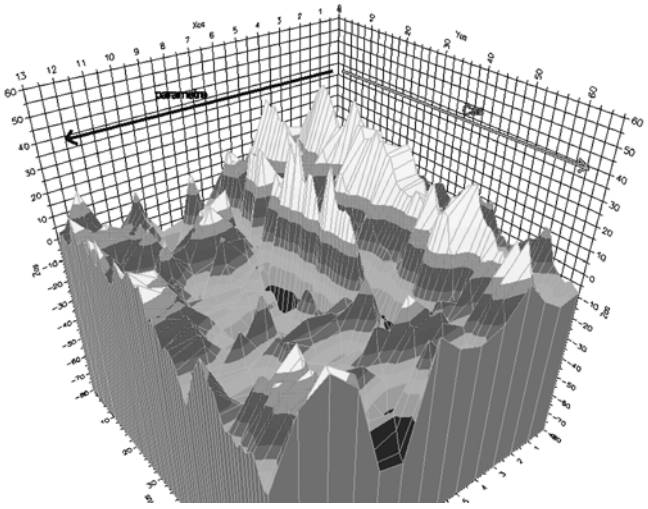
3. Využívaná L2 pamäť
- tabuľka modelov, rad pozorovacích vektorov,
- zámky a využívané premenné.



Obr.3 Bloková schéma rozdelenia programu



Obr.4 Parametre slova „jeden“, parametrizácia MFCC-10



Obr.5 Parametre slova „jeden“, parametrizácia MFCC-14

4. Využívaná L3 pamäť

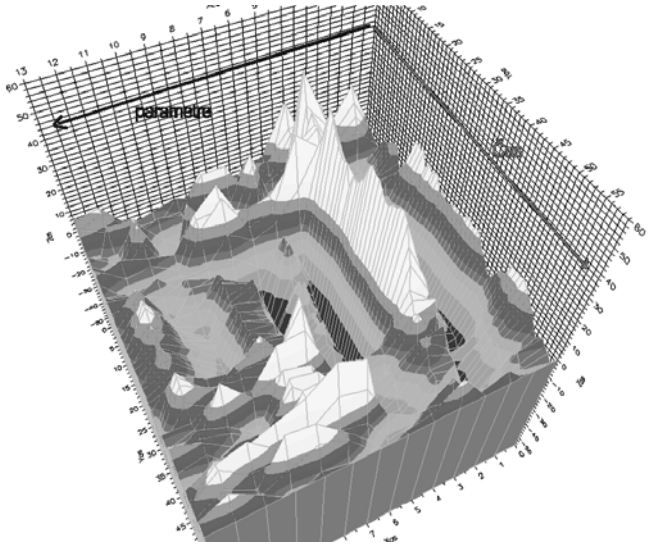
- knižnica samplov na prehrávanie a na testovanie,
- tabuľka na dekódovanie uLAW a aLAW, vysielací UART buffer.

Globálny projekt

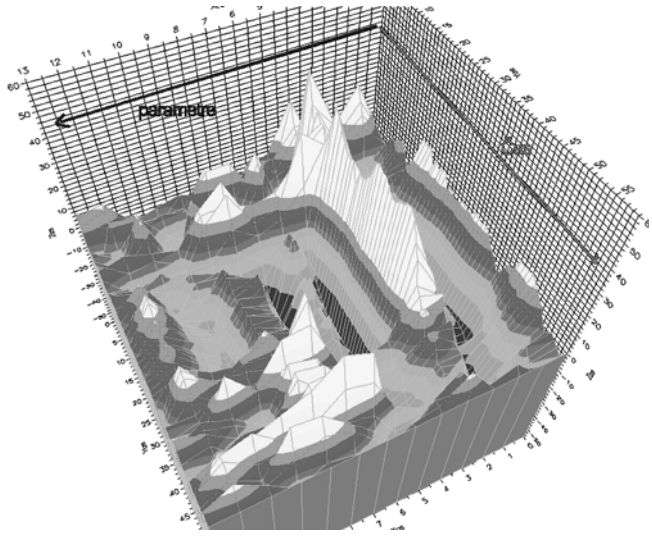
- komunikačné konštanty, nastavenie veľkosti bufferov, matematické konštanty,
- globálne štruktúry, systémové konštanty, inicializácia systémových hodín procesora, systémovej zbernice a L3 pamäte.

Pri implementácii programu bolo použitých 6 prerušení, všetky smerované do jadra A. Dôležité je pripomenúť, že použitý DSP Blackfin používa na uloženie celých čísel v pamäti taký istý formát ako x86 kompatibilný procesor – na úrovni znakov little endian, t. j. na nižšej pozícii byte je menej významný bit, na úrovni slov (16-bitový a 32-bitový) je to takisto. Toto značne uľahčilo komunikačný protokol s PC, ktorý nemusí zohľadniť konverzie medzi jednotlivými formátmi dát procesorov. Ďalším plusom bolo, že softvérovo emulovaný typ s pohyblivou desatinnou čiarkou – float (32-bitový) a double (64-bitový) – sú tiež na obidvoch procesoroch rovnako reprezentované v pamäti.

Na obr. 4, 5, 6 sú znázornené slová „jeden“ a „dva“ pri ich charakterizovaní kepstralnými koeficientmi Melovho frekvenčného spektra (MFCC feature extraction). Na obr. 4 je zobrazených 10 koeficientov slova „jeden“ a na obr. 5 je to 14 koeficientov toho istého slova. Na obr. 6 je zobrazených 14 MFCC koeficientov slova „dva“, na obr. 7 je to 24 koeficientov tohto slova. Z porovnania obr. 6 a 7 je zrejmé, že prvých 14 koeficientov dostatočne charakterizuje slovo a nie je nutné rozšírenie o ďalších 10 koeficientov.



Obr.6 Parametre slova „dva“, parametrizácia MFCC-14



Obr.7 Parametre slova „dva“, parametrizácia MFCC-24

Z porovnania obr. 4 a 5 je zrejmé, že 14 koeficientov poskytuje väčšie možnosti na charakterizovanie slova. Nakoniec z porovnania obr. 5 a 6 je viditeľný rozdiel medzi zobrazením MFCC koeficientov pre slová „jeden“ a „dva“.

Literatúra

[1] MATEIČKA, V.: Implementácia systému rozpoznávania reči nezávislého od rečníka do DSP Blackfin-BF561. Diplomová práca, FRI, KTK ŽU v Žiline, 2006

[2] HOLADA, M.: (1999) Comparing diferent parameterization methods for telephone speech recognition. In: Proc. of 9th Czech-German Workshop Speech Processing, Prague, September 1999.

[3] Survey of the State of the Art in Human Language Technology 1996 http://cslu.cse.ogi.edu/HLTsurvey/index.html

[4] Hidden Markov Toolkit – HTKBook http://htk.eng.cam.ac.uk/

Pokračovanie v budúcom čísle.

prof. Ing. Mikuláš Alexík, PhD.
Ing. Viliam Mateička

Žilinská univerzita
Fakulta riadenia a informatiky
Katedra technickej kybernetiky
Veľký diel, 010 26 Žilina
e-mail: Mikulas.Alexik@fri.utc.sk
viliam.mateicka@siemens.com