TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky a mezioborových inženýrských studií

Studijní program: B2612 – Elektrotechnika a informatika
Studijní obor: 2612R011 – Elektronické informační a řídící systémy

Návrh interaktivního hlasového rozhraní mezi člověkem a počítačem

Designing an interactive voice interface for human-computer interaction

Bakalářská práce

Autor: **Tomáš Klucho**
Vedoucí práce: Ing. Miroslav Holada, Ph.D.
Konzultant: Ing. Josef Chaloupka, Ph.D.

V Liberci 16. 5. 2008

Prohlášení

Byl(a) jsem seznámen(a) s tím, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 o právu autorském, zejména § 60 (školní dílo).

Beru na vědomí, že TUL má právo na uzavření licenční smlouvy o užití mé bakalářské práce a prohlašuji, že **s o u h l a s í m** s případným užitím mé bakalářské práce (prodej, zapůjčení apod.).

Jsem si vědom(a) toho, že užít své bakalářské práce či poskytnout licenci k jejímu využití mohu jen se souhlasem TUL, která má právo ode mne požadovat přiměřený příspěvek na úhradu nákladů, vynaložených univerzitou na vytvoření díla (až do jejich skutečné výše).

Bakalářskou práci jsem vypracoval(a) samostatně s použitím uvedené literatury a na základě konzultací s vedoucím bakalářské práce a konzultantem.

Datum: ……

Podpis: ……

Děkuji svému vedoucímu bakalářské práce Ing. Miroslavu Holadovi, Ph.D. za užitečné konzultace, rady a připomínky k této práci a především pak za poskytnutí již hotového programu k rozpoznávacímu systému Dundis a také za jeho trpělivost.

Abstrakt

Tato práce se zabývá problematikou návrhu interaktivního hlasového rozhraní mezi člověkem a počítačem. V rámci této práce byl vytvořen program, který je založen na rozpoznávání izolovaných slov. Aplikace má dvě hlavní fáze. První fází je návrh klientského prostředí. Snahou je vhodná volba parametrů tak, aby prostředí bylo uživatelsky příjemné, přehledné a hlavně přínosné pro uživatele v oblasti hlasového ovládání chodu počítač. Druhá fáze se zabývá možností jednoduchého vytvoření slovníku pomocí pohodlného editoru, který nám umožní, abychom se vyvarovali špatně zadaných informací rozpoznávači, který by s námi poté nekvalitně komunikoval a nerozpoznával naše promluvy.

Pro účely kvalitního rozpoznávání řeči byl použit již hotový rozpoznávací systém, který pro aplikaci umožnil využívat Ústav informačních technologií a elektroniky, Technické univerzity v Liberci. V rámci této práce není řešen problém detekce mluvčího, hlas řečníka v různých situacích ani měnící se akustické pozadí.

Práce je rozdělena tematicky na 3 části, kdy v první je uveden teoretický přehled problematiky návrhu hlasového dialogového systému. Jsou zde uvedeny možné způsoby přístupu k návrhu hlasového dialogového systému a také způsob, jakým vliv řeči ovlivňuje návrh hlasového dialogového systému.

Další část je zaměřena na popis jednotlivých modelů hlasového dialogového systému, které musí být implementovány v softwarové části hlasového systému. Jedná se o metody rozpoznávání řeči, kde je stále velký problém rozpoznat promluvu jakéhokoliv řečníka užívajícího libovolná slova daného jazyka a dialogový přístup. Vhodná kombinace těchto metod nabízí velmi dobrý poměr mezi efektivitou a složitostí, danou náročností implementace.

Poslední část je zaměřena na vlastní návrh a realizaci interaktivního hlasového rozhraní mezi člověkem a počítačem. Zde jsou porovnávány parametry návrhu a cílem je posoudit řadu atributů, které určují výslednou efektivitu a uživatelsky příjemnou ovladatelnost aplikace. Těmito parametry jsou např. volba rozpoznávacího systému, výběr vhodného uživatele pro hlasové rozhraní nebo design aplikace.

Klíčová slova: interaktivní hlasové rozhraní, rozpoznání řeči, distribuovaný rozpoznávací systém, řízení dialogu, dialogový systém s konečným počtem stavů.

Abstract

This work deals with problematics of designing an interactive voice control interface between human and a computer. Within this work it has been created computer program, which is based on recognition of isolated words of it's user. Application has two main parts. The first part is design client environment. The goal is suitable choice of parameters. Environment was user niceness, overview representation and mainly boon for users in area voice control computer. The second part is possibility of creating a simple dictionary using a comfortable editor, which guarantees not to provide wrong information to the recognizer, which could lead difficult communicated and poor recognized of our speech.

For purposes high-quality speech recognition it was use complete recognition system, which for application enabled use Institute of Information technology and electronics, Technical university of Liberec. In work it doesn't deal with a problem of speaker detection, voice of speaker under different conditions or variable acoustic background.

The work is thematically divided into three sections, where the first covers theoretical survey of voice dialog system design problematics. It presents possible approaches of voice dialog system design and also the influence speech on voice dialog system design.

The second section is focused on description of particular parts of voice dialog system design, which has to be implemented in software part of the recognition system. It's mainly speech recognition methods, where recognizing any speech of random speaker using arbitrary words of given language, and dialog approach. Suitable combinations of these methods provide very good trade off between effectivity and complexity of implementation.

The last section is focused on personal design and realization interactive voice dialog system. The parameters are compared and the goal is to examine several attributes, which determine resulting effectivity and user friendly controllability of the application. These parameters are for example a choice of recognition system, choice of specific user for an interface or application design.

Key words: interactive voice control interface, speech recognition, distributed recognition system, dialog approach, dialog system with final number status.

Obsah

Úvod

Problematika hlasového dialogového systému si v poslední době získává stále větší pozornost, především ve snaze vytvořit hlasový dialogový systém, který by bylo možné využít ve více oblastech najednou. V současné době stav vědní disciplíny i současné komerční aplikace hlasových dialogových systémů předpokládají, že hlasový dialog je omezen na zvolenou aplikační oblast. Takové systémy by bylo možné použít např. v komplexních službách cestovních kanceláří, kdy by mohly být využity pro objednání letenek, rezervaci ubytování, půjčení aut v závislosti na počasí (auto s klimatizací nebo bez), předpovědi počasí na dané období. Další uplatnění může být např. komplexní informace o odjezdech a příjezdech vlaků a autobusů, kde je potřeba zadat místo odjezdu a příjezdu, den a čas odjezdu eventuálně příjezdu a hlasový dialogový systém vyhledá možná spojení. Dále je možné systémy využít např. pro objednávání vstupenek na kulturní a sportovní akce, rezervační systém autobusových, vlakových a leteckých spojení, nákup věcí po telefonu, menší bankovní operace a hlasové ovládání. Jedním z důvodů intenzivního rozvoje těchto hlasových dialogových systémů je právě snaha o vytvoření systému, který by bylo možné využít ve více oblastech najednou (informace o letu a rezervování ubytování). V praxi je zatím využití dialogu omezené na jednotlivé oblasti.

Dalším z důvodů výzkumu hlasového dialogového systému je i fakt, že je veřejně prospěšný pro uživatele, kteří mají oči i ruce plně zaměstnány. Uživatel je vzdálen od systému a může využít pouze hlasové komunikace, a pokud je uživatel zdravotně handicapovaná osoba s pohybovými nebo zrakovými obtížemi. Systémy hlasové syntézy mohou zprostředkovat zrakově postiženým občanům přístup k informacím, které by jim byly jinak nedostupné. Systém rozpoznávání řeči může být zase využit pohybově postiženými lidmi jako hlasem řízené zařízení k ovládání světel, televize, rádia apod.

1 Základní problematika v hlasovém dialogu

1.1 Obecný přístup k návrhu dialogového systému

Hlasové dialogové systémy umožňují uživatelům komunikovat prostřednictvím vlastního hlasu s počítačovými nebo internetovými aplikacemi. Hlavním účelem hlasových dialogových systémů je vytvořit interaktivní rozhraní mezi počítačem, který řídí vytvořenou aplikaci a uživatelem, který komunikuje hlasem. Těmto požadavkům vyhovuje celá řada systémů hlasového dialogu. Systémy je možné dělit od jednoduše komunikujících systémů, které využívají menšího souboru slov a řečník je izolovaně vyslovuje až po systémy, které umožňují vést komunikaci s uživatelem souvislou přirozenou řečí. I přestože byl v problematice hlasového dialogového systému učiněn obrovský pokrok, přesto je v současné době nemožné udělat konstrukci aplikace hlasového dialogového systému tak, aby bylo možné hlasem ovládat více aplikačních oblastí najednou. Hlasový dialogový systém je omezen na zvolenou aplikační oblast (např. dialogový systém pro objednávání vstupenek, rezervační systém autobusových spojení, nákup po telefonu apod.), nejde tedy o přirozenou a neomezenou komunikaci člověka s počítačem na libovolné téma.

Pro úspěšné vedení komunikace člověka s hlasovým dialogovým systémem je potřeba mít dobře navržené a kvalitní moduly **rozpoznávání řeči**, **porozumění mluvenému jazyku**, **syntézy řeči** a modul zajišťující **řízení dialogu**. Modul porozumění mluvenému jazyku by měl pracovat tak, aby interpretoval rozpoznané promluvy v kontextu s dalšími známými informacemi a podával potřebné informace k uskutečnění odpovídajících akcí.

Hlasové dialogové systémy je také možné dělit podle přístupu použitého při řízení dialogu. Tyto metody je možné dělit na **systémy s konečným počtem stavů**, **systémy založené na rámcích** a **systémy založených na agentech**. Strategie řízení dialogu má přitom přímou souvislost s tím, jak dialogový systém zpracovává vstupní řečovou informaci od uživatele a jak zachází s případnými chybami rozpoznávání a interpretace.

1.2 Vliv řeči v hlasovém dialogu

Ještě než je schválena navržená strategie pro aplikaci řečového vstupu nebo výstupu, v interaktivním styku s počítačem, je nutné uvážit důsledek využití řeči jako komunikačního média. Řeč jako prostředek přenosu informace v interakci člověka a počítače má některé výhody, ale také mnohé nevýhody.

Mezi nevýhody patří zejména to, že řečový výstup zajišťuje menší rychlost přenosu informace, než je rychlost přenosu vizuální informace. Schopnost rychle prohlédnout dokument je při použité řeči výrazně redukována, neboť řeč je sériová prezentace informace a je proto hůře využitelná například pro rychlé vyhledávání požadované informace. Detailní informace obsažená v promluveném textu je rychle zapomínána a potřebuje být obvykle několikrát obnovena, což je většinou časově náročný proces. Je to způsobeno sériovým a tudíž pomíjivým charakterem řeči a omezenou kapacitou krátkodobé paměti člověka. Vizuální obnovení informace je naproti tomu velmi rychlý proces, který zahrnuje několikanásobné prohlédnutí pouze těch částí dokumentu, kde je informace obsažena.

Další nevýhodou řečového výstupu v hlasových dialogových systémech je povaha řeči v odpovědích systému. V případě, že je odpověď v textové podobě na monitoru a uživatel si ji nepřeje vidět, může ji jednoduše ignorovat. Ovšem v případě, že je odpověď v podobě zvuku, je obtížné jednoduše tuto skutečnost ignorovat. To může zapříčinit, že není možné využít zvukové odpovědi například v pracovním prostředí, protože by to zapříčinilo rušení ostatních pracovníku. Možné odstranění tohoto problému je využitím sluchátek s mikrofonem.

Přes všechny tyto zmíněné nedostatky a omezení, která se objevují při vedení komunikace mezi člověkem a počítačem, existuje mnoho aplikačních oblastí, kde bude využití komunikace s počítačem prostřednictvím řeči efektivní. Jde hlavně o situace, kdy:

- Oči i ruce uživatele jsou plně zaměstnány jinými úkoly.
- Uživatel má potřebu být pohyblivý a jiná vstupně-výstupní zařízení jsou neefektivní.
- Uživatel je vzdálen od systému a může využit pouze hlasové komunikace přes běžné telefonní nebo rádiové spoje

- Uživatel je zdravotně handicapovaná osoba s pohybovými, popřípadě zrakovými obtížemi.

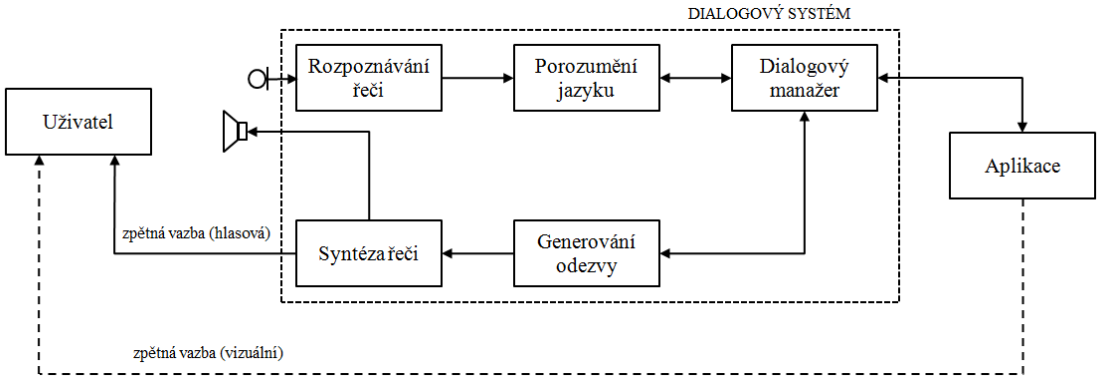
Pár důležitých rad a zásad, které by měly být respektovány při návrhu hlasového dialogového systému, aby výsledná aplikace byla pozitivně akceptována uživateli a stala se prospěšnou pro uživatele.

- Systém by měl být uživatelský příjemný a robustní ve způsobu interakce s uživatelem.
- Užití systému by mělo být přirozené a aktuálně prospěšné pro uživatele. Nemělo by být novinkou, která pouze vzbuzuje pozornost nebo dočasně zvyšuje prodej.
- Ideální systém by měl dovolit uživateli zeptat se na to, co potřebuje, jakýmkoli způsobem, kterým v dané situaci považuje za vhodný.
- Funkce rozpoznávání řeči musí být spolehlivá, tj. přesnost rozpoznání slov by měla být co nejvyšší, aby se dialog neprotahoval zbytečným opakováním a ověřováním nespolehlivě rozpoznaných promluv. Spolehlivý a rychlý přístup k informacím či rychlé a bezpečné vyřešení dané úlohy má pozitivní vliv na uživatele.
- Je velmi užitečné, když je uživatel hlasového dialogového systému vybaven včasnou a jednoznačnou odezvou systému, aby měl oprávněný pocit, že je součástí komunikačního procesu, nebo že se skutečně podílí na případných řídících akcích systému.

Výsledkem uplatnění lidského faktoru při návrhu hlasového dialogu by měla být spokojenost potenciálního uživatele s pohodlím a snadností obsluhy. Lidský faktor vstupuje do úlohy prostřednictvím rozumného návrhu na řízení dialogu, volbou přirozeného slovníku, využitím nápověd, časováním, užitím zpětné vazby k uživateli apod. I přes všechny tyto uvedené vlastnosti půjde vždy jen o tematicky omezenou komunikaci člověka s počítačem, která bude v návrhu řízení dialogu především soustředěna na úspěšné dokončení konkrétní úlohy a v žádném případě nebude mít snahu zajistit plný rozsah lidské komunikace. Takto konstruované dialogy jsou nazývány **praktické dialogy**.

2 Jednotlivé moduly hlasového dialogového systému

Na obrázku 2.1 je znázorněno blokové schéma pro tvorbu hlasového dialogového systému. Schéma se skládá z několika modulů, které zahrnují následující funkce – **rozpoznávání řeči, porozumění jazyku, generování odezvy** a také **syntézu řeči**. Hlavním a nejdůležitějším modulem je **dialogový manažer**, který řídí interaktivní strategii a také je zprostředkovatelem různých informací a znalostí získaných jinými moduly dialogového systému tak, aby byla úspěšně dokončena úloha, která je řešená dialogem. Ve většině případů aplikací je totiž velmi nepravděpodobné, aby uživatel při komunikaci s hlasovým systémem získal ihned potřebnou konečnou informaci pomocí jedné promluvy. Je to dáno tím, že dotaz uživatele může být nepřesný, neúplný, dvojznačný anebo také nenavazuje na předchozí promluvy uživatele a odpovědi hlasového systému. Ovšem i když je dotaz přesný, úplný, tak se může stát, že modul rozpoznávání řeči bude chybovat s nepřesností rozpoznání promluvy. Je proto potřeba, aby uživatel mohl v průběhu komunikace zajistit kompletaci a potvrzení všech správných informací, které vedou k úspěšnému dosažení cílů úlohy.



Obr. 2.1: Blokové schéma hlasového dialogového systému

2.1 Rozpoznávání řeči

Počítačové rozpoznávání mluvené řeči je předmětem zájmu výzkumných laboratoří již velkou řadu let. I přestože byl v rozpoznávání řeči učiněn obrovský pokrok, je nemožné udělat konstrukci zařízení, které by bylo bez problému schopno rozpoznat promluvu jakéhokoliv řečníka, který by vyslovoval libovolná slova, ještě poměrně vzdálenou budoucností. Hlavními důvody, které máji velký vliv na obtížnost rozpoznávání řeči, jsou typ řečníka, prostředí, ve kterém je promluva řečníka

rozpoznávaná, ale také i složitost úlohy. Stručnější popis jednotlivých důvodů, které mají vliv na rozpoznání řeči:

- Hlas jedné osoby se liší od hlasu jiných osob, protože každý uživatel má odlišné hlasové ústrojí a také artikulace je u všech uživatelů různá. Což je následkem, že každý člověk má jinou barvu hlasu, jiný přízvuk, odlišné tempo řeči apod. Systémy rozpoznávání řeči proto můžeme dělit na **systémy závislé na řečníku** a na **systémy nezávislé na řečníku**.
- Hlas jednoho řečníka se také mění v různých situacích. Promluva řečníka je jiná v případě, kdy člověk vysloví dotaz normálním hlasem, potichu, nahlas, šeptem apod. Proto je nemožné, aby jeden řečník pronesl jedno slovo stejným způsobem v různých situací. To je dáno tím, že v různých situacích má uživatel jiné časování vyslovení dotazu, změna frekvence hlasu je také jiná, tzv. **prozodie řeči**. V souvislé řeči se navíc projevuje také jev **koartikulace**, který může pozměnit fonetické vlastnosti začátku a konce slova v závislosti na kontextu okolních slov.
- Dalším důvodem pro obtížnost rozpoznání řeči je možnost různých prostředí. V promluvě uživatele se může projevit vliv okolního šumu (hlučné prostředí), nekvalita a rušení přenosového zařízení, které sníží kvalitu výrazu přeneseného do rozpoznávacího systému apod.
- Vliv na správnou funkci rozpoznávání řeči je způsob vyřčení promluvy. Je jasné, že **rozpoznávání izolovaných slov** (jednoslovných výrazu, povelů) z relativně malého slovníku bude určitě snazší, než **rozpoznávání diskrétního diktátu** (slova jsou vyslovována izolovaně s krátkou mezislovní pauzou), kdy slovník čítá tisíce slov. Nejobtížnější úlohou je **rozpoznávání souvislé řeči**, kdy slovník má desetitisíce slov.
- Dalším důležitým vlivem na správnou funkci rozpoznávání řeči je způsob promluvy. Zda jde o řeč **čtenou,** nebo zda jde o **spontánně pronášenou promluvu**. U čtené řeči, je větší šance a pravděpodobnost, že gramatická stavba vět bude vyhovovat spisovné češtině. Také by se zde neměli neprojevit tzv. **neřečové události** (hlasité váhání, nádechy apod.), protože je to čtené, tak by promluva měla být plynulá. U spontánně pronášené promluvy totiž řečník velmi často do promluvy vkládá, i když nevědomě, neřečové událostí. Také dochází, že se řečník při promluvě opakuje. Vysloví část slova či věty a poté se snaží o

jinou formulaci anebo začne mluvit o něčem jiném, aniž by předchozí myšlenku dokončil. Navíc dochází k vyslovování velkého množství hovorových slov a nespisovných gramatických vazeb, které také mají vliv na správnost rozpoznání promluvy.

Pro zmírnění a částečné odstranění těchto negativních vlivů při rozpoznávání promluvy uživatele, lze provést nějaké úkony pro zvýšení robustnosti:

- Při návrhu hlasového dialogového systému, kde je předpoklad hluku na pozadí nebo nekvalita a rušení přenosového zařízení (přenos po telefonní lince), lze navrhnout způsob potlačení nebo zmírnění vstupního šumu či vlivu zkreslení přenosového kanálu.
- Akustický model je potřeba trénovat z takových dat, která budou odpovídat prostředí, pro které je aplikace připravována.
- Krátké „nedořeky“ a nové zahajování komunikace lze zachytit modelem neřečových událostí.
- Jazykový model musí brát v potaz řešenou úlohu, ale také, že dialog bývá z většiny případů veden spontánní řečí, ve které dochází k velkému množství hovorových slov a nespisovných gramatických vazeb. Proto pro zvýšení robustnosti lze slova (spisovná i hovorová) ve slovníku brát jako výslovnostní varianty daného slova a snaha odhadu výskytu možných variant.
- Modul rozpoznávání řeči by měl při rozpoznávání promluvy uživatele využít **míry důvěry.** Jedná se o určení správnosti rozpoznání promluvy. Podle hodnoty míry důvěry je možné provést či zabránit akci. V případě zabránění akce, je nutná možnost znovu zadání výrazu, aby uživatel mohl získat požadované informace a dosáhl svého cíle.

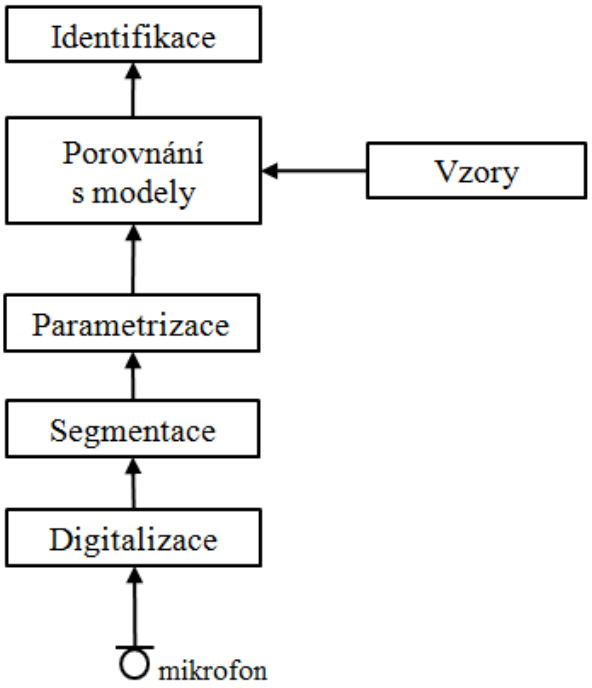
2.1.1 Prozodie ve spontánní řeči

U rozpoznávání vstupní spontánní promluvy řečníka hraje velkou roli prozodie promluvy. Jedná se o vliv základní frekvence hlasu uživatele nebo také o časování promluvy. Řečník může úmyslně nebo podvědomě ovládat frekvenci hlasu a tempo řeči, aby hlasem mohl vyjádřit svoje postoje a emoce. S prozodické analýzy se dá rozpoznat, o jaký typ promluvy se jedná – oznamovací, tázací, přikazovací charakter. Také se z analýzy dá určit pozornost uživatele, jeho postoj a emoce. Z výsledků analýzy prozodických charakteristik lze říci, že vyšší hodnota frekvence hlasu odpovídá spíše

typu dotazování, pokory či nejistoty. Kdežto nižší hodnota frekvence hlasu odpovídá typu autority, jistoty a panování. Vyhodnocení základního hlasivkového tónu a časování může být využito k určení konce věty či krátké promluvy.

2.1.2 Princip rozpoznávání řeči

Na obrázku 2.2 je znázorněno blokové schéma pro rozpoznávání izolovaných slov. Schéma se skládá z několika částí, kterými daná promluva v rozpoznávacím systému prochází. Těmito části jsou **digitalizace**, **segmentace**, **parametrizace**, **porovnání s modely** a **identifikace**.



Obr. 2.2: Blokové schéma rozpoznávání izolovaných slov

Vstupem do systému je promluva uživatele (analogový signál), která je přivedena na mikrofon, který je připojen k počítači. Analogový signál se moduluje do číslicové podoby (digitalizace), následně se číslicový signál rozdělí úseky o stejné délce (segmentace), tyto úseky se popíší přesnými parametry (příznaky vektoru) tak, aby se jednotlivé úseky od sebe lišily (parametrizace). Vektor příznaků je poté porovnáván se všemi modely jednotlivých slov (vzory zaznamenané ve slovníku) a slovo, které má nejdelší shodu s modelem slova, se stává rozpoznaným slovem (identifikace).

2.1.3 Metody rozpoznávání řeči

Metody pro rozpoznávání řeči lze dělit na ty, které pracují na principu **porovnávání se vzory** a na rozpoznávací systémy pracující s využitím **statistických metod**.

2.1.3.1 Princip porovnávání se vzory

Tato metoda se nejvíce vyvíjela v 70. a 80. letech, kdy byla hlavním modulem rozpoznání řeči pro izolovaně vyslovovaná slova. Slovo se zpracovává jako celek a jeho správné vyhodnocení se zajišťuje nejmenší vzdáleností k jeho vzoru. Rozhodující je určení správné vzdálenosti, která je určena pomocí metody dynamického programování. Tato metoda pracuje s efektem nelineární časové normalizace, kde kolísání v časové ose je modelováno časově nelineární „bortivou“ funkcí s přesně danými vlastnostmi. Časové rozdíly mezi dvěma řečovými obrazy jsou přitom eliminovány „borcením“ jedné z časových os tak, aby bylo dosaženo maximální shody s druhým obrazem.

2.1.3.2 Statistická metoda

Tato metoda rozpoznávání řeči je založena na statických metodách, kde jsou slova a celé promluvy modelovány pomocí tzv. skrytých Markovových modelů. Jednotlivá slova se mohou buď modelovat jako jeden celek jedním skrytým Markovovým modelem slova, nebo jsou ve většině případů vytvořeny skryté Markovovy modely subslovních jednotek (slabik, fonémů, trinomů apod.) a promluva je modelována zřetězením těchto subslovních jednotek. Každá subslovní jednotku má nastaveny parametry odpovídajícímu Markovovu modelu. Trénování jednotek je na základě trénovací množiny promluv.

2.2 Porozumění mluvenému jazyku

Modul rozpoznávání řeči pracuje s velkými slovníky a jeho úkolem a cílem je převést promluvu uživatele (řečový signál) na řetězec rozpoznaných slov. U modulu porozumění mluvenému jazyku je navíc potřeba zajistit porozumění nebo vhodnou interpretaci rozpoznaného řetězce slov a tím tak správně provést požadovanou akci.

Do modulu porozumění mluvenému jazyku vstupuje řetězec slov, který je výstupem modulu rozpoznávání řeči a nemusí být vždy úplně správně celý řetězec slov rozpoznán. Nepřesnosti rozpoznání řetězce slov je způsobeno tím, že rozpoznávací

modul rozpoznává spontánní řeč běžných uživatelů hlasového dialogového sytému. Více o rozpoznávání řeči v odstavci 2.1. Promluva také může obsahovat mnoho dvojznačností.

Modul porozumění mluvenému jazyku v hlasových dialogových systémech se skládá ze tří složek, které mají na porozumění mluvenému jazyku různou míry vlivu. Je to jednak formální **syntaktická analýza**, **reprezentace znalostí** a také **interpretace významu** rozpoznané promluvy.

Existuje několik typů znalostí, které do značné míry ovlivňují porozumění mluveného jazyka při komunikaci. Je proto nutné při nahrazování člověka počítačem v hlasovém dialogovém systému, aby modul měl co nejvíce znalostí v sobě zabudovaných. Jde zejména o:

- **Akusticko-fonetické znalosti**, které vyjadřují vztah mezi akustickými daty a fonetickým přepisem vyslovené promluvy. Jde tedy v širších souvislostech o problematiku akusticko-fonetického modelování.
- **Lexikální a fonologické znalosti**, jejichž úlohou je transformovat fonetickou mřížku na přijatelnou posloupnost slov.
- **Syntaktické znalosti** vyjadřující vztah symbolů, tj. slov, frází a vět, k sobě navzájem. Jsou nejčastěji reprezentovány vhodnou gramatikou.
- **Sémantické znalosti**, jež vyjadřují vztah symbolů k realitě. Jejich úkolem je odhalit takové vazby mezi slovy nebo frázemi, které pomohou stanovit smysl věty.
- **Pragmatické znalosti** obsahující znalosti účastníků o rozhovoru i o jeho historii, znalost o řešené úloze, včetně obecné znalosti o světě.

2.2.1 Formální syntaktická analýza

Formální syntaktická analýza řeší problematiku vzájemných vztahů a řazení symbolů. Na jednotlivé typy symbolů (fonémy ve slově, slova ve větě) můžeme nahlížet buď jako na posloupnost kvantitativních charakteristik (vyjádřeny číselnými příznaky získanými akustickou analýzou), anebo jako na posloupnost kvalitativní (nečíselný) charakter.

2.2.2 Reprezentace znalostí

Reprezentace znalostí je vhodně zvolený formalismus, který umožňuje jak znalosti uchovávat, tak na základě stávajících znalostí odvozovat znalosti nové. Z hlediska struktury modelů reprezentace znalostí jsou nejvíce využívány **logické** a **síťové modely**.

2.2.2.1 Logické modely

Základní myšlenka logického přístupu k reprezentaci znalosti vychází z předpokladu, že celá množina znalostí je vyjádřena souborem faktů nebo tvrzení převedených do formulí nějaké logiky, nejčastěji **predikované logiky prvního řádu**. Formule tak reprezentují znalosti, které je možné libovolně a podle potřeby přidávat či odebírat. Logické metody reprezentace znalostí jsou vybaveny účinným formalismem odvozovacích pravidel, který umožňuje na základě stávajících znalostí odvozovat znalosti nové.

2.2.2.2 Síťové modely

Základní idea síťového přístupu k reprezentaci znalostí vychází z předpokladu, že se problémová oblast zkoumá jako soubor objektů a vzájemných vztahů mezi nimi. Nejznámějšími síťový modely jsou **sémantické sítě** a **sémantické rámce**. Rozdíl mezi těmito modely je především v hloubce strukturovanosti souboru znalostí. Sémantické sítě jsou organizovány kolem významných jednotek až v průběhu řešení úlohy. Vychází z koncepce sémantických rámců a ještě obecněji koncepce **scénářů** při organizaci znalostí z vyšších struktur.

2.3 Generování odezvy

Generování odezvy je úkon, kdy je uživateli hlasového dialogového systému sdělována odpovídající odpověď na danou promluvu, která je získána z nějakého zdroje dat. Hlavním úkolem je určit, jak danou odpověď podat. V současné době se většinou využívá pro aplikace hlasových dialogových systémů konstrukce promluv založených na sémantických rámcích. Odpověď, která se má vytvořit pro uživatele, je v textové podobě, ve které jsou nevyplněné části, do kterých se vloží odpovídající informace, aby odpověď byla celistvá a měla všechny náležitosti, které si řečník v promluvě určil. Je také potřeba zajistit, aby vložené informace do odpovědi byly v gramatické shodě

s okolními slovy. Pokud je počet generových odezev konečný, je možné dané věty (odpovědi) předem namluvit a uložit je do slovníku řečových odpovědí systému. Výsledná odpověď je poté tvořena tak, že systém po rozpoznání slov začne vkládat informace do připravené věty tak, aby byla vytvořena odpovídající odpověď systému na požadavky uživatele. V případě, že je nějaká z volných pozic nevyplněna odpovídajícím a očekávaným způsobem, tak je nutné podání konkrétního dotazu na chybějící část, aby uživatel mohl doplnit potřebnou informaci.

Chytřejší a lépe propracované systémy generování odezvy mohou při vytváření odpovědi pro uživatele, také využít znalost o míře důvěry. Jedná se o určení správnosti rozpoznání promluvy. Podle hodnoty míry důvěry je možné provést či zabránit akci. Hodnota se mění v průběhu rozpoznávání dotazu uživatele.

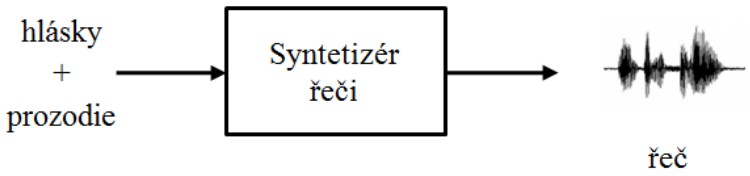
Dalším z důležitých parametrů při návrhu generování odezvy hlasového dialogového systému je využití diskursu. Jeho úkolem je zabezpečit, aby generovaná odezva byla v kontextu s průběhem vedení dialogu mezi uživatelem a hlasovým dialogovým systémem a také aby nemohlo dojít k dvojznačnému odkazu.

2.4 Syntéza řeči

Syntéza řeči je další z důležitých částí v problematice hlasového dialogového systému. Tato oblast je předmětem zájmu výzkumných laboratoří již velkou řadu let. Jde o úkon, kde je snaha uměle vytvořit řeč, která se co nejvíce podobá lidskému hlasu a aby komunikace mezi počítačem a uživatelem byla co nejvíce přirozená, která by odpovídala komunikaci mezi lidmi. Proč se lidé touto problematika tak intenzivně zabývají je z důvodu, že řeč je nejideálnější a nejpřirozenější způsob, jak hlasový dialogový systém může odpovídat řečníkovi.

Zařízení, které je potřebné k vytváření řeči, nazýváme **syntetizér řeči.** U každého modulu syntézy řeči je syntetizér řeči hlavním blokem pro konverzi textu na řeč. Výsledná odpověď je tvořena na základě vstupních informací. Těmi jsou fonetická a prozodická informace. Fonetická informace je dána posloupností fonémů. Prozodická informace udává průběhy prozodických charakteristik při vytváření řeči. Vstupní informace mohou být doplněny o direktivy či tagy, které ještě zvyšují kvalitu vytvářené řeči. Výsledkem syntézy řeči je vytvořit kvalitní řeč, aby se nelišila od řeči člověka.

Syntetická řeč by měla mít prozodické prvky, samozřejmostí je přirozenost a také by měla řečníka do jisté míry upoutat. Schéma typického syntetizéru řeči je na obrázku 2.3.



Obr 2.3: Zjednodušené schéma typického syntetizéru řeči

2.4.1 Základní přístupy

Základní přístup syntézy řeči se nejčastěji dělí podle způsobu modelování, které je využito k vytvoření výsledné řeči na tři typy: artikulační, formantovou a konkatenační syntézu.

- **Artikulační syntéza –** používá fyzikální model produkce řeči, který zahrnuje například jednotlivé artikulátory. Modeluje se celý systém vytváření řeči. Je nejobecnější metodou syntézy řeči. Jelikož se modeluje celý systém najednou, tak se jedná o složitější návrh syntézy řeči a proto není téměř využíván pro vytváření řeči.
- **Formantová syntéza -** je založena na teorii zdroje a filtru. Syntéza řeči se ve většině případů dělá pomocí sady pravidel, která převádějí fonetickou informaci na vstupu systému na posloupnost parametrů syntetizéru. Těmito parametry jsou formanty a jiné akustické či artikulační parametry. Tato metoda byla dlouhou dobu nejpoužívanější v oblasti syntézy řeči.
- **Konkatenační syntéza –** před syntézou řeči je potřeba mít vytvořen inventář řečových jednotek, z něhož se při vytváření výsledné syntézy řeči řečové jednotky vybírají. Jelikož tyto jednotky jsou uloženy v inventáři s určitým prozodickým charakterem, tak je potřeba výslednou odpověď systému prozodicky modifikovat, aby výsledná odpověď byla přirozená, srozumitelná a se správně zvolenými prozodickými vlastnostmi. Výsledná řeč se vytváří řetězením různých řečových jednotek a segmentů přirozené řeči. Tato metoda je v současné době nejpoužívanější v oblasti syntézy řeči.

U artikulační syntézy je cílem modelovat celý systém vytváření řeči najednou, kdežto zbylé typy se soustředí na přesnou imitaci hlasového ústrojí člověka, vlastní řečový signál a jeho modelování.

Přístup k syntéze řeči je také možno dělit i dalšími způsoby. Například podle míry zapojení řečníka a tomu odpovídající se zapojení do vývoje systému syntézy řeči. Tyto metody je možné dělit na syntézu podle pravidel a syntézu řízenou daty.

- **Syntéza podle pravidel** – parametry pro syntézu řeči je vytvářejí ručně uživatelem na základě sady manuálně odvozených pravidel. Tato metoda se považuje za formantovou analýzu.
- **Syntéza řízená daty** – parametry pro syntézu řeči je vytvářejí automaticky na základě řečových dat, ovšem pokud je inventář řečových jednotek vytvářen automaticky. Dříve se používalo ruční vytváření inventáře. Tato metoda se považuje za konkatenční syntézu.

V současné době se využívá automatická tvorba inventáře řečových jednotek na základě rozsáhlých řečových korpusů. Tento přístup syntézy řeči se nazývá **korpusově orientovaná syntéza.** Tato metoda se považuje za syntézu řízenou daty.

2.5 Řízení dialogu

Řízení dialogu je popsáno jako rozhodovací proces, který probíhá v prostoru stavů, akcí a strategií dialogu. **Stav dialogu** je všechna znalost, kterou systém získal vzájemným působením jednotlivých modulů hlasového dialogového systému s uživatelem a aplikací. Stav dialogu se mění v případě, že se provede **akce dialogu.** To je možné uskutečnit pomocí všech možných akčních zásahů, které může dialogový systém provést. **Strategie dialogu** určuje, jaká se provede příští akce, odpovídající danému stavu, v němž se dialogový systém nachází.

Dialogový manažer je zodpovědný za získání potřebné informace o stavu dialogové úlohy, výběru vhodné strategie a odpovídajících akcích dialogového systému. Úkolem dialogového manažera je organizovat vazby mezi jednotlivými moduly hlasového dialogového systému, využívat znalostní zdroje a zajistit komunikaci s uživatelem, aby bylo zaručeno cílové chování.

Dialogový manažer využívá různé **zdroje znalostí**. Zapojení těchto zdrojů do systému umožňuje správné rozpoznání promluvy. V případě nejasnosti při rozpoznávání se podílí na vytvoření dotazu, který žádá o opakování či doplnění informací.

2.5.1 Vedení dialogu

Na vedení dialogu lze nahlížet z různých hledisek. Jedním hlediskem může být způsob, jakým dialogový manažer zachází s promluvou uživatele, jak reaguje na nejasnosti v promluvě (chyba rozpoznání, nekvalitní kanál, vliv prostředí či špatná artikulace uživatele). Dalším hlediskem může být, jakým způsobem dialogový manažer ověřuje, že rozpoznávání a vyhodnocení promluvy je stejné s výrazem řečeným uživatelem.

2.5.1.1 Modelování zdrojů znalostí

Zdroje znalostní obsahují dvě části, které závisí na aplikační oblasti. První část je statická část modelu, která je navržena při přípravě dialogového systému. Druhá část je dynamická, která se vytváří po celou dobu dialogu mezi uživatelem a systémem.

MODEL PROSTŘEDÍ

Modelování prostředí je vhodné v případech, kde je dialog veden různými přenosovými kanály nebo v prostředích, kde se projevuje šum na pozadí. V případě, že rušení překročí určitou mez, kdy již není hlasový dialogový systém schopen rozpoznávat promluvu uživatele, bude muset dialogový manažer pozastavit dialog.

MODEL INTERAKCE SYSTÉMU

Modelování interakcí dialogového systému je vhodné v případech, kdy systém ví o všech funkcích a omezeních jednotlivých modulů systému. V průběhu dialogu by měl mít dialogový manažer informaci o tom, jaký modul je právě aktivní a která akce je právě prováděna. Takovýto model by také měl být schopen, v případě nutnosti, okamžitě vyřešit aktuální stav nebo celou úlohu.

MODEL UŽIVATELE

Model uživatele je vhodné využít v případě, kdy jsou známy informace o uživateli, které byly získány před začátkem dialogu, nebo v průběhu dialogu. Model

uživatele může být použit pro zlepšení efektivity vedení dialogu, protože umožňuje přizpůsobit dialogový systém individuálním požadavkům uživatele.

MODEL DIALOGU

Model dialogu je převážně založen na implicitně předpokládaném kooperativním chování jak uživatele, tak systému. **Soubor kooperativních principů**, který se skládá z tzv. konvenčních maxim, by měl uživatel dialogového systému v dialogu dodržovat. Jsou jimi:

- **kvantita** – uživatel se snaží podat co nejvíce informací (tolik, kolik je potřeba)
- **kvalita** – řečník podává pravdivé informace a nepodává informace, které nejsou podpořeny fakty
- **závažnost –** uživatel sděluje informace potřebné pro danou diskusi
- **způsob** – řečník se snaží podat promluvu co možná nejstručnější, jasně formuluje své cíle a vyhýbá se nesrozumitelným a mnohoznačným výpovědím

V případě nedodržení těchto konvenčních maxim je nutné uskutečnit několik výměn mezi uživatelem a systémem, aby se aktualizovali společné znalosti.

2.5.1.2 Ověření správnosti rozpoznání promluvy

Zapojení znalostních zdrojů do systému umožňuje vyrovnávat malé chyby vzniklé rozpoznáváním či neideální promluvou uživatele. Ovšem, aby systém provedl požadovanou akci, je nutné, mít potvrzení uživatele o správném rozpoznání promluvy, která se shoduje se záměrem uživatele. Existují dvě hlavní metody ověřování, že byl záměr správně rozpoznán – **explicitní** a **implicitní ověření.**

- **Explicitní ověření –** Využívá se otázek, na které uživatel odpovídá *ano/ne*. U této metody je problém s častým potvrzováním promluvy a dialog se prodlužuje. Při ověření celé promluvy najednou může, při jedné nebo více chybách, docházet ke komplikaci dalšího vedení dialogu.
- **Implicitní ověření –** Ověření správnosti rozpoznání se provádí tak, že následující otázka je upravena tak, že se v ní objeví ověření rozpoznané promluvy. To umožňuje uživateli ihned opravit špatně rozpoznanou promluvu. V případě, že se v odpovědi neobjeví žádná oprava, je to vnímáno za správně rozpoznané předchozí výrazy. Tato metoda ověření je rychlejší a přirozenější.

Ovšem je také důležité mít kvalitnější moduly rozpoznávání řeči a porozumění mluvenému jazyku.

2.5.2 Strategie řízení dialogu

Strategie dialogového systému určuje, jaká se provede akce, která odpovídá danému stavu, v němž se dialogový systém nachází. Každý stav úlohy má vlastní subdialog, který se snaží vyřešit daný dílčí cíl. Většinou se zaměřují na některou z následujících úloh – **potvrzení** (zjištění správnosti rozpoznané promluvy), **zotavení z chyby** (náprava chyby, co systém špatně rozpoznal), **opětovná pobídka** (systém neobdržel informaci), **dokončení** (zjištění chybějících informací), **omezení** (redukce rozsahu požadavku), **uvolnění** (zvětšení rozsahu požadavku), **zjednoznačnění** a **pozdrav/zakončení** (začátek a konce komunikace).

Strategie řízení dialogu je závislá na tom, který z účastníků dialogu má větší iniciativu. Těmito strategiemi jsou – **strategie s iniciativou systému, strategie se smíšenou iniciativou a strategie s iniciativou uživatele.**

2.5.2.1 Dialog s iniciativou systému

V dialogu s iniciativou systému zajišťuje systém řízení celého dialogu. Modul vybírá cíl i obsah, které musí být splněny, aby došlo k úspěšnému ukončení dialogu. U tohoto dialogu uživatel odpovídá na požadavek či položené otázky. Jednotlivé dotazy systému jsou uloženy ve slovníku a uživateli jsou nabízeny možné odpovědi formou nápovědy.

2.5.2.2 Dialog se smíšenou iniciativou

V dialogu se smíšenou iniciativou zajišťuje začátek dialogu systém, který podle získaných promluv od uživatele usoudí, kdy je možné přenechat iniciativu řečníkovi. Menší cíle dialogu mezi uživatelem a systém jsou vytvářeny podle odpovědí uživatele. Systém ovšem stále kontroluje, aby řešením menších cílů bylo dosaženo celkového cíle. V případě, že tomu tak není, přebírá systém řízení zpět do své režie.

2.5.2.3 Dialog s iniciativou uživatele

V dialogu s iniciativou uživatele zajišťuje řízení dialogu uživatel. Promluva řečníka se interpretuje bez omezení či stanovení celkového cíle. U této strategie vedení

dialogu může často docházet k přepínání témat hovoru. Tento poznatek klade vyšší nároky na kvalitu jednotlivých modulů hlasového dialogového systému. Také je potřeba mít systémy s větší podporou různých znalostních zdrojů.

2.5.3 Typy dialogových systémů

Jednotlivé typy dialogových systémů lze dělit podle způsobu zpracování promluvy uživatele. Dialogové systémy je možné rozdělit na tři skupiny:

- dialogové systémy s konečným počtem stavů
- dialogové systémy využívající strukturu rámců
- dialogové systémy založené na agentech

2.5.3.1 Dialogový systém s konečným počtem stavů

U dialogových systémů s konečným počtem stavů je struktura dialogu založena na tvaru stavově přechodové sítě. Každý uzel představuje konkrétní stav dialogu, ve kterém je možné získat informaci. Přechody mezi jednotlivými uzly sítě určují možné cesty sítí a tomu odpovídající povolené dialogy. Získání či potvrzení potřebné informace pro daný stav dialogu je uskutečněno prostřednictvím dílčích subdialogů. V dialogových systémech s konečným počtem stavů je většinou využívána strategie řízení s iniciativou systému. Pro tyto systémy jsou typické jednoduché nápovědy, které uživateli v každém stavu dialogu pomáhají vybírat, z jakých slov nebo krátkých frází má vytvořit svoji odpověď, aby úspěšně vyřešil dílčí úlohu a postupoval k dosažení celkového cíle.

Výhodou těchto dialogových systémů je jednoduchost dialogu. Možná odpověď uživatele v každém stavu je omezena souborem promluv (uloženy ve slovníku). Tento systém nemá velké nároky na jednotlivé moduly hlasového dialogového systému.

Nevýhodou těchto dialogových systémů je jejich malá pružnost. Problémy vznikají, když uživatel potřebuje změnit již vloženou položku. Uživatel také nemůže vložit více informací najednou, aby urychlil dosažení požadovaného cíle.

Dialogový systém s touto strukturou se využívá pro rozpoznávání izolovaných slov. Úlohy pro tento systém jsou závislé na malém počtu variant dalšího pokračování – jednoduché bankovní operace, informace o počasí, hlasové ovládání počítače apod.

2.5.3.2 Dialogové systémy využívající strukturu rámců

U dialogových systémů využívajících strukturu rámců je dialog konstruován tak, že je na základě analýzy řešené úlohy navržen rámec či struktura rámců, jejichž sloty představují jednotlivé parametry pro vyřešení úlohy. Tyto pozice musí být úplně vyplněny, aby mohla být vykonána požadovaná akce. Dialog je řízen dialogovým manažerem tak, aby uživatel mohl vložit informace v různém pořadí, a aby v promluvě mohl sdělit více informací najednou. Tímto se zásadně liší od systému s konečným počtem stavů.

Tento dialogový systém s rámcovou strukturou neomezuje uživatele na vyslovení pouze izolovaných slov nebo předem připravených frází, poskytuje větší pružnost vedení dialogu. Ovšem i tak je dialog stále omezen, protože záleží na výsledku analýzy předchozí promluvy uživatele. Tomu odpovídá, zda byly vyplněny volné části ve větě. Složitější komunikace mezi uživatelem a systémem, tak není možné vytvořit, protože systém mohou využívat i méně zkušení uživatelé, kteří nemají takové znalosti v oblasti hlasových dialogových systémů. Úspěšné dosažení cíle je v tomto případě také mnohem rychlejší, protože uživatel v promluvě může sdělit více informací najednou.

Nevýhodou tohoto systému jsou problémy s nejednoznačnostmi, s opravami již akceptovaných údajů, s klasifikací promluv, které modul rozpoznávání není schopen opakovaně zpracovat apod. Proto je užitečné, aby dialogový systém včas rozpoznal potíže se získáváním údajů od řečníka a přepnul strategii dialogu se smíšenou iniciativou na dialog s iniciativou systému a vhodně formulovanými a položenými otázkami došel s uživatelem k požadovanému cíli.

Složitější dialogové úlohy si občas vyžádají využití systému několika rámců, v případě vytvoření dialogového systému, který má nabídnout uživateli služby ve více oblastí (zajištění komplexních cestovních služeb). Nevýhodou je v tomto případě velké množství pravidel a kontextů.

2.5.3.3 Dialogové systémy založené na agentech

Základní myšlenkou vzniku dialogových systémů založených na agentech je snaha souběžné a nezávislé komunikace uživatele s více informačními zdroji. V případě tvorby hlasového dialogového systému se setkáme s obtížemi spojenými s přirozenou

nutností zadávat a odebírat všechny informace. Přináší to obtíže jak uživateli, tak i systému.

Ve snaze navrhnout systém pro komunikaci ve více oblastech najednou, narazíme na problémy spojené s růstem složitosti systému. Musí se provést rozšíření hlasového dialogového systému na novou oblast. Bude potřeba rozšířit slovník, natrénovat nové jazykové modely, rozšířit znalostní zdroje a také vytvořit nové scénáře pro generování promluv.

Efektivnější možnost návrhu systému na více oblastí je využití spolupracujících agentů. Tato možnost zajišťuje nezávislé vyvíjení jednotlivých agentů. K úspěšnému dosažení celkového cíle uživatele je potřeba spolupráce jednotlivých agentů. Úspěšná a účinná komunikace mezi agenty vede nejen k výměně dat, ale i znalostí.

Využití techniky spolupracujících agentů při návrhu hlasových dialogových systémů je vhodná v takových případech, kdy chceme, aby agent vyřešil danou úlohu komplexně a uživatel nemusel řešit postupně dílčí cíle, ale vše bylo vyřešeno najednou.

Technologie konstrukce inteligentních agentů je velmi intenzivně rozvíjena v rámci vědní disciplíny umělá inteligence. Při tvorbě multidoménového dialogového systému lze využít buď **centralizovaný model**, nebo **distribuovaný model**.

U informačního dialogového systému s centralizovaným modelem lze navrhnout soubor agentů, kteří vyhledají potřebné informace, zajišťují potřebné služby apod. V případě plánování dovolené lze navrhnout agenta zajišťujícího dopravu do destinace, agenta zajišťujícího ubytovací služby, agenta pro vypůjčení automobilu apod. Tento model je efektivní jen pro dialogy s menším počtem výměn.

U informačního dialogového systému s distribuovaným modelem je architektura založena na modelu klient–agent–server. V rámci této architektury je určen agent uživatelského rozhraní, který je připojen na další agenty systému. Agenti mluveného dialogu pak přistupují k příslušným databázovým serverům, aby zprostředkovali danou službu. Tyto systémy obvykle pracují se smíšenou strategií, proto je dobré mít kvalitní modul porozumění mluveného jazyku.

3. Návrh interaktivního rozhraní

Hlavním účelem hlasových dialogových systémů je vytvořit rozhraní mezi počítačem řízenou aplikací a uživatelem, který komunikuje hlasem. Při vytváření interaktivního hlasového rozhraní je nutné vhodně zvolit strategii návrhu. Je potřeba určit jednotlivé parametry pro návrh interaktivního hlasového rozhraní, které zaručí, aby aplikace byla uživatelsky příjemná, přehledná, byla přínosná pro uživatele a snadno se ovládala.

Je potřeba vybrat kvalitní rozpoznávací systém, který bude zvolen pro rozpoznávání promluvy, tj. přesnost rozpoznávání slov by měla být co nejvyšší, aby se dialog neprotahovat zbytečnými chybami v nerozpoznaných promluvách uživatelů. Dále je nutné rozlišovat pro jakou kategorie lidí a komunikační oblast témat (typy úloh) je návrh tvořen. Pokud se jedná o zkušené uživatele hlasových dialogových systémů, není třeba vkládat časté nápovědy. V opačném případě se nápovědy vloží tak, aby se s interaktivním rozhraním a hlasovým dialogovým systémem naučil nezkušený, či méně zkušený uživatel a nedocházelo ke špatným promluvám, které rozpoznávací systém není schopen rozpoznat. Tím je zabráněno nežádoucímu prodlužování dialogu. U tvorby interaktivního rozhraní také záleží na oblasti (typu úlohy), kde bude využívána. V současné době není možné vytvořit aplikaci, která by pomocí hlasového digitálního systému mohla ovládat více oblastí najednou.

S parametrem kategorie lidí souvisí další parametr pro návrh a tím je design aplikace. V případě tvorby rozhraní pro kategorii lidí ve věku mladých dětí a mladistvých je nutné, aby aplikace uživatele do určité míry zaujala (barevný vzhled, animované obrázky apod.) a byla uživatelsky příjemná. Výsledkem při návrhu interaktivního hlasového rozhraní by měla být spokojenost potenciálního uživatele s pohodlím a snadností obsluhy aplikace.

3.1 Rozpoznávací systém

Pro účely kvalitního rozpoznávání řeči byl použit již hotový rozpoznávací systém, který pro aplikaci umožnil využívat Ústav informačních technologií a elektroniky, Technické univerzity v Liberci. Rozpoznávací systém byl využíván pro rozpoznávání izolovaných slov či krátkých slovních spojení promluv řečníků. Měl by zajistit co možná nejvyšší přesnost rozpoznávání slov uživatelů, aby nedocházelo

k nežádoucímu protahování dialogu mezi uživatelem a interaktivním hlasovým rozhraním vlivem špatně rozpoznaných promluv v rozpoznávacím systému a tedy zároveň špatným odpovědím rozpoznávacího systému.

Pro rozpoznávací systém byla zvolena funkce nahrávání. Aby rozpoznávací systém spustil nahrávání, je potřeba provést úkon k dosažení zapnutí příjmu rozpoznávacího systému. V aplikaci je navrženo zapnutí nahrávání pomocí ovládací ikony a pohybu myši. Tato funkce zajistí, aby rozpoznávací systém nebyl trvale v režimu nahrávání a přijímání promluv od uživatele. V případě, že by tato funkce nebyla vytvořena, tak by se uživateli na monitoru počítače zobrazovala jedna odpověď systému za druhou, protože rozpoznávací systém by trvale přijímal a snažil se rozpoznávat dané promluvy.

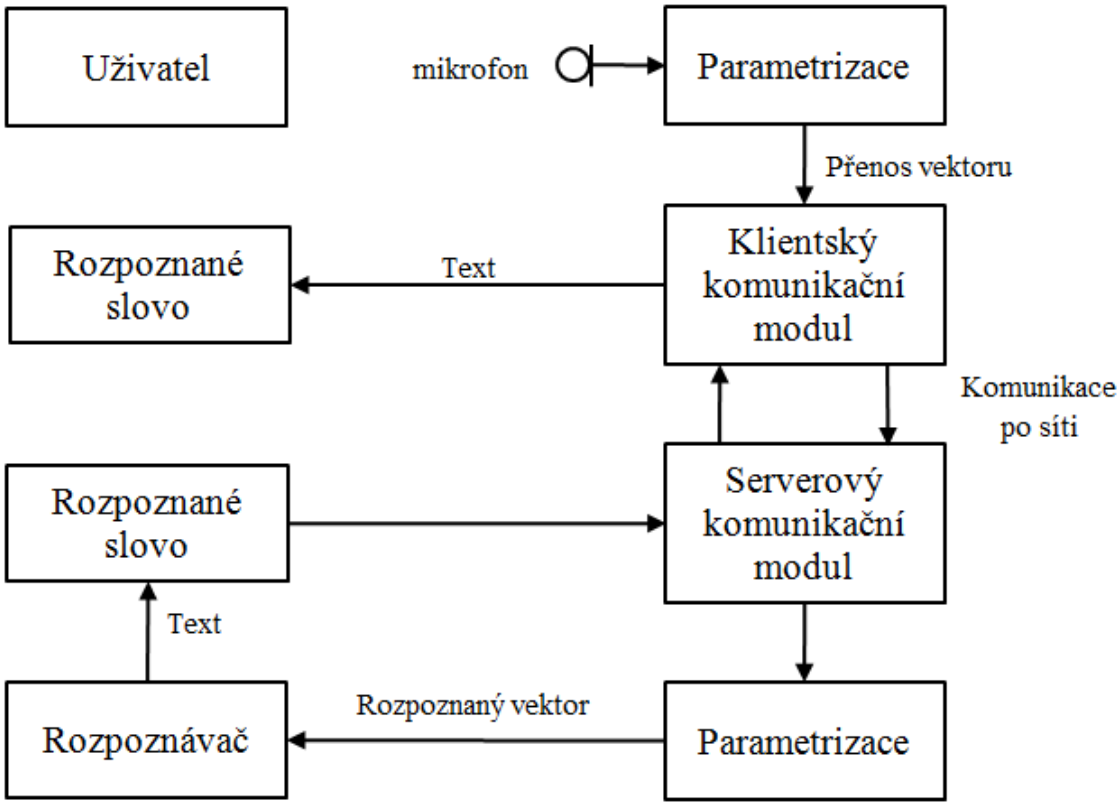
3.1.1 Princip rozpoznávacího systému

Rozpoznávací systém, který pro aplikaci umožnil využívat Ústav informačních technologií a elektroniky, Technické univerzity v Liberci, je vybaven řadou vhodných funkcí, které zabraňují nepřesnému rozpoznávání promluvy a tím zaručují rychlejší průběh komunikace. Těmito funkcemi jsou hlídání šumu a hluk (v případě nějakého šumu či hluku na pozadí a při zapnutém nahrávání rozpoznávací systém nevygeneruje nějakou náhodnou odpověď). Další funkcí je, že rozpoznávací systém má možnost pracovat se dvěma slovníky. Jeden slovník je nahrán v rozpoznávacím systému a uživatel nemá možnost ho měnit. Jeho kapacita slov se pohybu kolem statisíců slov a druhý slovník, který je v počítači uživatele, může čítat až dva tisíce uživatelských výrazu.

Princip rozpoznávání promluvy pomocí rozpoznávacího systému je následující. Uživatel vysloví promluvu na mikrofon, který je připojen k počítači. V případě, že je aktivováno nahrávání rozpoznávacího systému dojde k základní parametrizaci rozpoznávání dotazu a přenosu vektoru do klientského komunikačního modulu, který je v počítači uživatele. Ten přenese vektor do rozpoznávacího modulu pomocí internetové sítě. Rozpoznávací modul není v počítači uživatele a přistupuje se k němu vzdáleně pomocí internetové sítě. Promluva je opět parametrizována, tentokrát jinými algoritmy. Rozpoznávač vyhodnotí rozpoznávání promluvy uživatele pomocí vyhodnocení největší shody se vzorem uloženým ve slovníku. Rozpoznaná promluva má danou odpověď,

která odpovídá požadavku uživatele. Tato odpověď na rozpoznanou promluvu řečníka je poslána zpět uživateli a zobrazí se ve spuštěné aplikaci na počítači.

Struktura rozpoznávacího systému je znázorněna na obrázku 3.1.



Obr. 3.1: Struktura rozpoznávacího systému klient-server distribuovaného systému

3.2 Výběr uživatelů

Při vytváření interaktivního rozhraní je nutné uvážit pro jaké věkové, jazykové, oborově zaměřené kategorie lidí je aplikace interaktivního rozhraní vytvářena. Parametr věku lidí, aplikaci ovlivňuje hlavně v designu návrhu a také možnosti ovladatelnosti. S parametrem jazyku lidí je nutné počítat při vytváření aplikace, aby bylo možné pohodlné ovládání a nebyl problém, že uživatele nejsou schopni rozumět jinému jazyku (např. anglický, německý apod.). Menu aplikace a veškerý text, odpovědi rozpoznávacího systému a nápovědy musím být čitelné pro zvolenou skupinu lidí. Dalším parametrem je oblast, ve které bude aplikace využívána. Možnost oblastí je velká řada (např. informace o dopravě, malé bankovní operace, hlasové ovládání počítače apod.) a v současné době není možné vytvořit aplikaci hlasových dialogových systémů, která by mohla ovládat více oblastí najednou.

Jedním z důležitých parametrů pro návrh aplikace je výběr oblasti lidí z pohledu, jak jsou uživatelé schopní využívat a pracovat s interaktivním rozhraním aplikace a s dialogovým systémem. Uživatelé, kteří pravidelně používají dialogový systém, nepotřebují různé nápovědy, které by dialog zbytečně prodlužovaly a proto je snaha dialog co nejrychleji ukončit. Podobně je třeba předpokládat, že k dialogu přistoupí i uživatelé, kteří mají malé anebo žádné zkušenosti s danou aplikací, anebo nemají žádnou představu o tom, jaké jsou možnosti současných počítačů. Pro takovéto typy uživatelů je potřeba důležitá formulace nápověd, aby v průběhu dialogu postupně upřesňovaly pravděpodobný cíl komunikace nezkušeného uživatele se systémem.

Při tvorbě aplikace byly tyto parametry pro návrh zohledněny. Jednotlivé parametry byly nastaveny na:

- *věková kategorie lidí* – dospělí
- *jazyková kategorie lidí* – český jazyk
- *obor témat* – hlasové ovládání počítače
- *schopnost uživatelů se systémem* – středně zkušení lidé

Dané hodnoty byly při návrhu aplikace zohledněny. Tyto parametry by měli vést k tomu, aby se jednalo o pohodlně ovladatelnou aplikaci pro zvolenou věkovou kategorii lidí a měla by se stát prospěšnou pro uživatele ve zvolené oblasti témat.

3.3 Design aplikace

Volba designu aplikace je jednou z částí, jak zaujmout uživatele pro využívání interaktivního hlasového rozhraní. Vzhled aplikace působí na první dojem uživatele a podle toho se odvíjí další pozitivní či negativní vnímání. Dalším vlivem působení na uživatele může být snadnost a přirozenost ovladatelnosti. V případě špatného návrhu designu a ovladatelnosti určitě uživatele nezaujme a on raději zůstane buď u svého původního, nebo bude vyhledávat nějaké jiné interaktivní rozhraní. V opačném případě je vidět, že dané parametry designu aplikace byly zvoleny správně.

V případě tvorby rozhraní pro děti či mladistvé je nutné, aby aplikace uživatele do určité míry zaujala svým vzhledem. Pro děti by základní panel mohl mít tvar a pozadí animované postavičky či něčeho co děti zaujme. Pro mladistvé je dobré zvolit pozadí odpovídající věku a pro dospělé je vhodné upoutání zajímavým barevným vzhledem. V případě mladých dětí je také potřeba, aby ovladatelnost aplikace byla velmi jednoduchá, protože se jedná o nové nezkušené uživatele interaktivních

hlasových rozhraní, kteří si například ještě nemohou přečíst a využívat nápovědy. V jiných věkových kategoriích by ovladatelnost a obsáhlost aplikace mohla být větší, ale nesmí to být na úkor uživatelsky příjemného ovládání. Je tu také možnost využití nápovědy, která může značným způsobem ovlivnit a pomoci uživateli práci s interaktivním rozhraním pro správné a rychlé dosažení požadovaných cílů.

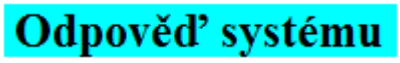
Design a ovladatelnost aplikace byly navrženy převážně pro dospělé lidi (základní panel má jednoduchý barevný vzhled), kteří se již s prací interaktivního rozhraní alespoň částečně seznámili. Je možné využít doplňujících a pomocných nápověd pro snadnou práci a pohodlné, rychlé dosažení cílů promluvy.

Design celé aplikace je možné vidět na obrázku 3.2, 3.3 a 3.4.

Obr. 3.2: Hlavní panel aplikace v barvě červeno-modro-bílé

a) b)

Obr. 3.3: Ovládací ikona pro rozpoznávací systém
a) režim nenahrávání, **b)** režim nahrávání



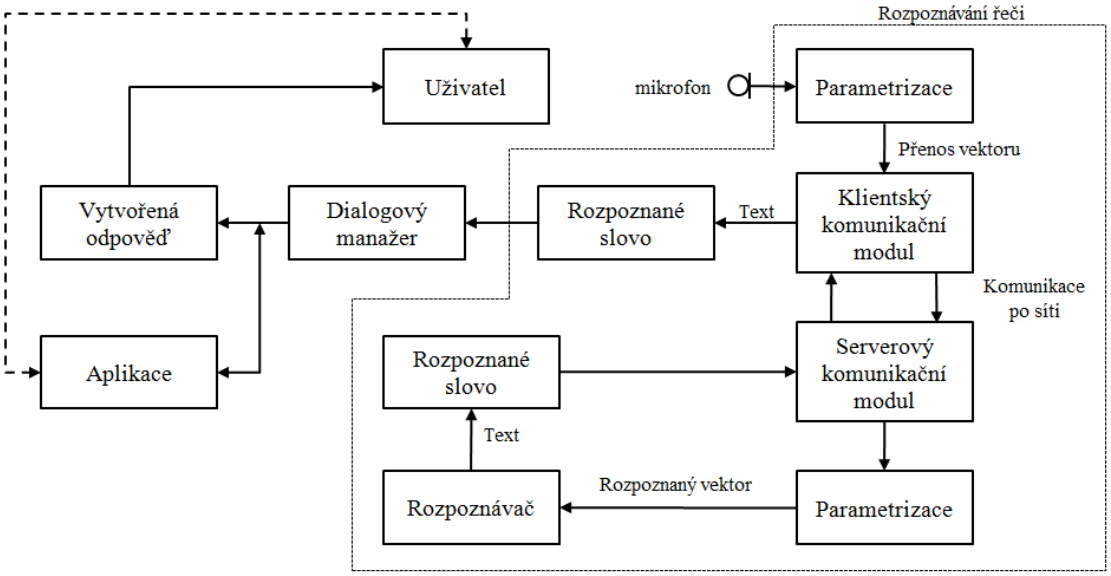
Obr. 3.4: Text pro odpověď rozpoznávacího systému

3.4 Struktura hlasového systému

Při zvolené strategii návrhu interaktivního hlasového rozhraní je také potřeba zvolit vyhovující strukturu pro kompletní hlasový systém. Nejdůležitější částí je uživatel. Jeho možností je pracovat přímo s aplikací interaktivního rozhraní anebo pomocí mikrofonu vytvářet promluvy pro rozpoznávací systém, který výraz přijímá a porovnává ho se slovy zaznamenanými ve slovníku. Správné rozpoznání promluvy je pomocí vyhodnocení největší shody výrazu se slovem ze slovníku. Po rozpoznání je

vytvořena odpověď pro uživatele, která je pro daný výraz jednoznačně daná a zaznamená ve slovníku pro rozpoznávání promluvy. Jedná se o odpověď buď v podobě textu zobrazeného na monitoru, nebo spuštěním odpovídající aplikace.

Bloková struktura hlasového systému je vidět na obrázku 3.5.



Obr. 3.5: Struktura hlasového systému

4 Realizace interaktivního hlasového rozhraní

4.1 Programovaní interaktivního hlasového rozhraní

Strategie návrhu interaktivního hlasového rozhraní byla realizována v prostředí Microsoft Visual Studio C++ .NET verze 2005. V současné době se jedná o jeden z nejpoužívanějších programovacích jazyků. Protože má vlastní zkušenost s programovacím jazykem v začátku vytváření bakalářské práce nebyla velká, bylo postupováno od studií daného jazyku, jeho struktury, zápisu, přes vytváření menších lehčích aplikací, které by bylo možné využít pro vložení do samotné aplikace až po samotné programování navržené strategie aplikace interaktivního hlasového rozhraní.

Jednou z menších lehčích aplikací byl naprogramován Editor slovníku, který je vhodný k využití pro hlavní aplikaci. V této aplikace bylo naprogramováno vkládání slov „Co se má psát“ a „Co se má říkat“ do seznamových rámečků a možnost zvýraznění si výrazů v „Co se má psát“ a tomu odpovídající výraz v „Co se má říkat“ a naopak. Byla vytvořena funkce pro správné zapsání cesty internetové stránky, aby rozpoznávací systém při uživatelově promluvě správně rozpoznal promluvu a správně provedl úkon otevření internetové stránky. Dalšími funkcemi bylo načítání již vytvořených slovníků z libovolného místa na disku do připravených seznamových rámečků nebo uložení seznamových rámečků do uživatelsky zvoleného souboru.

V programování bylo pokračováno navrženou strategií interaktivního hlasového rozhraní s využitím naprogramované aplikace editor slovníku. Rozpoznávací systém nebylo potřeba programovat, protože byl využit již hotový systém, který pro aplikaci umožnil využívat Ústav informačních technologií a elektroniky, Technické univerzity v Liberci. Bylo potřeba naprogramovat, jak bude vypadat odpověď rozpoznávacího systému na promluvu uživatele, co se má vypsat do textového pole odpovědi rozpoznávacího systému, eventuálně co a jakým způsobem se provede.

Aby rozpoznávací systém zahájil nahrávání je potřeba najet myší na nahrávací ikonu po dobu alespoň 0,5 vteřiny. Nahrávání je znázorněno zeleným orámováním ikony. V režimu nenahrávání rozpoznávacího sytému je orámování červené. Po spuštění nahrávání rozpoznávacího systému se automaticky za 3 vteřiny nahrávání vypne, aby nedošlo k trvalému sledování promluvy uživatele v případě, že nebyla uvolněna

nahrávací ikona. Pro nové nahrávání je potřeba opět znovu na ikonu najet. Při rychlém přejetí ikony se nahrávání nespustí. Nahrávací ikonu má také naprogramovány funkci posunutí (pomocí levého tlačítka myši) a rozbalovací menu (pomocí pravého tlačítka myši), kde je možné například vybrat velikost ikony a její průhlednost.

Do hlavního panelu bylo vloženo a naprogramováno menu, ze kterého je pohodlný přístup ke všem možným funkcím ovládaným z hlavního panelu. Všechny funkce hlavního panelu, které bylo potřeba vhodný způsobem naprogramovat, aby aplikace působila na první dojem uživatele, byla uživatelsky příjemná, jsou: načtení slovníku (možnost načíst vlastní slovník do rozpoznávacího systému), zahájení a ukončení komunikace (zahájení a ukončení nahrávání rozpoznávacího systému), editor slovníku, přehled slovníků (stručný přehled všech vytvořených slovníku s krátkými popisky pro lepší orientaci), výběr prohlížeče (volba prohlížeče pro zobrazování internetových stránek). Do editoru slovníku byly vloženy další vhodné funkce, které mají doplnit funkce již vytvořené. Úkolem nových funkcí je dopomoci uživateli k vhodnému a jednoduchému vkládání do nově tvořeného slovníku. Správná struktura jednotlivých výrazů (jednoduchých a složitých – vestavěné příkazy exe, www a file) umožňuje, aby rozpoznávací systém při uživatelově promluvě správně rozpoznal výraz a vykonal odpovídající a správnou odpověď. To umožní, aby se uživatel rychle a spolehlivě dostal k očekávanému cíly promluvy. Veškeré funkce editoru slovníku, které bylo potřeba vhodný způsobem naprogramovat, jsou: nový slovník, otevřít slovník, připojení slovníku, uložit, uložit jako, přidat slovo, přidat WWW stránku, přidat aplikaci, přidat prohlížeč a odpovídajícím souborem, úprava výrazu v seznamovém rámečku „Co se má říkat“ nebo změnu výrazů v obou seznamových rámečcích, automatické vytvoření telefonního seznamu pracovníků TUL, přehození slov u „Co se má říkat“ a funkce, které se automaticky přidávají do všech vytvářených slovníku při ukládání. Těmito funkcemi jsou: přidat stránku (rychlé vložení WWW stránky do slovníku), zobrazení slovníku a editace slovníku.

Pro aplikaci interaktivního hlasového rozhraní byla naprogramována nápověda, která slouží pro uživatele v případě výskytu nějakých nejasností při práci s aplikací. Je nezbytnou součástí programu, aby se uživatel mohl zorientovat v navršené aplikaci a nebránilo to rychlému postupu uživatele dosáhnout požadovaného cíle. Uživatelé, kteří pravidelně používají dialogový systém, nepotřebují tak obsáhlé nápovědy. Podobně je třeba předpokládat, že k programu přistoupí i uživatelé, kteří mají malé nebo žádné

zkušenosti s danou aplikací, nebo jinými interaktivními hlasovými rozhraními. Pro různé typy uživatelů je důležitá správná formulace nápověd, které jsou pomocníkem při učením se s novou aplikací.

4.1.1 Ukázka stěžejních zdrojových kódů

Ukázka stěžejního zdrojového kódu programu pro spuštění libovolné aplikace (*.exe), která se nachází ve slovníku, který je nahrán do rozpoznávacího systému, je uvedena níže. Při zapnutém nahrávání rozpoznávacího systému a rozpoznání promluvy uživatele vybere systém nejlepšího kandidáta promluvy z příslušného slovníku. Poté program pokračuje výběrem vhodné předpony (v našem případě exe, další možné jsou www nebo file), která určuje co se dalšího bude provádět. Po rozdělení celého řádku slovníku do dvou proměnných je vykonáno spuštění aplikace pomocí příkazu Process::Start(parST, parND); a zároveň je do textu pro odpověď rozpoznávacího systému napsána cesta k danému spouštěcímu souboru.

```
if (Nahravani->Record) // pokud je nahrávání rozpoznávacího systému zapnuté
{
    System::String ^t = gcnew String(this->rozpoz->voc.SayToWrite((char *) this->rozpoz->bestCandidate->word));
    if (t->Length > 4) // v případě, že je do proměnné t zapsáno slovo delší jak 4 znaky
    {
        if ((t[0] == 101) && (t[1] == 120) && (t[2] == 101) && (t[3] == 40)) // rozpoznání předpony exe pro
        {                                                                    spuštění aplikace
            int i;
            char *t1 = this->rozpoz->voc.SayToWrite((char *) this->rozpoz->bestCandidate->word);
            for (i = 0; i < (int) strlen(t1); i++)
                if (t1[i] == 40)
                    break;
            char t2[1000];
            strcpy(t2, t1 + i + 1);
            for (i = 0; i < (int) strlen(t2); i++)
                if (t2[i] == 41)
                    break;
            t2[i] = 0;
            System::String^ parND = "";
            System::String^ parST = gcnew String(t2);
            Process::Start(parST, parND); // spuštění aplikace
            toolStripStatusLabel2->Text = parST;
            VypisAkci->listBox1->Items->Add(parST);
            Nahravani->label1->Visible = true;
```

```
            Nahravani->label1->Text = parST;
            timer1->Enabled = true;
            delete parST; delete parND;
        }
    }
}
```

Ukázka stěžejního zdrojového kódu programu pro načtení jednotlivých položek do seznamových rámečků je uvedena níže. Hlavním příkazem pro správné zajištění načtení položek do slovníku je příkaz t = a.RNI(); jedná se o přečtení nové položky (Read Next Item) ze slovníku a automaticky je v této funkci zajištěna kontrola, zda se výraz ve slovníku vyskytuje v uvozovkách (jednoduchý výraz) nebo je testován jako vestavěný příkaz – exe, www a file (složený výraz). Promluva uživatele a odpověď systému je uložena do proměnných. „Co se má říkat“ tedy výraz uživatele je testován se všemi vloženými výrazy v seznamu rámečku „Co se má říkat“, aby nedošlo k vložení stejného výrazu promluvy uživatele pro více možný odpovědí systému. Zvyšuje to robustnost aplikace, aby nedocházeno k nežádoucím odpovědím rozpoznávacího systému.

```
soubor = openFileDialog2->FileName;
private: System::Void NacitaniRadku()
{
    String^ t1;
    char souborznak[256];
    sprintf(souborznak, "%s", soubor->ToCharArray());
    CTextIO_voc07 a;
    char *t;
    int x = a.Load(souborznak);
    do
    {
        t = a.RNI();
        if(t)
        {
            t1 = gcnew String(t);
        }
        t = a.RNI();
        if(t)
        {
            System::String ^t2 = gcnew String(t);
            if (listBox2->Items->IndexOf(t2) == -1)
            {
```

```
                listBox1->Items->Add(t1);
                listBox2->Items->Add(t2);
                delete t1; delete t2;
            }
        }
    }
    while(t);
}
```

4.2 Aplikace interaktivního hlasového rozhraní

Výsledná aplikace interaktivního rozhraní byla naprogramována pro rozpoznávání izolovaných slov či krátkých spojení slov k hlasovému ovládání počítače. V programu je možné, vytvořit si jednoduchým způsobem vlastní slovník s vlastními výrazy pro ovládání a spouštění všech možných aplikací a souborů na počítači. Jedná se o přínosného pomocníka při práci na počítači. Je zde poskytnuta řada funkcí, které by měli být prospěšné uživateli a měli by pomoci pro rychlejší a snadnější dosažení cílů. Zároveň bylo dosaženo uživatelsky příjemného prostředí.

Jednou z nejdůležitějších částí aplikace je rozpoznávání promluvy uživatele, které zajišťuje rozpoznávací systém. Pro aplikaci jej umožnil využívat Ústav informačních technologií a elektroniky, Technické univerzity v Liberci. Dalšími funkcemi je editor slovníku, který umožňuje jednoduchým způsobem si vytvořit uživatelský slovník pro kompletní a pohodlnou ovladatelnost počítače. Po nahrání do rozpoznávacího systému uživatel již ovládá zvolenými příkazy celý svůj počítač. Editor má v sobě zahrnuty další funkce (viditelné i skryté). Přehled vytvořených slovníků nám poskytuje informace o všech možných slovnících, které byly dosud vytvořeny, i s krátkými popisky pro rychlou orientaci mezi nimi. Jedna z dalších funkcí je možnost výběru prohlížeče internetových stránek. Protože každý uživatel dává přednost jinému prohlížeči internetových stránek, ať už podle oblíbenosti, zkušenosti s ovládáním, či vyšším stupněm zabezpečení při vstupu do internetové sítě. Je možné si vybrat ze dvou přednastavených internetových prohlížečů (Internet Explorer a Mozilla Firefox), nebo si zvolit vlastní prohlížeč vybráním cesty k danému spouštěcímu souboru pro vámi vybraný prohlížeč internetových stránek.

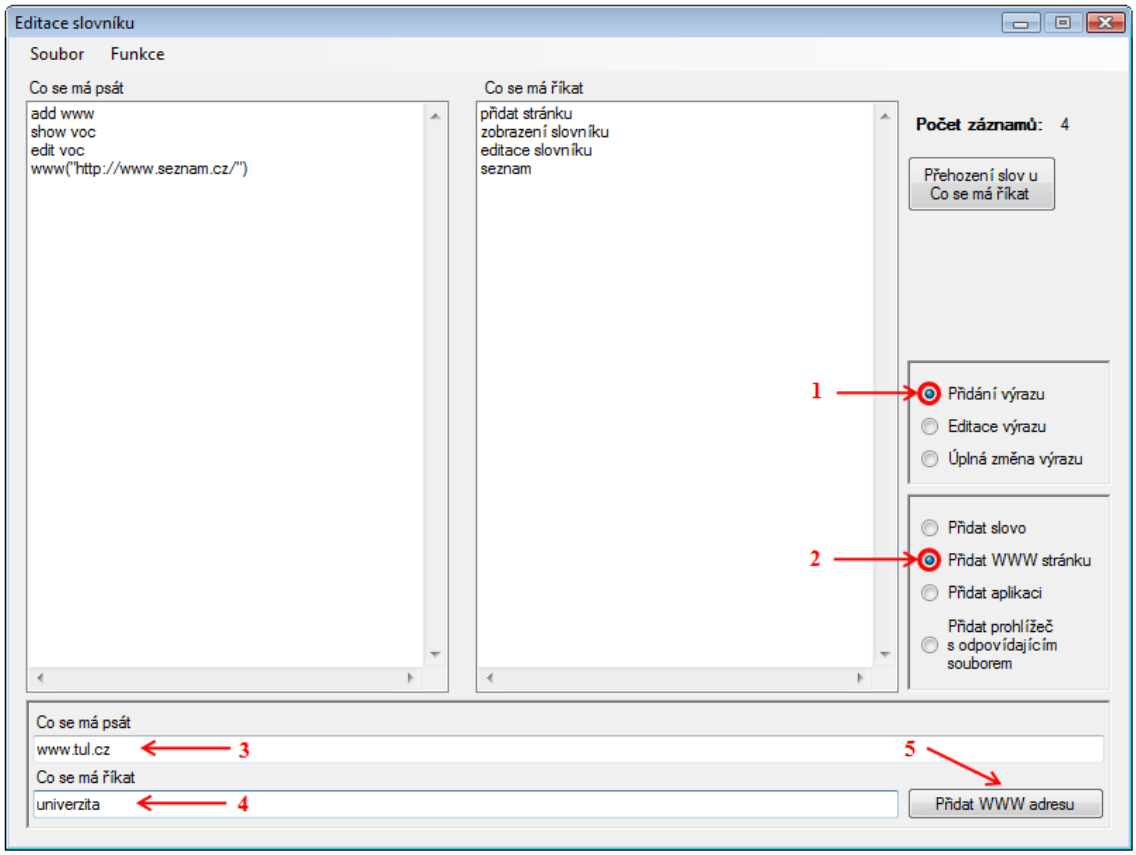
4.2.1 Editor slovníku

Druhý významný bod aplikace je vlastní editor slovníku. S touto pomocí má uživatel poměrně značně ulehčenou práci při tvorbě vlastního slovníku s vlastními výrazy a příkazy. Jedná se uživatelsky příjemné a přehledné prostředí, které je velmi prospěšné uživateli k ušetření času. Je zde také zajištěna správná struktura slovníku (jednotlivých výrazů a složených výrazů), aby po nahrání do rozpoznávacího systému nedošlo k chybě při rozpoznávání, jako by se tomu mohlo stát při ruční tvorbě slovníku. Kromě ušetřeného času uživatele při tvorbě, je zde zaručena i rychlá odpověď rozpoznávacího systému a tedy rychlé a přesné dosažení žádaného cíle.

Editor slovníku má pro snadné ovládání a tvorbu slovníku řadu viditelných a pár skrytých funkcí. Viditelné funkce jsou:

- **přidat slovo** – tato funkce umožňuje přidání jednoduchého výrazu. Odpovědí rozpoznávacího systému je textová informace podle dané promluvy uživatele (např. snadné a rychlé zjišťování telefonních čísel pracovníku TUL).
- **přidat WWW stránku** – tato funkce umožňuje přidání internetové stránky do slovníku (obr 4.1). Odpovědí rozpoznávacího systému je zobrazení dané internetové stránky v závislosti na rozpoznané promluvě uživatele v odpovídajícím prohlížeči a také výpis v textové poli dané internetové stránky.

  Postup: 1 … vybrat možnost Přidání výrazu
  2 … vybrat možnost Přidat WWW stránku
  3 … zadat název internetové stránky (tvar: www.server.domena)
  4 … zadat promluvu uživatele k dosažení cíle
  5 … použít tlačítko Přidat WWW stránku k vložení do slovníku

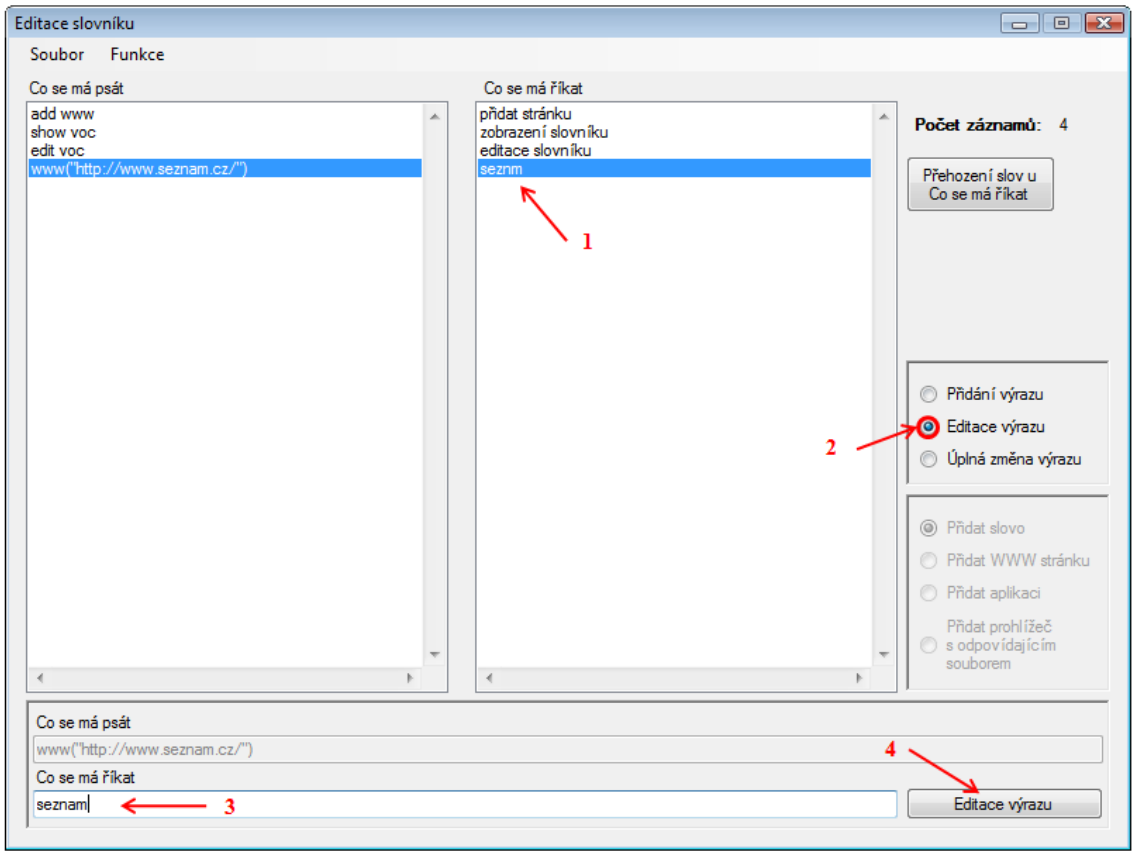


Obr. 4.1: Přidání WWW stránky do editoru slovníku

- **přidat aplikaci** – tato funkce umožňuje přidání jakékoliv spouštěcí aplikace (*.exe) do slovníku. Odpovědí rozpoznávacího systému na základě promluvy je spuštění konkrétní aplikace a také výpis v textovém poli, který určuje cestu ke spouštěcímu souboru.
- **přidat prohlížeč s odpovídajícím souborem** – tato funkce umožňuje přidání jakékoliv spouštěcí aplikace (*.exe) a jemu odpovídajícímu souboru pro spuštění v dané aplikaci do slovníku (např. C:\Program Files\Microsoft Office\Office12\WINWORD.EXE, D:\Moje\Bakalářka\Bakalářka\Bakalářka.docx). Odpovědí rozpoznávacího systému podle rozpoznané promluvy uživatele je spuštění konkrétní aplikace (v našem případě Microsoft Office Word) a souboru, který se v dané aplikaci otevře (v našem případě dokument bakalářské práce). Také do textového pole se vypíše cesta ke spouštěcímu a otevíranému souboru.
- **úprava či změna vložených výrazů** - tato funkce má dvě možnosti. První možností je pouhá změna výrazu v seznamovém rámečku „Co se má říkat“ a

odpověď rozpoznávacího systému zůstává stejná (obr. 4.2) a druhá je kompletní změna výrazů v obou seznamových rámečcích.

Postup: 1 … vybrat položku, která se má upravit (špatná promluva uživatele)
2 … vybrat možnost Editace výrazu
3 … zadat novou správnou promluvu uživatele
4 … použít tlačítko Editace výrazu k návratu výrazu do slovníku



Obr 4.2: Změna výrazu v editoru slovníku

- **automatické vytvoření telefonního seznamu pracovníků TUL** – tato funkce umožní automatické vytvoření telefonního seznamu všech pracovníků na Technické univerzitě v Liberci. Celý telefonní seznam je k dostání na internetové stránce http://telefon.tul.cz/ odkud je nakopírován a uložen do textového souboru. Telefonní seznam má následující sloupcovou strukturu. První sloupec je na všechny dosažené tituly, ve druhém sloupci je příjmení a jméno, ve třetím sloupci je telefonní číslo na dané pracoviště a v posledním sloupci je místo pracoviště. Mezi jednotlivými sloupci je mezera vytvořená

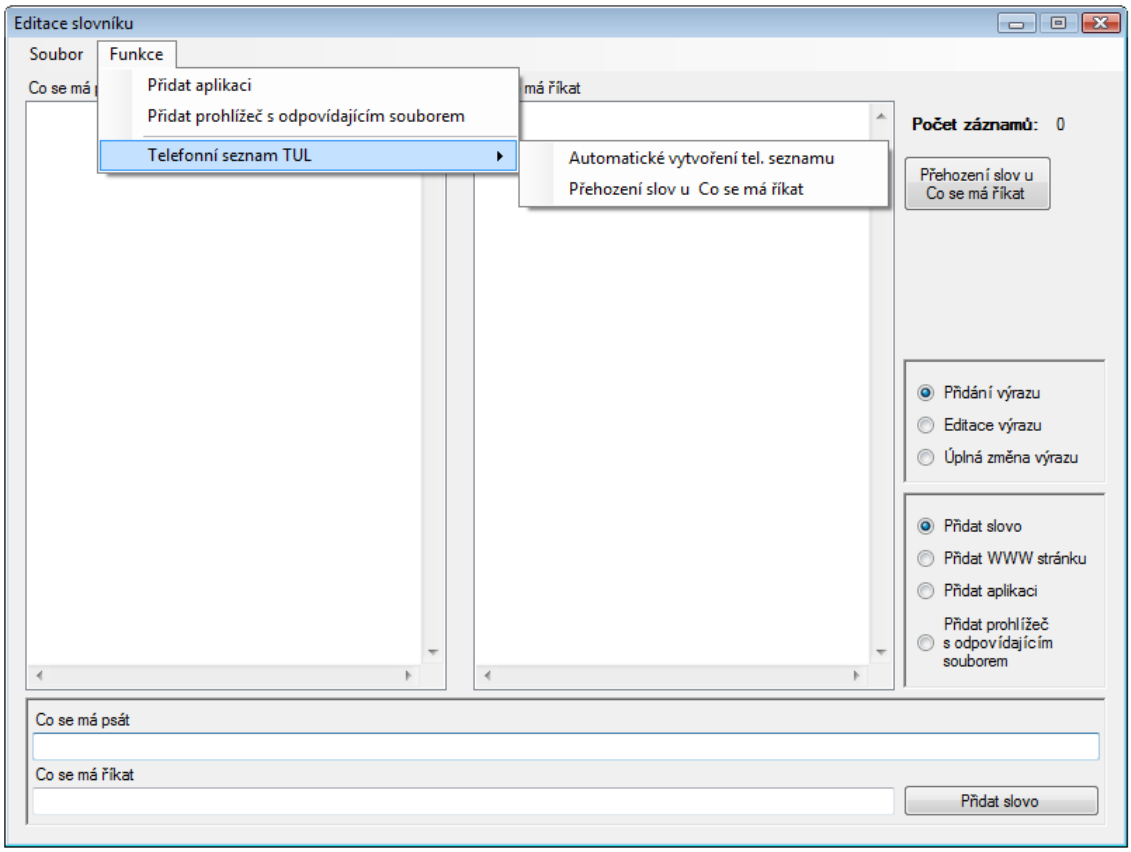
pomocí tabelátoru. V případě že chybí například titul, je opět vytvořena mezera tabelátorem. Struktura nutného sloupcového seřazení pro automatické vytvoření telefonního seznamu pracovníků na Technické univerzitě v Liberci je vidět v tabulce 4.1.

Tab.4.1: Struktura telefonního seznamu pro automatické vytvoření

1.sloupec	2.sloupec	3.sloupec	4.sloupec
tituly	**příjmení a jméno**	**tel. číslo**	**pracoviště**
Ing.	Holada Miroslav, Ph.D.	3080	Ústav ITE
Ing.	Chaloupka Josef, Ph.D.	3099	Ústav ITE
prof. Ing.	Konopa Vojtěch, CSc.	3483	Ústav MTI
	Pánková Věra	3429	Děkanát fakulty mechatroniky

- **přehození slov u Co se má říkat** – tato funkce nám umožní, přehození slov v seznamovém rámečku „Co se má říkat“. Původní výraz se zachová a vytvoří se nový, na který bude rozpoznávací systém reagovat stejně, akorát k tomu budou možné dvě promluvy uživatele. V případě, že je slovo pouze jedno v „Co se má říkat“, tak není možnost přehodit slova a nic se nevykoná. Tato funkce je značně spojena s funkcí automatického vytvoření telefonního seznamu pracovníku TUL. Do pole promluvy uživatele je automaticky vytvořen výraz z příjmení a jména pracovníka a pro rychle a pohodlné přehození slov na jméno a příjmení stačí využití této funkce.
- dalšími známými funkcemi jsou – **nový slovník** (načtení nových prázdných seznamových rámečků), **otevřít slovník** (načtení jednotlivých výrazů do odpovídajících seznamových rámečku), **připojit slovník** (spojení slovníku buď s dalším slovníkem a vytvoření jednoho velkého slovníku, nebo s novými ještě neuloženými výrazy), **uložit a uložit jako** (uložení jednotlivých výrazů ze seznamových rámečků do souboru s příponou *.vcb)

Přehled viditelných funkcí editoru slovníku je vidět na obrázku 4.3.



Obr. 4.3: Editor slovníku a jeho jednotlivé funkce

Neviditelné (skryté) funkce jsou automaticky vkládány do všech tvořených a vytvořených slovníku v momentě ukládání. Proto je dobré je brát jako funkce editoru slovníku. Opět slouží k pohodlnému a uživatelsky příjemnému ovládání a práci s interaktivním hlasovým rozhraním. Neviditelnými funkcemi jsou:

- **přidat stránku** – jedná se o funkci, která umožní rychlé přidání WWW stránky bez použití editoru slovníku. Je nutné požadovanou internetovou adresu načíst do clipboardu (CTRL + C) a pomocí nahrávání rozpoznávacího systému a promluvy Přidat stránku se přidá do slovníku. V případě, že je stránka špatně načtena do clipboardu, je možné uložení neprovést. V opačném případě je stránka přidána do připraveného slovníku pro rychlé vytváření WWW stránek. Název soubor je Slovnik_s_WWW_strankama.vcb
- **zobrazení slovníku** – tato funkce nám umožní zobrazení aktuálního slovníku načteného v rozpoznávacím systému. Tato funkce je z důvodu možnosti projít si všechny možné promluvy uživatele a tím i rychlejší možnosti dosažení

požadovaného cíle. Vyvolání prohlížeče s aktuálním slovníkem je pomocí nahrávání rozpoznávacího systému a promluvy Zobrazení slovníku.

- **editace slovníku** – tato funkce nám umožní editaci aktuálního slovníku načteného v rozpoznávacím systému, aniž bychom museli použít hlavní panel aplikace. Vyvolání editoru slovníku s načteným aktuálním slovníkem je pomocí nahrávání rozpoznávacího systému a promluvy Editace slovníku.

4.3 Testování aplikace

Pro zjištění informací k doladění malých detailů aplikace (kvalita rozpoznávání, kvantita funkcí, design, ovladatelnost) a možnosti pro další vývoj, byla aplikace rozšířena mezi skupinu nových uživatelů, kteří aplikaci využívali určitou dobu a poté zhodnotili jednotlivé vlastnosti a funkce.

Jelikož se jednalo o zcela nové uživatele interaktivních hlasových rozhraní, tak měli značné výhrady k nedostupné nápovědě. Daný problém byl ihned napraven a nápověda byla naprogramována. Hlavní panel na první dojem také nevzbuzoval nic zajímavého a zvláštního a proto byla snaha o zjednodušení a zvýšení zajímavosti volbou vhodných barev panelu. Naopak editor slovníku se setkal s velkou oblibou. Byl uživatelsky příjemný a přehledný, ovladatelnost byla jednoduchá, usnadnění a urychlení práce při tvorbě slovníku byla veliká a zároveň bylo zaručeno správné schéma slovníku nutné pro rozpoznávací systém.

4.4 Shrnutí

Aplikace měla vhodně zvolenou strategii návrhu interaktivního hlasového rozhraní. Byly zvoleny vhodné parametry, které vytvořené aplikaci vyhovují a tvoří ji zajímavou a hlavně uživatelsky příjemnou a prospěšnou pro hlasové ovládání počítače.

Rozpoznávací systém pro rozpoznávání promluvy řečníka, byl velmi kvalitní a izolovaná slova a krátké spojení slov rozpoznával s velmi velkou přesností a nedocházelo k nežádoucím chybám rozpoznávacího systému vlivem špatných odpovědí a tedy i zbytečnému prodlužování dialogu mezi uživatelem a systémem.

Modul dialogového manažera je součástí hlasového dialogového systému. Jedná se o typ dialogového systému s konečným počtem stavů. Ovšem nejde zcela o konečný počet stavů, protože tento systém pracuje se dvěma slovníky. Jeden slovník je nahrán

přímo na serveru s hlasovým dialogovým systémem. Obsahuje běžně využívané izolované promluvy či krátké spojení slov a uživatel nemá možnost přístupu. Kapacita tohoto slovníku je mnohonásobně větší než kapacita druhého slovníku (první slovník asi statisíce slov a druhý slovník dva tisíce slov), který je uživatelský. Tento slovník nám umožňuje dynamické rozšiřování dialogového systému s konečným počtem stavů. Více o dialogovém systému s konečným počtem stavů je v odstavci 2.5.3.1

Závěr

Jelikož realizace strategie návrhu interaktivního hlasového rozhraní byla v prostředí Microsoft Visual Studio C++ .NET, který má nové struktury a lepší programovatelné vlastnosti než starší programovací jazyky, bylo možno jednodušším způsobem dosáhnout zajímavé a uživatelsky prospěšné aplikace pro hlasové ovládání chodu počítače. Aplikace byla programována ve verzi 2005, protože novější Microsoft Visual Studio 2008 mělo určité problémy, které měly negativní vliv na možnosti návrhu designu aplikace, aniž by došlo k chybnému programování.

Strategie návrhu interaktivního hlasového rozhraní je ovlivněna řadou parametrů, které ovlivňují výslednou aplikaci. Pro jednotlivě navržené parametry interaktivního hlasového rozhraní se aplikace stala uživatelsky příjemná, přehledná a hlavně přínosná pro uživatele v oblasti hlasového ovládání chodu počítače, což se podle průzkumu z testování aplikace do velké míry povedlo a spokojenost uživatelů s aplikací byla. Důležité také je, že rozpoznávací systém rychle a s velkou pravděpodobností správně rozpoznával promluvy řečníků a tedy i správně navazoval na dialog svými odpověďmi. To vede k dalšímu zvýšení spokojenosti uživatele.

Výhodou vytvořeného programu je řada zajímavých a užitečných funkcí. Funkce nahrávání rozpoznávacího systému zajišťuje větší robustnost aplikace. Nedochází k trvalému rozpoznávání promluv a tím k odezvám systému, které by negativně působili na uživatele. Další funkcí je možnost editoru vlastního uživatelského slovníku, který je velmi prospěšný pro jednoduchý a přesný návrh promluv uživatele a odpovědí systému. Zajímavou možností je automatické vytvoření slovníku s telefonním seznamem. Tato funkce je vhodná pro větší firmy, organizace či školy, kteří mají velký počet zaměstnanců a vyhledávání konkrétního čísla v telefonním seznamu je náročné.

Při malé úpravě zdrojového kódu pro nahrávání rozpoznávacího systému by bylo možné aplikaci vhodně využít i pro zdravotně handicapované osoby s pohybovým omezením ke snadné a pohodlné práci s počítačem pomocí hlasového ovládání.

Strukturu hlasového dialogového systému by bylo možné doplnit o modul syntézy řeči, který by umožňovat hlasovou odpověď uživateli na požadovanou promluvu. Toto nebylo v zadání a řešení by bylo nad rámec tvořené práce.

Z práce také vyplývá, že problematika návrhu hlasového dialogového systému je neobyčejně rozsáhlá a složitá a nabízí mnoho prostoru pro další vývoj nových a

vylepšování stávajících dialogových systémů, ať už v oblasti rozpoznávání řeči, porozumění mluvenému jazyku, generování odezvy, syntézy řeči nebo dialogového řízení.

Seznam použité literatury

[1] Chalupa A.: *1001 tipů a triků pro Visual C++*. Computer Press, Brno, 2003, ISBN 80-7226-842-2

[2] Corera A.: *Visual C++ .NET pro programátory v C++*. Computer Press, Brno, 2003, ISBN 80-7226-860-0

[3] Hui P.Y., Meng H.M.: *Joint Interpretation of Input Speech and Pen Gestures for Multimodal Human-Computer Interaction*, [cit. Interspeech 2006], ISSN 1990-9772

[4] Kačmář D.: *Programujeme .NET aplikace ve Visual Studiu .NET*. Computer Press, Praha, 2001, ISBN 80-7226-569-5

[5] Kadlec V.: *Učíme se programovat v jazyce C*. Computer Press, Praha, 2002, ISBN 80-7226-715-9

[6] Koenig A., Moo Barbara E.: *Rozumíme C++*. Computer Press, Praha, 2003, ISBN 80-7226-656-1

[7] Kruglinski D.J.: *Mistrovství ve Visual C++*. Computer Press, Praha, 1999, ISBN 80-7226-132-0

[8] Levin E., Mané A.M.: *Voice User Interface Design for Automated Directory Assistance*, [cit. Interspeech 2005], ISSN 1018-4074

[9] Matoušek D.: *Visual C++ 6.0*. BEN, Praha, 2003, ISBN 80-7300-130-6

[10] Nouza J.: *Počítačové zpracování řeči.* Sborník článků, Liberec, 2001, ISBN 80-7083-551-6

[11] Psutka J.: *Komunikace s počítačem mluvenou řečí*. Academia, Praha, 1995, ISBN 80-200-0203-0

[12] Psutka J.: *Mluvíme s počítačem česky.* Academia, Praha, 2006, ISBN 80-200-1309-1

[13] Reynolds-Haertle R.A.: *OOP – objektově orientované programování*. Mobil Media, Praha, 2002, ISBN 80-86593-25-8

[14] Wikipedia: *Human-copmuter interaction*, [online], [cit. 2008-05-13], URL: < http://en.wikipedia.org/wiki/Human-computer_interaction>

[15] Wikipedia: *Voice User Interface*, [online], [cit. 2008-04-08], URL:< http://en.wikipedia.org/wiki/Voice_User_Interface>

Obsah přiloženého CD

- Text bakalářské práce ve formátu PDF
- Vytvořený program návrhu interaktivního hlasového rozhraní
- Zdrojové kódy programu vytvořeného v prostředí Visual Studio
- Telefonní seznam osob pracujících na Technické Univerzitě v Liberci
- Datové soubory obsahující již vytvořené databáze slov