

VOICE CONTROLLED CALCULATOR

Ota Pavelek

Bachelor Degree Programme (3), FIT BUT

E-mail: xpavel08@stud.fit.vutbr.cz

Supervised by: František Grézl

E-mail: grezl@fit.vutbr.cz

Abstract: This paper describes implementation of calculator, which can be controlled by both voice and normal way. Speech recognizing is realized with BSCore and has recognition network restricted to words needed by calculator. When the recognition is done, recognized sentence is transformed to expression and shown to user, so it can be corrected (especially in case of erroneous recognition). Expression is evaluated on user's request. The purpose of voice control is to make usage of calculator more effective.

Keywords: calculator, speech recognition

1 ÚVOD

Tato bakalářská práce se zabývá implementací hlasem ovládané kalkulačky. Hlasové ovládání se v poslední době dostává do popředí v oblasti uživatelských rozhraní [1]. Má obecně za cíl zefektivnit komunikaci (zejména nezkušených) uživatelů s počítačem. Místo pomalé manipulace s klávesnicí a myší komunikuje uživatel s počítačem přirozeným jazykem. Tento systém se dnes používá například v rezervačních systémech, callcentrech a mobilních telefonech[2]. Použitelnost systému se odvíjí od přesnosti použitého rozpoznávače. Přesnost rozpoznávačů mimo jiné závisí na tom, má-li rozpoznávač poznávat hlasy více lidí nebo jen jednoho člověka. Objevuje se též snaha o použití rozpoznávání hlasu pro autentizaci - například projekt MOBIO¹.

Mým cílem je implementovat kalkulačku, která by byla kompletně ovládána hlasem. Rozpoznávač je popsán v sekci 2, implementace v sekcích 3 a 4.

2 ROZPOZNÁVAČ

Použitý rozpoznávač je odvozen z rozpoznávače pro plynulou řeč s velkým slovníkem. Ten modeluje elementární řečové jednotky - kontextově závislé fonémy - pomocí skrytých markovových modelů (HMM). Z těchto jednotek se pak skládají slova podle tzv. výslovnostního slovníku. Napojení slov se řídí „jazykovým modelem“ udávajícím pravděpodobnosti N-tic slov. Vše je zkompilováno do jedné rozpoznávací sítě, kterou pak dekodovací algoritmus hledá nejpravděpodobnější cestu[3]. Vstupem rozpoznávače je nahrávka 8kHz MONO PCM, jeho výstupem je sekvence rozpoznávaných slov.

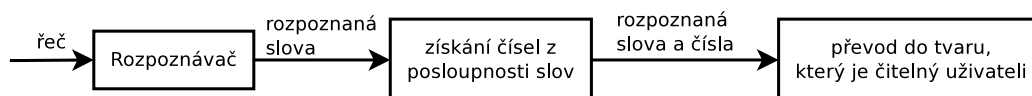
Pro potřeby kalkulačky byl slovník omezen na 66 slov a byla vytvořena nová rozpoznávací síť, která umožňuje libovolnou posloupnost slov. To výrazně urychlí proces rozpoznávání a zaručí rychlou odezvu i při zpracování dlouhé nahrávky. Správnost posloupnosti rozpoznávaných slov je kontrolována při převodu do reverzní polské notace (vizte kapitola 4). Rozpoznávač je zapouzdřen do rozhraní BSApi, které je volně dostupné pro nekomerční použití².

¹<http://www.mobioproject.org>

²<http://www.phonexia.com/docs/bsapi/>

3 ZÍSKÁNÍ VSTUPU OD UŽIVATELE A JEHO ZOBRAZENÍ

Vstup z mikrofonu je zpracován rozpoznávačem. Výstupem rozpoznávače je sekvence slov. Tato sekvence slov se zobrazí v čitelné formě uživateli, aby měl možnost opravit chybně rozpoznaná slova.



Obrázek 1: Znázornění převodu slov

3.1 ZOBRAZENÍ ČÍSEL

V sekvenci rozpoznáných slov se nejdříve hledá posloupnost číslovek tvořící číslo - např. číslo 128 rozpoznávač rozezná jako tři slova - „sto dvacet osm“. Algoritmus, který hledá posloupnosti číslovek tvořící číslo, je řízen tabulkou (příklad vizte tabulku 1). V této tabulce je na řádce stávající číslovka, ve sloupci následující číslovka. Pokud výskyt následující číslovky za současnou má smysl, je považována za součást čísla - např. „sto deset“ je považováno za jedno číslo, zatímco „deset sto“ už jsou čísla dvě.

	0	1 - 9	10	11 - 19	20, 30, ..., 90	100
8	F	F	F	F	F	F
20	F	T	F	F	F	F
100	F	T	T	T	T	F

Tabulka 1: Příklad tabulky, která řídí hledání posloupnosti čísel pro číslo 128

3.2 ZOBRAZENÍ FUNKCÍ

Kromě číslovek také vyžaduje pozornost zobrazení funkcí. Je totiž odlišit jednotlivé parametry funkcí a následně je oddělené čárkami vložit do závorek. Například „*logaritmus základu deset čísla sto*“ se zobrazí jako $\log(10, 100)$. U posledního parametru ve výše uvedeném výrazu se v případě delších výrazů řeší problém, co je ještě parametr a co už je další část výrazu. Problém je vyřešen tak, že se jako parametr bere pouze první číslo (nebo funkce) za slovem „čísla“ (vizte výše uvedený příklad). Pokud chce uživatel jako poslední parametr použít nějaký složitější výraz, musí ho uzavřít závorkami.

4 VÝPOČET VÝRAZU

Pokud se ve výrazu vyskytují funkce, tak je nejdříve vypočítána jejich hodnota. Parametry goniometrických funkcí lze zadat buď ve stupních, radiánech nebo gradientech. Vypočtenou hodnotou je pak nahrazen výskyt funkce ve výrazu. Tato náhrada je provedena jen za účelem kontroly výrazu a pro pozdější výpočet, uživateli se však stále zobrazuje funkce v původním tvaru.

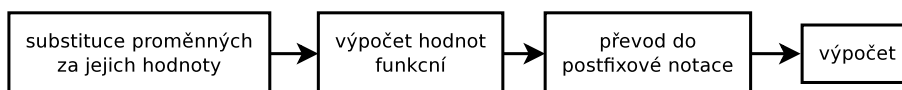
Výraz zobrazený uživateli je v *infixové notaci*³. Pro elegantní výpočet výrazu pomocí zásobníku je potřeba výraz převést do *postfixové notace*⁴[4]. Při převodu do postfixové notace dochází ke kontrole posloupnosti slov. V případě úspěšného převodu je výraz vypočítán a výsledek je zobrazen uživateli.

Při pokusu o výpočet špatně zadaného výrazu (např. chybí-li operátor) je uživatel informován o chybě. Zadaný výraz zůstane nezměněn, tudíž ho má uživatel možnost opravit. Ovšem i v případě úspěšně vypočítaného výrazu je uživateli ponechána možnost tento výraz opětovně zobrazit a editovat.

³operátor se nachází mezi svými operandy

⁴synonymum je reverzní polská notace, operátor se nachází za svými operandy

Vypočítaný výraz se dá uložit do proměnné. Tyto proměnné jsou v programu celkem tři (x , y a z). Proměnné lze poté používat v dalších výrazech.



Obrázek 2: Znárodnění postupu při výpočtu výrazu

5 ZÁVĚR

Implementace je provedena v jazyce C++. Na grafické uživatelské rozhraní je použit framework Qt, k rozpoznávání řeči knihovna BSCore. Pro rozpoznávač je vytvořena síť, která rozpoznává pouze slova potřebná pro tuto kalkulačku. Syntaxe výrazů se snaží přiblížit co nejvíce přirozenému jazyku. Vzhledem k nepřesnosti rozpoznávače je zde však snaha o použití co nejmenšího počtu slov pro zadávání výrazů. Pokud uživatel při nahrávání zvuku dobře artikuluje, úspěšnost rozpoznání je většinou stoprocentní. Problémy s rozpoznáváním však nastávají, pokud je v pozadí hluk.

Kalkulačka pracuje s čísly v desítkové soustavě a umožňuje rozsah čísel v typu *double*. Kalkulačka umí pracovat s desetinnými čísly s přesností na devět desetinných míst. Zvládá operace sčítání, odčítání, násobení, dělení a mocniny, resp odmocniny. Umožňuje používání předdefinovaných proměnných a používání závorek. Umí též počítat goniometrické funkce a logaritmy o libovolném základu, přičemž funkce umožňují mít jako svoje parametry jiné funkce.

Kalkulačka byla testována sedmi uživateli, z nichž byli 3 muži a 4 ženy. Úspěšnost rozpoznání se pohybovala ze začátku kolem 70%. Poté, co se uživatelé začali snažit mluvit zřetelněji, se úspěšnost rozpoznání výrazně zlepšila. S možností ovládní hlasem byli všichni dotazovaní uživatelé spokojeni.

PODĚKOVÁNÍ

Chtěl bych poděkovat vedoucímu své bakalářské práce, panu Ing. Františku Grézlovi, Ph.D., za veškerou pomoc, kterou mi poskytl, a ochotu, s níž mi věnoval svůj čas. Tato práce vznikla za částečné podpory Výzkumného záměru (MSM0021630528).

REFERENCE

- [1] GOODHEAD, Paul. Bit-tech.net [online]. 2011-02-22 [cit. 2011-03-03]. Voice controlled games demonstrated. Dostupné z WWW: <<http://www.bit-tech.net/news/gaming/2011/02/22/plymouth-based-developer-sets-sights-on-voi/1>>.
- [2] HARDY, Hilda, et al. Data-driven strategies for an automated dialogue system. In SCOTT, Dornia. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Barcelona, Spain : Association for Computational Linguistics, 2004. s. 8. Dostupné z WWW: <<http://acl.ldc.upenn.edu/P/P04/P04-1010.pdf> >.
- [3] YOUNG, Steve, et al. The HTK Book [online]. [s.l.] : [s.n.], 1995 [cit. 2011-03-03]. Dostupné z WWW: <http://www.eee.bham.ac.uk/jancovic/teaching/EE4H_Nov2010/HTKManual.pdf >.
- [4] HONZÍK, Jan Maxmilián. Algoritmy [online]. [s.l.], 2007. 262 s. Studijní opora. FIT VUT v Brně.