

**Univerzita Palackého v Olomouci**

**Přírodovědecká fakulta**

**Katedra geoinformatiky**

**Martin KADLČÁK**

**NÁVRH IMPLEMENTACE JAZYKOVÉHO  
MODULU PRO OVLÁDÁNÍ MAPY HLASEM**

**Magisterská práce**

**Vedoucí práce: Ing. Zdena Dobešová Ph.D.**

**Olomouc 2012**

### **Čestné prohlášení**

Prohlašuji, že jsem magisterskou práci magisterského studia oboru Geoinformatika vypracoval samostatně pod vedením Ing. Zdeny Dobešové Ph.D.

Všechny použité materiály a zdroje jsou citovány s ohledem na vědeckou etiku, autorská práva a zákony na ochranu duševního vlastnictví.

Všechna poskytnutá i vytvořená digitální data nebudu bez souhlasu školy poskytovat.

V Olomouci 18. 4. 2012

podpis\_\_\_\_\_

Děkuji vedoucímu práce Ing. Zdeně Dobešové Ph.D za podněty a připomínky při vypracování práce. Dále děkuji své přítelkyni a celé své rodině za podporu během celého studia na vysoké škole.

Vložený originál **zadání** bakalářské/magisterské práce (s podpisy vedoucího katedry, vedoucího práce a razítkem katedry). Ve druhém výtisku práce je vevázána fotokopie zadání.

# OBSAH

|  |           |
|--|-----------|
| <b>ÚVOD</b> .....  | <b>8</b>  |
| <b>1 CÍLE PRÁCE</b> .....  | <b>9</b>  |
| <b>2 POUŽITÉ METODY A POSTUPY ZPRACOVÁNÍ</b> .....                   | <b>10</b> |
| 2.1 Použité programy .....   | 11        |
| <b>3 ÚVOD DO PROBLEMATIKY</b> .....                                  | <b>12</b> |
| 3.1 Historie ovládání počítače hlasem .....                          | 12        |
| 3.2 Oblasti využití hlasového modulu.....                            | 14        |
| 3.2.1 Rozpoznání příkazů a vykonání určité činnosti.....             | 14        |
| 3.2.2 Přepis mluveného slova do psané podoby.....                    | 14        |
| 3.2.3 Překlad mluvy do požadovaného jazyka .....                     | 14        |
| 3.2.4 Bezdrátové ovládání spotřebičů pomocí hlasu .....              | 14        |
| 3.2.5 Syntéza psané podoby do mluvené.....                           | 15        |
| 3.2.6 Informační a komunikační služby uskutečňované po telefonu..... | 15        |
| 3.2.7 Rozpoznávání řečníka.....                                      | 15        |
| 3.2.8 Voice Browser .....  | 15        |
| 3.2.9 Uživatel je handicapovaná osoba.....                           | 16        |
| <b>4 TEORIE OVLÁDÁNÍ POČÍTAČE HLASEM</b> .....                       | <b>17</b> |
| 4.1 Základní oblasti zpracování řeči .....                           | 17        |
| 4.1.1 Syntéza řeči.....  | 17        |
| 4.1.2 Rozpoznávání řeči .....  | 18        |
| 4.1.3 Hlasové dialogové systémy .....                                | 19        |
| 4.2 Jak se tvoří hlas .....  | 20        |
| 4.2.1 Vytváření řeči člověkem.....                                   | 20        |
| 4.2.2 Modelování vytváření řeči.....                                 | 21        |
| 4.3 Problémy se zpracováním lidské řeči.....                         | 22        |
| 4.3.1 Problematika větné skladby a skloňování.....                   | 22        |
| 4.3.2 Problematika prozodických vlastností.....                      | 22        |
| 4.4 Okolí mluvčího .....   | 23        |
| <b>5 PROGRAMY NA OVLÁDÁNÍ POČÍTAČE HLASEM</b> .....                  | <b>24</b> |
| 5.1 MyVoice.....   | 24        |
| 5.2 JetVOICE.....  | 25        |
| 5.3 Respondings Heads .....  | 26        |
| 5.4 VoiceMate Professional .....                                     | 27        |
| 5.5 Pearl-box Voice.....   | 28        |
| 5.6 Tyle .....   | 29        |
| 5.7 e-Speaking.....  | 30        |

|           |  |           |
|-----------|--|-----------|
| 5.8       | Dragon Naturally-Speaking .....                              | 31        |
| 5.9       | Rozpoznávání hlasu ve Windows .....                          | 32        |
| 5.10      | Zhodnocení programů .....                                    | 33        |
| <b>6</b>  | <b>OVLÁDÁNÍ MOBILNÍHO ZAŘÍZENÍ HLASEM .....</b>              | <b>34</b> |
| 6.1       | Hlasové vytáčení .....                                       | 34        |
| 6.2       | Hlasové vyhledávání .....                                    | 34        |
| 6.3       | Ovládání GPS navigací v mobilním telefonu pomocí hlasu ..... | 36        |
| 6.4       | Hlasové ovládání GPS navigací .....                          | 37        |
| <b>7</b>  | <b>ARCSOUND .....</b>  | <b>38</b> |
| 7.1       | Popis aplikace ArcSound .....                                | 38        |
| 7.2       | Funkce ArcSound .....  | 39        |
| 7.2.1     | Spouštění příkazu hlasem .....                               | 39        |
| 7.2.2     | Zvuky v mapě .....   | 40        |
| 7.3       | Hodnocení ArcSound .....                                     | 41        |
| <b>8</b>  | <b>MOŽNOSTI UPLATNĚNÍ V GIS .....</b>                        | <b>43</b> |
| 8.1       | Ovládání mapy hlasem .....                                   | 43        |
| 8.2       | Vyhledávání v mapách .....                                   | 43        |
| 8.3       | Interaktivní mapy .....                                      | 44        |
| 8.4       | Použití hlasového modulu u GPS .....                         | 44        |
| 8.5       | Využití v DPZ .....  | 44        |
| 8.6       | Využití ve 3D .....  | 45        |
| 8.7       | Hlasové ovládání totální stanice .....                       | 45        |
| 8.8       | Informační tabule .....                                      | 46        |
| 8.9       | Ovládání mobilního GISu hlasem .....                         | 46        |
| 8.10      | Hlasové ovládání animací .....                               | 46        |
| 8.11      | Hlasem plně ovládaný mapový program .....                    | 47        |
| 8.12      | Virtuální mapy .....   | 47        |
| <b>9</b>  | <b>NÁVRH A REALIZACE APLIKACE NA OVLÁDÁNÍ MAPY HLASEM</b>    | <b>48</b> |
| 9.1       | Použití JetVoice .....                                       | 48        |
| 9.2       | Použití ArcExplorer .....                                    | 48        |
| <b>10</b> | <b>DEFINOVÁNÍ HLASOVÝCH POVELŮ .....</b>                     | <b>50</b> |
| 10.1      | Snadná výslovnost .....                                      | 50        |
| 10.2      | Snadné zapamatování .....                                    | 51        |
| 10.3      | Zvolené povely ve vytvořené aplikaci .....                   | 52        |
| <b>11</b> | <b>VÝSLEDKY .....</b>  | <b>56</b> |
| <b>12</b> | <b>DISKUZE .....</b>   | <b>58</b> |
| <b>13</b> | <b>ZÁVĚR .....</b>   | <b>60</b> |

**POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE**  
**SUMMARY**  
**PŘÍLOHY**

## ÚVOD

Komunikace pomocí řeči patří mezi lidmi k základnímu způsobu dorozumívání se mezi sebou. Důvodem toho, že je tak využívána je především to, že můžeme pomocí pár vyslovených slov sdělit vše, co sdělit chceme a to sdělení je tudíž rychlejší, než třeba sdělení pomocí psaného projevu.

Díky tomu, že lidé řeč využívají tak hojně, se v posledních letech snaží také o to, aby mohli určitým způsobem komunikovat i s počítačem. Vědci a vývojoví technici se proto snaží, aby počítače rozuměli, co jim člověk říká, aby mohli plnit pokyny, které jim člověk sděluje a hlavně, aby se v mluveném dialogu stal počítač člověku plnohodnotným partnerem. Tento cíl je sledován zejména proto, že takový způsob komunikace může být člověku velmi prospěšný a často mu může i výrazně usnadnit život [16].

Člověk využívá v dnešní době počítač prakticky každodenně a víceméně ke každé činnosti. Pokud by využíval k ovládní počítače pouze svůj hlas a ovládali by jej pomocí hlasových pokynů, jejich práce by se podstatně urychlila a zjednodušila. Samozřejmě by takový způsob ovládní počítače zcela nenahradil klasický způsob ovládní (pomocí myši a klávesnice), jelikož některé složitější úkony (např. kreslení v grafických programech) by byly pro změnu obtížnější ovládat pomocí hlasu a hlasových příkazů.

Bohužel je třeba smutně konstatovat fakt, že i přes intenzivní vývoj těchto řečových technologií, dnes neexistuje takový nástroj, díky kterému by člověk mohl plnohodnotně ovládat počítač hlasem. Není to proto, že by vědci vykazovali malé úsilí takový plnohodnotný systém na ovládní počítače hlasem vytvořit, ale především proto, že vytvoření takového systému s sebou nese mnoho algoritmicky a technicky velice složitých úloh, které se týkají zejména zpracování mluvené řeči a rozpoznávání řeči.

Samozřejmě, že dnes existuje mnoho systémů, které již ovládat počítač hlasem umí, ale nejedná se o plně sofistikované systémy použitelné napříč všemi oblastmi použití počítačů. Vždy se jedná jen o tematicky zaměřená řešení, která jsou aplikovatelná jen na konkrétní příklad, jelikož takový systém má u sebe uložen slovník, ve kterém je určitý počet slov (příkazů) a k těmto jednotlivým slovům existují konkrétní příkazy, co a jak má počítač udělat. Tato řešení jsou v poslední době poměrně hojně využívána v praxi, avšak k plně sofistikovanému systému mají stále poměrně dost daleko.

Široce využívány jsou i dílčí komponenty hlasových dialogových systémů, tj. moduly syntézy a rozpoznávání řeči. Všeobecné využití nacházejí například různé systémy ovládní strojů a zařízení hlasovými povely či automatický přepis diktátu (vhodné zejména jsou-li oči i ruce člověka zaměstnány jinými činnostmi nebo v případě využití tělesně handicapovanými lidmi).

Vědci se stále snaží vyvinout systém, který bude umět plnohodnotně ovládat počítač pouze mluveným projevem a pokud se jim to podaří, bude to znamenat další průlom v oblasti výpočetní techniky a zdolání zdánlivě nezdolatelné překážky ve smyslu větší integrace mezi člověkem a počítačem.



# 1 CÍLE PRÁCE

Cílem magisterské práce je popsat a navrhnout možnosti implementace hlasového ovládání digitálních map. Cíle jsou rozděleny na cíle v teoretické části práce a cíle v praktické části diplomové práci.

Cílem teoretické části je zhodnotit dostupné programové prostředky, které již v oblasti ovládání počítače hlasem v současné době existují. Hodnocení bude probíhat na základě předem daných parametrů a výsledkem tohoto hodnocení bude tabulka s přesně zhodnocenými parametry a celkovým výsledkem. Dalším dílčím cílem této části diplomové práce je popsat metody, které jsou používány pro zpracování lidského hlasu počítačem a také bude popsáno možnosti spojení s mapami či GIS softwary. V neposlední řadě budou nastíněny oblasti, které jsou určitým způsobem propojeny s geoinformatikou, a kde by právě hlasový modul na ovládání mapy hlasem byl přínosem, popř. kde by ulehčil práci jejím uživatelům.

Cílem praktické části je ze získaných znalostí o hlasových technologiích vytvořit ukázkový příklad, na kterém budou ukázány možnosti ovládání mapy lidským hlasem. Nebude se jednat o aplikaci, která by celá byla vytvořena napsáním zdrojového kódu, ale bude vytvořena z již existujícího programu na ovládání počítače hlasem, a která bude modifikována na požadavky ovládání mapy hlasem. Druhým cílem praktické části je vytvořit vhodné slovní pokyny pro ovládání mapy a pro získávání informací z mapy. Také bude provedena analýza vhodných hlasových povelů pro ovládání mapy hlasem. Bude zhodnoceno, který jazyk je nejvhodnější pro definování hlasových povelů a také, které významné prvky ovlivňují úspěšnost fungování hlasového povelu.

## 2 POUŽITÉ METODY A POSTUPY ZPRACOVÁNÍ

Metody a postupy použité při zpracování této diplomové práce lze rozdělit na tři části: rešeršní část a doporučení, kde všude v geoinformatice by mohl být hlasový modul použit, vytvoření aplikace na ovládání mapy hlasem, analýza a definování vhodných slovních povelů pro ovládání mapy hlasem.

Nejdříve byla provedena rešerše dostupné literatury na téma obecně ovládání počítače hlasem. Při výběru vhodné literatury bylo dbáno na to, aby literatura co nejvíce vypovídala o zkoumaném oboru. Studovány byly také odborné články na toto téma, ale také články, které byly zaměřeny spíše vědecko-populárně. Byla vybírána literatura jednak česká a jednak cizojazyčná, především anglická. Samotné studium této literatury zabralo nemálo času, jelikož odborná literatura na toto téma je velice složitá na čtení a ve většině případů se zaměřuje na fyzikální podstatu toho, jak ovládání počítače hlasem funguje. Nejvíce mi při studiu a pochopení celé problematiky ovládání počítače hlasem pomohly především vědecko-populární články, které popisovaly nejaktuálnější vývoj v této oblasti a poskytly mi velkou inspiraci.

Dalším důležitým prvkem při zjišťování vývoje v oblasti ovládání počítače hlasem byla nejrůznější videa. Jednak to byla instruktážní videa k jednotlivým aplikacím, která určitým způsobem využívala hlasový modul pro jejich ovládání, a jednak to byla videa uživatelů, kteří ukazovali, jak daná aplikace funguje reálně. Video byla použita pro studium především z toho důvodu, že jsem neměl přístup k těmto aplikacím, jelikož jejich používání je zpoplatněno, nebo tyto aplikace již nejsou k dispozici ke stažení a vyzkoušení.

Dále bylo zapotřebí otestovat již vytvořené programy, které umí ovládat počítač pomocí hlasu. Celému otestování předcházelo vyhledání příslušných programů na internetu. Hledání probíhalo především pomocí odkazů, které byly uvedeny v člancích k problematice ovládání počítače hlasem. Na základě vyhledávání jsem si vybral 9 programů, u kterých jsem se rozhodl udělat otestování funkčnosti. Některé programy se mi bohužel nepodařilo zprovoznit a u některého zase byla tak vysoká cena za použití, že jsem byl donucen program hodnotit jen na základě zkušenosti uživatelů, kteří s programem již pracovali a také na základě videí k programu. Ostatní programy jsem již vyzkoušel opravdu fyzicky, kde jsem zkoumal především úspěšnost rozpoznávání hlasových povelů a také možnosti jednotlivého nastavení programů. K celkovému zhodnocení jsem poté použil srovnávací tabulku, kdy na základě zvolených kritérií jsem jednotlivé programy porovnal a určil pomocí procentuálního ohodnocení funkčnost jednotlivých programů. Výslednou tabulku s hodnocením jsem poté vložil jako jednu z příloh diplomové práce.

Samostatné testování bylo provedeno u programu ArcSound, který je přímo vytvořen k ovládání programu ArcMap za pomoci hlasu. Testování bylo zastaveno již v počátku, kdy se mi nepodařilo program nainstalovat. Zhodnocení tak bylo opět provedeno díky videím na internetu a ohlasů uživatelů, kteří s aplikací již v minulosti pracovali.

Po důkladném prostudování a hlavně pochopení celé problematiky ovládání počítače hlasem jsem již začal pracovat na tom, ve kterých oblastech geoinformatiky by bylo možné hlasový modul použít. Bylo zvaženo jak použití v mobilních aplikacích, desktopových aplikacích i nejrůznější použití v terénu. Byly brány v potaz téměř všechny oblasti geoinformatiky, kde by hlasový modul byl opravdovým přínosem.

Důležitou částí práce byla analýza a definování vhodných hlasových příkazů pro aplikaci na ovládání mapy hlasem. Analýza probíhala především tak, že nejdříve bylo zapotřebí zjistit, který jazyk je pro použití hlasových povelů nevhodnější. Byly vybrány dva jazyky, čeština a angličtina. V každém jazyku bylo vytvořeno 20 hlasových povelů, které významem odpovídají tomu, co přesně ovládají. Poté již proběhlo zkoušení jednotlivých jazykových souborů. Každý soubor příkazů byl vyzkoušen v 10 jednotlivých testováních a na základě celkové procentuální úspěšnosti každého souboru byl vybrán úspěšnější jazyk.

Dalším důležitým krokem bylo definování vhodných povelů. To probíhalo tak, že opět na 10 testováních byla vypočtena úspěšnost jednotlivých povelů a pokud byla u některého povelu tato úspěšnost výrazně menší než u ostatních, tak bylo pro daný povel vybráno jiné slovo a opět vyzkoušeno a pokud byla úspěšnost vyšší, tak slovo bylo ponecháno v souboru. Celá funkčnost souboru byla testována dvěma lidmi, jedním zástupcem mužského pohlaví a jedním ženského pohlaví, kvůli tomu, aby testování bylo opravdu věrohodné.

V poslední části diplomové práce šlo o vytvoření aplikace na ovládání počítače hlasem. Nejdříve musela být opět provedena analýza ke zjištění toho, který program na ovládání počítače hlasem by byl vhodný pro náš příklad. Byly vybrány 3 programy a u nich postupně byla zkoušena funkčnost. Ta byla zkoušena pomocí úspěšností rozpoznávání jednotlivých povelů. Důležitým faktorem při výběru vhodného programu byla také zohledněna možnost různého dalšího nastavení programu. Po zvážení veškerých kladů a mínusů jednotlivých vybraných programů byl vybrán jeden program, který nejlépe odpovídal daným požadavkům.

Při výběru programu, který slouží jako prohlížečka pro předem vytvořené mapy byl již dopředu vybrán jako nejlepší kandidát program ArcExplorer, díky své vysoké profesionalitě a díky velkým možnostem na zobrazování nejrůznějších podkladových map.

## **2.1 Použité programy**

Byl používán zejména software na ovládání počítače hlasem. Dále kancelářský balík MS Office, konkrétně programy Word, Excel a PowerPoint. Obrázky pro diplomovou práci byly zpracovávány v programu FastStone Capture. Pro tvorbu internetových stránek o diplomové práci byl použit program PSPad editor.

## 3 ÚVOD DO PROBLEMATIKY

### 3.1 Historie ovládání počítače hlasem

Myšlenka, že by člověk mohl komunikovat s počítačem vzrušuje člověka již několik století. První doložené záznamy o „jakémisi“ druhu komunikace s počítačem byly popsány již v druhé polovině 18.století experimenty s mechanickým syntetizérem lidského hlasu.

Experimenty prováděné von Kempelenem, rakouským vynálezcem, byly založené na konstrukci stroje, který imitoval lidské dýchací ústrojí: měchy napodobovaly plíce, místo hlasivek používal regulovatelnou trubici a další trubice nahrazovala nosní dutinu se dvěma nosními dírkami [27]. Stroj měl také jakési „ústa“, která byly vyrobené z gumy. Limitujícím pro von Kempelena bylo to, že se mu nepodařilo nalézt vhodný materiál, kterým by napodobil jazyk, měkké patro a také lidské rty.



Obr.1 Rekonstrukce Kempelenova mluvícího stroje (zdroj: Wikipedia)

Podle dobových záznamů dokázal stroj vyslovovat různé slabiky, jejichž spojením bylo možné sestavit věty v různých jazycích. Všechny zvuky byly tvořeny vibracemi slonovinového jazýčku, kterým procházel vytlačovaný vzduch.

Dalším významným krokem v této oblasti bylo vynalezení Fourierovy analýzy na konci 18.století. Fourierova analýza byla známá pouze v její analogové verzi. Opravdové praktické uplatnění lze sledovat až s nástupem číslicových počítačů, kdy se uplatnily a rozvinuly metody digitalizace a číslicového zpracování akustického řečového signálu [16].

Propracovanější a technicky náročnější systémy, které pracují v reálném čase, což znamená že, např. ihned provádějí převod z psané podoby jazyka do mluvené, se

rozvíjely až s nástupem signálových procesorů v první polovině devadesátých let. Dnešní výpočetní výkon běžného počítače, v němž je instalován modul zpracování řečového signálu, už vůbec nevyžaduje podpurné předzpracování signálu externím signálovým procesorem, ale zvládá bez problémů celý výpočet mnohabodové Fourierovy transformace v reálném čase, včetně navazujícího zpracování [16].

Odlišný přístup ke klasifikaci mluvené řeči, který využíval v té době nejnovějších poznatků z oblasti umělé inteligence, přinesly systémy vyvinuté v letech 1971 až 1976 v rámci projektu DARPA-SUR [17]. Tento projekt byl zaměřen na vývoj systému, který by rozuměl souvislé lidské řeči. Byly poprvé navrženy a experimentálně ověřeny některé postupy, které později přispěly k pokroku v oblasti rozpoznávání lidské řeči.

Dalším milníkem ve vývoji systémů pro komunikaci mezi člověkem a počítačem byla v 80. letech technika klasifikace řeči založená na statistickém přístupu, která je vhodná pro rozpoznávání souvislé řeči. Zcela nové možnosti rozpoznávat souvislou řeč se slovníkem desítek i stovek tisíc slov zde přitom poskytl nový způsob modelování řeči, který je založen na použití tzv. Markovových modelů, které modelují nikoli celá slova, ale kratší subslovní jednotky (např. fonémy, alofony, trifony apod.). Systém akustických Markovových modelů je obvykle trénován na vhodně připravené a rozsáhlé řečové databázi a z jednotlivých natrénovaných modelů subslovních jednotek jsou pak podle potřeby sestavovány modely slov i celých promluv. Tento způsob modelování a rozpoznávání zejména souvislé řeči zcela převládl a je využíván i v současné době, neboť nyní poskytuje prokazatelně nejlepší výsledky [17].

Koncem 20. století také došlo na poměrně velké využití hlasových dialogových systémů, které využívají technologie syntézy a rozpoznávání řeči. Dokonce byl v roce 2003 vyvinut počítačový jazyk VoiceXML, který podporuje a usnadňuje návrh a realizaci dialogových aplikací. Tento jazyk slouží k popisu scénářů pro automatizovaný dialog mezi člověkem a počítačem. Lze pomocí něj popsat informace, které je potřeba od uživatele získat [26]. Síla jazyka spočívá v možnosti vést s uživatelem dialog a přizpůsobovat svou komunikační strategii aktuální situaci. Umožňuje zpřístupňovat obsah Internetu a informací obecně pomocí hlasu a telefonu, vytvářet dialogové systémy založené na mluvené řeči a další věci spojené s řečí.

V dnešní době je vývoj a výzkum v oblasti hlasových systémů soustředěn především na rozvoj systémů, které komunikují s člověkem s použitím neomezeného přirozeného jazyka. Oblast automatického dialogového systému je stále omezena na komunikaci, pro kterou je dialog připraven, neboli systém opět umí reagovat pouze na to, na co je naprogramovaný.

Samozřejmě, že se pracuje i na systému umělé inteligence, který by uměl reagovat na veškeré uživatelské pokyny a také s uživatelem plnohodnotně komunikovat. Tato oblast výzkumu je spíše ještě teoretická a tak trochu science-fiction, jelikož to bychom se poté bavili již o počítačích a strojích, které umějí samy přemýšlet a tady ten vývoj je stále ještě během na dlouhou trať.

## **3.2 Oblasti využití hlasového modulu**

### **3.2.1 Rozpoznání příkazů a vykonání určité činnosti**

Rozpoznáváním izolovaných příkazů se rozumí aplikace, která v sobě obsahuje databázi předvolených hlasových příkazů a k těmto příkazům je vybrána konkrétní činnost v počítači, která se má provést.

Dnes se tato oblast využití používá v širokém spektru oborů. Existují programy, které umožňují ovládat téměř celý počítač. Velkou výhodou těchto systémů je to, že každý uživatel si může vytvořit své vlastní příkazy tak, jak vyhovují jemu a u počítače si může vybrat ty činnosti, které on sám potřebuje.

Dokonce i operační systém Windows obsahuje modul na ovládání počítače hlasem, avšak pouze pro anglickou lokalizaci systému.

### **3.2.2 Přepis mluveného slova do psané podoby**

Přepis mluveného slova do psané podoby je využíván především u soudů a také v kancelářích, např. při diktování dopisů. Jedná se o aplikaci, kdy program umí zaznamenat vše, co člověk řekl. V dnešní době je tato oblast tak propracovaná, že kvalita zaznamenané řeči odpovídá takové kvalitě, jako kdyby přepis prováděl člověk.

### **3.2.3 Překlad mluvy do požadovaného jazyka**

Překlad mluvy do požadovaného jazyka pracuje tak, že uživatel do programu mluví ve svém jazyce, program si pro sebe zaznamená, co uživatel řekl a poté to přeloží do požadovaného jazyka a tento překlad je následně přečten samotným programem v tom daném jazyce [22]. Jedná se o jakési nahrazení překladatelů. Tato aplikace je poměrně dosti náročná, jelikož musí mít v sobě zabudován modul pro rozpoznávání řeči, dále překladač a v neposlední řadě syntetizátor lidského hlasu, který umí přečíst to, co je přeloženo. I přes velkou náročnost již dnes existují sofistikované programy, které tuto oblast využití hlasového modulu zvládnou.

### **3.2.4 Bezdrátové ovládání spotřebičů pomocí hlasu**

Už v praxi poměrně dosti využívané je ovládání určitých spotřebičů hlasem. Tímto způsobem jsou nejvíce ovládány domácí spotřebiče hlasem [13]. Běžně jsou ovládány světla v domě za pomoci příkazů, dále také nastavení teploty v domě pomocí hlasových příkazů, ale i třeba různých přehrávačů, videa a televize.

Tato úloha pracuje na principu jednoduchého rozpoznávání jednotlivých příkazů. Tudíž tyto spotřebiče obsahují pouze jeden jednoduchý modul na rozpoznávání izolovaných slov, avšak tento modul na druhou stranu musí pracovat spolehlivě a bez nějakého zaváhání.

### 3.2.5 Syntéza psané podoby do mluvené

Syntéza psané podoby do mluvené znamená, že aplikace tohoto hlasového modulu umí přečíst napsaný text v digitální podobě jako člověk. Jinak řečeno, tato aplikace umí simulovat lidský hlas a tudíž člověk, který poslouchá čtení takového textu má opravdu pocit, že daný text čte člověk a ne jen počítač.

Uplatnění nachází všude tam, kde je text neustále předčítán a také měněn. Je to například na vlakovém nádraží, kdy počítač s využitím této aplikace předčítá, kam přijede daný vlak. Je to zjednodušení, jelikož kdyby takový systém neexistoval, musely by být nahrávány všechny možnosti příjezdů vlaků v celé republice. Díky tomuto systému všechny hlášení v celé republice vypadají téměř stejně.

### 3.2.6 Informační a komunikační služby uskutečňované po telefonu

Informační a komunikační služby uskutečňované po telefonu jsou poměrně perspektivní aplikační oblastí. Je to díky tomu, že mohou být uskutečňovány automaticky bez nutnosti lidské obsluhy i bez požadavku na speciální telefonní přístroje v kteroukoliv denní dobu.

Jedná se o služby, kdy uživatel zavolá na určité telefonní číslo, kde uslyší automat, který s ním komunikuje na základě toho, jaká klíčová slova uživatel vysloví [7].

Systém je využitelný např. při zjišťování určitých informací. Konkrétní aplikace byla vyvíjena na Katedře kybernetiky Západočeské univerzity v Plzni. Aplikace se jmenuje *Nádraží* a funguje tak, že člověk zavolá na telefonní číslo a chce zjistit odjezdy vlaků. Normálně plynule vede dialog s automatem na druhé straně telefonní linky a automat detekuje slova, která uživatel říká a pokud se objeví nějaké konkrétní slovo, na které má reagovat tak na něj reaguje podle předepsaného scénáře. Tento systém pracuje, aniž by u telefonu musel sedět člověk, tudíž není vázaný na pracovní dobu a může pracovat pořád, celý den [15].

### 3.2.7 Rozpoznávání řečníka

Tuto úlohu lze obecně definovat jako rozpoznávání lidí na základě jejich hlasu. Spočívá v tom, že máme záznam hlasu nějaké neznámé osoby a víme, za koho se tato osoba vydává. Úkolem je tedy ověřit, zda hlas neznámé osoby je dostatečně podobný hlasu člověka, za kterého se neznámá osoba vydává [1]. Většinou systém požaduje říci nějakou specifickou větu, na které poté ověří, o koho se jedná.

Tento systém je možno využít k přístupu do nějakého objektu nebo třeba k přístupu do nějakého informačního systému přes mobilní telefon.

### 3.2.8 Voice Browser

Pod termínem Voice Browser se skrývá nový výzkumný směr W3C, který je zaměřen na výzkum jazyka VoiceXML. Má sloužit pro procházení internetových stránek pomocí hlasu [26]. Procházení by bylo interaktivní, kdy uživatel internetových stránek by mluvil

a internetová stránka reaguje na jeho mluvu, buď pomocí předem připraveného zvukového souboru nebo přečtení určitých textových informací syntetizérem.

### **3.2.9 Uživatelem je handicapovaná osoba**

Systemy, které využívají hlasové moduly, jsou velice prospěšné pro handicapované osoby, ať už s pohybovými, popřípadě zrakovými obtížemi. Systemy, které jsou vytvářeny speciálně pro handicapované, mohou zprostředkovat zrakově postiženým lidem přístup k informacím, které by jim byly jinak nedostupné. Dále také mohou být tyto systémy užity pohybově postiženými lidmi jako hlasem řízená zařízení k ovládní kolečkových křesel, světel, televize apod.



## 4 TEORIE OVLÁDÁNÍ POČÍTAČE HLASEM

Proniknout do teorie ovládání hlasu počítačem je velice náročné, jelikož již samotný princip, kterým je ovládán počítač hlasem, je velmi komplikovaný. Pochopení veškerých zákonitostí a pravidel, kterými se tato oblast vědy řídí, vyžaduje rozsáhlé studium mnoha vědních disciplín, mezi nimi jsou statistika, přenos signálu, zpracování signálu a mnoho dalších odvětví fyziky, která jsou nezbytné k pochopení celé problematiky [28].

V této kapitole budou uvedeny základní poznatky k základnímu porozumění celého procesu, jakým ovládání počítače hlasem funguje.

### 4.1 Základní oblasti zpracování řeči

Zpracování řeči má v dnešní době velké uplatnění v řadě oborů. Avšak existují 3 základní úlohy, ze kterých všechny aplikace vycházejí. Jedná se o úlohy: syntéza řeči, rozpoznávání řeči a hlasové dialogové systémy.

#### 4.1.1 Syntéza řeči

Syntézou řeči myslíme proces, při němž se uměle vytváří lidská řeč. V případě počítačové syntézy se řeč uměle vytváří počítačem. Uměle vytvořená řeč si klade za cíl „zpřirozenit“ komunikaci člověka s počítačem a stát se tak rovnocenným partnerem tradiční vizuální komunikace [17]. Konečným cílem syntézy řeči je vytvářet řeč v takové formě a kvalitě, aby nebyla rozpoznatelná od řeči řečníka. Syntetická řeč by teda neměla působit jednotvárně, měla by znít přirozeně a její poslech by neměl unavovat ani vyžadovat zvýšenou pozornost. Syntetizérem řeči máme na mysli zařízení, které proces vytváření řeči provádí [20].

Syntéza řeči je obecně velice složitý proces, který je ve většině případů řešen tak, že je mu dopředu dána nějaká vstupní informace. Zpravidla to bývá informace o promluvě, která se má generovat. Taková informace je poté složena ze dvou základních částí, z fonetické a prozodické informace. Fonetická informace reprezentuje posloupnost hlásek a popisuje, jaká řeč se má vytvořit. Prozodická informace definuje průběhy základních prozodických charakteristik, kterými jsou melodie, časování a intenzity promluvy a popisuje, jak se má výsledná řeč vytvářet [17].

Nejčastější oblastí lidské řeči je syntéza řeči z textu. Tento způsob konverze textu na řeč patří mezi nejobecnější, ale také nejkomplikovanější úlohu počítačové syntézy řeči [19]. Základním smyslem této úlohy je schopnost systému automaticky převést libovolný psaný text (kniha, časopis, SMS zprávy apod.) na mluvenou řeč a to pokud možno v takové podobě a kvalitě, aby uživatel měl pocit, že text čte opravdu člověk s dobrým přednesem. Tudíž by nemělo být posláním takového systému na syntézu řeči, aby byl text přednesen monotónním počítačovým hlasem, ale aby systém uměl generovat přirozený lidský hlas, jehož poslech by neunavoval a nevyžadoval zvýšenou pozornost člověka [24].

Pokud požadujeme, aby syntéza řeči byla opravdu kvalitní a aby člověk došel k názoru, že poslouchaný hlas patří člověku, i když je generovaný počítačem, tak použitý syntetizér musí být navrhnut velice kvalitně.

#### 4.1.2 Rozpoznávání řeči

Rozpoznávání řeči znamená, že se podle zaznamenaných vzorků lidského hlasu rozpozná, co bylo řečeno. Tato úloha je předmětem výzkumu již padesát let a i když byl v této oblasti vykonán poměrně velký pokrok, tak je stále konstrukce systému, který bude schopen rozpoznat promluvu jakéhokoliv řečníka užívajícího libovolná slova, ještě poměrně vzdálenou budoucností.

Problémy jsou především v rozdílných intonacích a výslovnostech jednotlivých osob. Hlas jedné osoby se totiž dost liší od hlasu jiných osob, což znamená, že každý člověk má obvykle jinou barvu hlasu, jiný přízvuk, každý mluví jinak rychle apod. [17]. Dalším důvodem proč je to tak složité je fakt, že hlas jednoho člověka je sice stejný, ale může být jiný za odlišných podmínek, v odlišných situacích. Reálně tudíž skoro není možné, aby člověk vyslovil něco stejně v odlišné situaci [5]. Problémem je také akustické prostředí, tj. okolní šum, který se neustále mění. Například při vyšším šumu je pro systém rozpoznávání řeči daleko složitější rozpoznat začátky a konce slov.

Úloha rozpoznávání řeči je doposud nejvíce propracovaná pro anglický jazyk. Je to z toho důvodu, že tento jazyk není ohebný, na rozdíl třeba od češtiny. Ohebný jazyk totiž musí obsahovat daleko větší slovníky slov než jazyky neohebné.

Jak již bylo zmíněno, rozpoznávání řeči je závislé na člověku, který k systému promlouvá. Tyto systémy mají natrénovány hlas jednoho konkrétního řečníka anebo malé skupiny řečníků. Systémy, které jsou závislé na řečníkovi, dosahují pro danou osobu daleko lepší výsledky, ale jejich hlavním problémem je to, že řečník musí namluvit několik hodin nahrávek a taky to, že daný systém je použit pouze pro jednoho nebo malý počet osob.

Na druhou stranu existují systémy, které na řečníkovi závislé nejsou. Tyto systémy jsou založené na natrénování na hlasy stovek i tisíců řečníků. Tyto systémy nejsou poté limitované jen konkrétním řečníkem, ale může pracovat s téměř jakýmkoli řečníkem. Naproti tomu mají tyto systémy také jednu velkou nevýhodu oproti systémům závislých na řečníkovi, nedosahují totiž takové úspěšnosti v rozpoznávání řeči.

#### **Metody rozpoznávání řeči:**

##### *a) algoritmus HMM (Hidden Markov Model- Skryté Markovovy modely)*

Princip této metody modelování řeči vychází z představy o vytváření řeči. Při generování řeči člověkem si lze představit, že hlasové ústrojí je během krátkého časového intervalu (např. 10 ms) v jednom z konečného počtu stavů artikulačních konfigurací (např. nastavení pro vyslovení určité hlásky). V uvažovaném intervalu je pak hlasovým ústrojím produkován krátký signál, který závisí na stavu artikulačního ústrojí a může být popsán určitými spektrálními charakteristikami, jež jsou reprezentovány vhodnými příznaky [18].

#### b) Viterbiho algoritmus

Tato metoda se používá pro spojité rozpoznávání řeči, což je mnohem komplexnější a náročnější úloha než rozpoznávání izolovaných slov. Patří mezi metody založené na prohledávání stavového prostoru do šířky využívající výhodné vlastnosti stavového prostoru na úrovni slov a časového kroku, kterou je topologická uspořádanost grafu stavového prostoru podle časového kroku. Pro každé slovo (frázi) spočítá cena minimální cesty příslušným grafem. Nakonec se jako výsledek rozpoznávání vybere slovo či fráze, jehož minimální cesta je ze všech vyšetřovaných slov (frází) [18].

#### c) metoda DTW (Dynamic Time Warping- Dynamické borcení času)

Metoda vhodná pro rozpoznávání izolovaných slov (ovládání počítače pomocí hlasových příkazů), jelikož je zapotřebí mít ve slovníku uloženou modelovou nahrávku každého použitého slova. Metoda měří podobnost mezi dvěma nahrávkami slov a snaží se najít nejlepší shodu. Daným „borcením“ je myšleno, eliminace časových rozdílů mezi dvěma řečovými obrazy [18].

### 4.1.3 Hlasové dialogové systémy

Hlasové dialogové systémy dovolují uživatelům komunikovat prostřednictvím hlasu s počítačovými či internetovými aplikacemi, jako jsou databázové systémy a expertní systémy či systémy automatického řízení, ovládání a monitorování [17]. Účelem hlasových dialogových systémů je vytvořit rozhraní mezi počítačem řízenou aplikací a člověkem, který komunikuje s danou aplikací pomocí lidské řeči.

Mezi tyto systémy může patřit celá řada různých systémů, od poměrně jednoduchých systémů, které umí komunikovat pouze pomocí určitého malého souboru izolovaně vyslovených příkazů, až po relativně sofistikované systémy, které již umí s uživatelem vézt s uživatelem plynulou řeč.

Avšak problémem všech těchto systémů je v tom, že jsou pokaždé omezeny jejich aplikační oblastí, do které jsou nasazeny a nejde tudíž o přirozenou konverzaci člověka s počítačem na libovolné téma.

Tyto systémy v sobě totiž musejí obsahovat oba dva systémy předchozí, modul syntézy řeči i modul rozpoznávání řeči, a navíc musejí obsahovat modul, který zajišťuje řízení dialogu. Tudíž požadavky na hlasové dialogové systémy jsou daleko vyšší, jelikož nepracují pouze s jedním modulem, ale musejí spojit dohromady více složitých modulů a všechny moduly musejí spolu pracovat navzájem provázaně.

Pokud opravdu chceme, aby se systém hlasového dialogu stal plnohodnotným pomocníkem při práci s počítačem, tak daný systém by měl splňovat několik kritérií:

- měl by být uživatelsky příjemný a interaktivní

- neměl by na uživatele působit tak, že s ním komunikuje počítač, ale měl by mít pocit, že s ním opravdu komunikuje člověk
- měl by se uživatele zeptat na to, co potřebuje
- klasifikátor rozpoznávání řeči musí být spolehlivý, tzn., že by měl klasifikovat jednotlivá slova přesně, aby systém neměl zbytečné prodlevy a uživatel tak nemusel zbytečně opakovat to, co již řekl.

I přesto, že systémy hlasového dialogu nejsou zatím dokonalé, je jejich aplikační uplatnění v dnešní době poměrně rozsáhlé. Jsou používány především tam, kde mohou nahradit člověka a usnadnit tak práci celého systému. A i když budou splňovat systémy hlasového dialogu všechny výše vyjmenované kritéria, tak vždy půjde zatím jen o tematicky omezenou komunikaci, která bude v návrhu řízení dialogu prioritně soustředěna na úspěšné dokončení dané úlohy a v žádném případě nebude mít takový systém ambice zajistit plný rozsah lidské konverzace [17].

## 4.2 Jak se tvoří hlas

Každý jazyk existuje ve dvou základních podobách: mluvené a psané. Cílem obou podob je především sdělovat myšlenky a zprostředkovávat komunikaci mezi lidmi [23].

Starší z obou podob jazyka je podoba mluvená. Existovala dávno předtím, než se lidé naučili zaznamenávat různé věci písmem. I proto patří řeč mezi nejčastěji používanou formu komunikace mezi lidmi. Velkou výhodou řeči oproti písmu je to, že všichni účastníci rozhovoru se mohou současně věnovat jiným činnostem. Nevýhodou naopak je, že řečník svou myšlenku nikdy nezopakuje přesně jako v té podobě napoprvé [1].

### 4.2.1 Vytváření řeči člověkem

Pro vytváření řeči existuje v lidském těle několik skupin orgánů, které se souhrnně nazývají artikulační orgány.[17] Tyto orgány plní v lidském těle různé, spolu nesouvisející funkce.

Artikulační orgány jsou složeny ze tří základních ústrojí, dechové, hlasové a artikulační ústrojí.

#### **Dechové ústrojí**

Dechové ústrojí slouží jako zdroj energie pro samotnou řeč a je umístěno v hrudníku. Při nádechu dochází k pohybu vzduchu, který tak poskytuje zdroj energie pro řeč. Při výdechu potom v plicích vzniká výdechový proud vzduchu, který je základním materiálem pro tvorbu řeči. [16]

Vzduch je při výdechu z plic odváděn průdušnicí a pak prochází hrtanem, kde se modifikuje, a jako řečový signál je vypouštěn do okolního světa. To, jakou silou člověk vydechne vzduch, má vliv na sílu hlasu i na jeho výšku.

## **Hlasové ústrojí**

Hlasové ústrojí je uloženo v hrtanu. Z hlediska tvorby řeči jsou jeho nejdůležitější částí hlasivky.

Jestliže člověk mlčí, pak hlasivky drží hlasivkovou štěrbinu odkrytou, takže jí může bez odporu procházet vzduch k dýchání. Pokud chce člověk mluvit, musí hlasivky zaujmout tzv. hlasové postavení. Vzduch z plic postupuje z plic a v hrtanu se objeví překážka vytvořená hmotou hlasivek, které cestu vzduchu uzavřou [17]. Tyto stažené hlasivky jsou tudíž pod tlakem a začínají kmitat. V důsledku toho se vzduch z plic „rozdrobí“ a vzniká tzv. vzduchová vlna, kterou člověk vnímá jako zvuk.

## **Artikulační ústrojí**

Artikulační ústrojí je posledním ústrojím, které se podílí na tvorbě řeči. Jeho význam spočívá v tom, že umožňuje vytvářet velké množství různých zvuků, které charakterizují mluvený jazyk.

Skládají se z nadhrtanových dutin a také z artikulačních orgánů. Nadhrtanové dutiny se účastní tvorby řeči pasivně, kdežto artikulační orgány jsou aktivní, tzn., že svým pohybem mění velikosti nadhrtanových dutin. [17]

Mezi nejvýznamnější artikulatory patří především jazyk, rty a měkké patro. Nejdůležitější je určitě jazyk, díky tomu, že je pružný a přizpůsobivý a tudíž schopen tvořit mnoho tvarů. Právě variabilita jazyka vede k vytváření různých zvuků řeči.

### **4.2.2 Modelování vytváření řeči**

Při počítačovém zpracování řeči bývá užitečné modelovat procesy produkce i vnímání řeči a řeč následně reprezentovat pomocí parametrů zvoleného modelu. [16] Cílem samotného modelování vytváření řeči je nalézt matematické vztahy, které by mohly být využity pro reprezentaci akustických fyzikálních dějů spojených s touto produkcí, a to s požadavkem minimální složitosti a maximální přesnosti modelu.[16]

Díky tomu, že lidská řeč je poměrně dosti proměnlivá a záleží na mnoho okolnostech jak je řeč vnímána počítačem, je modelování samotné řeči složitý úkol a není divu, že zatím nebyl předložen odborníky na toto téma jediný univerzální model, který by respektoval všechny zákonitosti lidské řeči.

Modelování lidské řeči vychází z několika oborů a propojuje tyto jednotlivé obory. Obor fyziologie, který je důležitý pro pochopení vzniku lidského hlasu, dále fyzikální obory akustika a zpracování signálu, které slouží především k pochopení, jak se hlas přenáší a jak je vnímán posluchačem.

Důkladné pochopení všech těchto dílčích oblastí je nezbytné k tomu, aby se výzkum posunul dále v modelování lidské řeči a aplikaci těchto poznatků do oblasti jako je rozpoznávání řeči počítačem.

### **4.3 Problémy se zpracováním lidské řeči**

Jak už bylo popsáno výše, lidský hlas je velice komplikovaný a dokonalé rozpoznávání lidské řeči je pro počítač stále ještě problematické. Problémů, které omezují počítač v tom, aby bez problémů poznal, co člověk říká je hned několik.

V první řadě se jedná o problematiku větné skladby a skloňování slov. Druhým zásadním problémem jsou rozdílné prozodické vlastnosti každého uživatele.

#### **4.3.1 Problematika větné skladby a skloňování**

Tento problém se týká především jazyků, které jsou ohebné, takovým jazykem je i čeština. Pro neohebné jazyky, např. angličtina, je tento problém téměř úplně eliminovatelný, díky tomu, že tyto jazyky mají pevně daná pravidla pro tvorbu vět a v těchto jazycích nedochází ke skloňování podstatných jmen, ani k časování sloves.

Čeština je díky tomu, že větná skladba je poměrně dosti volná pro modelování a rozpoznávání lidské řeči značně náročná [9]. A tudíž pro systém, který pracuje s rozpoznáváním češtiny, existují de facto dvě možnosti jak zajistit, aby úspěšnost byla na srovnatelné úrovni se systémem, který pracuje pouze s angličtinou. První možností je mít daleko obsáhlejší databázi slov, která bude obsahovat slova ve všech jeho obměnách, což při současném stavu počtu českých slov, který se odhaduje na 200 000 a při všech obměnách těchto slov je prakticky nemožné. Druhou možností je mít poměrně dokonalý systém umělé inteligence, který pozná, že se jedná o dané slovo, avšak jinak vyskloňované. Systém, který by toto uměl, dnes ještě neexistuje.

V současné době se problematika větné skladby a skloňování řeší jakousi mutací obou těchto řešení. Neboli existuje rozsáhlá databáze slov a systém částečné umělé inteligence, který „domýšlí“ o jaké slovo se jedná [21]. Úspěšnost skloubení těchto dvou systémů není stoprocentní, ale s rozvojem umělé inteligence se bude dozajista zvyšovat.

#### **4.3.2 Problematika prozodických vlastností**

U problematiky větné skladby a skloňování slov nás zajímalo, co člověk říká, neboli význam řeči. U problematiky prozodie nás zajímá, jak člověk určité slova pronáší, neboli jakým způsobem.

Termínem prozodie se společně označují takové vlastnosti řečového signálu, které souvisí především s frekvencí základního hlasivkového tónu (výškou hlasu), intenzitou (hlasitostí) a časováním. [5] Další vlastnosti jsou především rytmus a rychlost řeči.

Prozodie jako taková význam slov nepřenáší, ale slouží spíše jako doprovodná složka řeči, která pomáhá posluchačům pochopit význam přenášené zprávy. Prozodie slouží například k rozpoznání věty oznamovací či tázací, k rozpoznání emocí mluvčího (rozčilení či radost), ale může také vyjadřovat postoj řečníka k myšlence, kterou vyjadřuje.

### **Časování řeči**

Časování patří mezi důležité prozodické vlastnosti řeči.[5] Záleží především na tempu řeči, přízvuku a na používání pauz.

Čím vyšší (rychlost) řeči, tím obvykle je řeč méně srozumitelná a počítač s takovou řečí poté hůře pracuje. Stejný problém nastává, pokud mluvčí mluví s určitým přízvukem, i poté je pro počítač těžší rozpoznat jednotlivá vyslovená slova. Neméně důležitým je správné používání pauz v řeči. Pauzy slouží k odlišení jednotlivých slov, a tudíž zvyšují srozumitelnost celé řeči.

### **Intonace**

Intonací je v tomto případě myšlena změna hlasivkového tónu, někdy též nazývána melodii řeči. Ve většině jazyků nemá intonace vliv na význam slov, ale slouží pouze k modifikaci jejich akustické realizaci. [5] Naopak v jazycích asijských a afrických je intonace daleko důležitější, jelikož význam slov je měněn v závislosti na intonaci, s jakou bylo slovo vysloveno.

Ve většině jazyků je intonace brána jen jako tzv. informátor o citovém rozpoložení mluvčího a také jako informátor o oddělení jednotlivých větných celků.

### **Intenzita**

Intenzita neboli hlasitost, patří mezi nejméně důležité aspekty řeči. Ve většině případů závisí na emocionálním stavu řečníka.

## **4.4 Okolí mluvčího**

Důležitým faktorem při rozpoznávání řeči počítačem hraje prostředí, ve kterém se mluvčí vyskytuje. Je zřejmé, že pokud je v prostředí, ve kterém počítač rozpoznává řeč, naprostý klid, počítač nebude mít s rozpoznáváním řeči sebemenší problém [2]. Pokud však v prostředí bude nějaký šum, bude to rozpoznávání ztěžovat, jelikož se do rozpoznávání budou přimíchávat další okolní zvuky a nebude jednoznačné.

Tento problém je v dnešní době poměrně lehce eliminován tím, že aplikace na rozpoznávání řeči nejprve na začátku spuštění projde kalibrací. Neboli na určitý čas aplikace zaznamenává šum v okolí. Z tohoto celkového šumu je poté vypočtena určitá střední hodnota šumu v okolí. Tato střední hodnota je již při samotném rozpoznávání odečítána od zaznamenané řeči mluvčího a je tudíž používán poměrně přesný záznam řeči mluvčího.

Samozřejmě, že pokud aplikace pracuje po delší dobu je pravděpodobná změna šumu v prostředí. Tento problém je v dnešní době také již překonaný, díky tomu, že aplikace má nastavenou automatickou kalibraci po určitém čase. To znamená, že střední hodnoty šumu jsou stále aktuální ve vztahu k danému prostředí.

## 5 PROGRAMY NA OVLÁDÁNÍ POČÍTAČE HLASEM

V dnešní době již existuje poměrně široká škála programů a aplikací, které umožňují uživatelům ovládat počítač pomocí hlasových příkazů. Ve většině případů se jedná o programy, které lze aplikovat podle uživatelského přání na široké spektrum příkazů ovládající celý počítač, avšak některé jsou využitelné pouze pro konkrétní aplikační oblast.

Mnoho z těchto programů patří mezi tzv. „pionýrské programky“ v oblasti ovládání počítače hlasem. Je tím myšleno to, že dané aplikace vznikly jako takové pokusy, jestli vůbec něco takového půjde vytvořit a s jakou funkcionalitou celého programu. Tyto programy jsou již poměrně staršího data vytvoření, avšak ten základní požadavek na takový program většina splňuje, ať už pomocí předvolených či uživatelem vytvořených příkazů ovládat počítač hlasem. Musíme však konstatovat, že většina těchto aplikací je pro běžné užívání téměř nepoužitelná a uživatelé je používají spíše jako odreagování a hračku, než jako usnadnění ovládání počítače.

Na druhou stranu již dnes existují poměrně sofistikované programy na ovládání počítače hlasem, které mají za sebou dlouhý vývoj a dnešní podoba těchto programů je již na profesionální úrovni. Možnosti uplatnění v běžném jsou rozsáhlé a běžně používané v nejrůznějších oblastech, např. zdravotnictví, vojenství, telekomunikace a jiné. Ve většině těchto aplikací jsou již moduly na rozpoznávání hlasu plně integrovány a jejich použití je již zaběhlé.

### 5.1 MyVoice

Program MyVoice byl vyvinut na Technické univerzitě v Liberci v Laboratoři počítačového zpracování řeči. Původně se jednalo o vědecký projekt, který měl pomoci handicapovaným lidem v přístupu k počítačové technice. Dnes se jedná o software, který je licencován a komerčně využíván [15].

Program je určený pro operační systém Windows a dokáže ovládat jakýkoliv program v počítači [12]. Sám obsahuje velké množství příkazů, které jsou nejčastěji v počítači vykonávány. Avšak uživatel si může sám vytvořit své vlastní příkazy podle toho, jak mu budou vyhovovat.

Dva základní směry využití programu:

- 1) ovládání počítače pomocí hlasových příkazů- pomocí předem připravených nebo uživatelem definovaných příkazů ovládat jakýkoli program v počítači. Dokáže ovládat také akce, které jsou realizovány pomocí klávesnice či myši.
- 2) diktování textu- umožňuje diktovat text do textových editorů a jiných oblastí kde je text používán. Umí rozpoznat diktování samostatných slabik i jednotlivých



slov, avšak tady je omezení v malém, předefinovaném slovníku nejčastěji používaných českých slov (kolem 10 000). Uživatel si však může zvolit své vlastní slova k diktování.

Výhodou celého programu je především to, že u diktování si uživatel může pod jediný příkaz schovat celé věty a diktování je tudíž o dost rychlejší a efektivnější. Další výhodou je určitě to, že dokáže reagovat na český jazyk, což jiným programům chybí. Na druhou stranu je nevýhodou, že program je stále závislý na mluvčím a na vnějším prostředí, neboli program správně reaguje jen na jednu osobu a okolní prostředí výrazně ovlivňuje přesnost rozpoznávání příkazů.

## 5.2 JetVOICE

Aplikace JetVOICE byla vyvinuta v rámci maturitní práce na SPŠE v Pardubicích v roce 2001. V roce 2002 dokonce aplikace vyhrála Středoškolské odborné činnosti jako nejlepší studentský projekt.

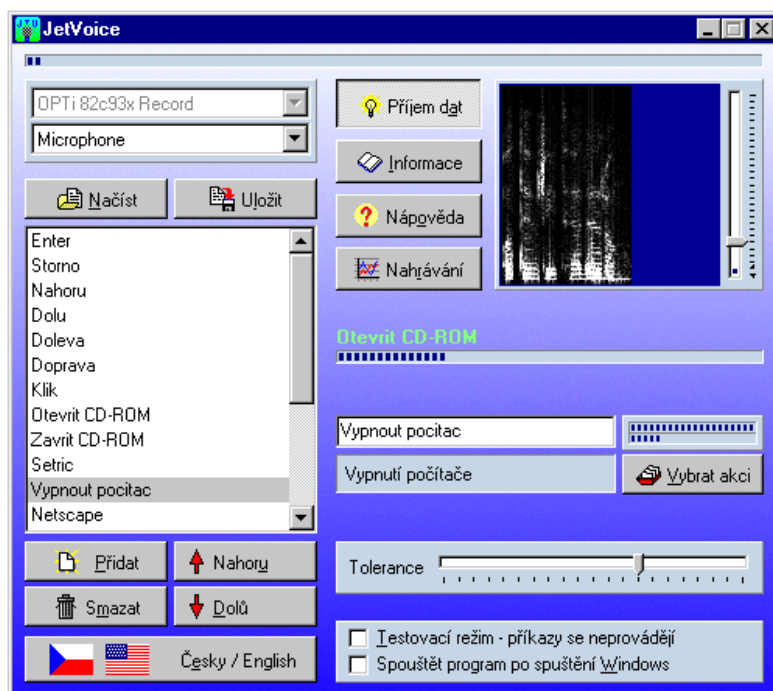
Jedná se také o program na ovládání počítače hlasem. Program je pro normální uživatele licencován pod shareware licenci, která má pouze jedno omezení, a to, že při každém spuštění programu musí uživatel aktualizovat časový klíč z webových stránek, jehož platnost je 8 a půl dne. Tato verze je pouze pro nekomerční využití, a pokud chce uživatel používat program komerčně, musí si domluvit s autorem individuální cenu používání. Tato nekomerční verze by měla v brzké době přejít na verzi, která bude již zcela zdarma i pro komerční využití. Termín vydání této verze však ještě nebyl stanoven.

Program v sobě nemá žádné předvolené příkazy a uživatel si může všechny příkazy vytvořit vlastním nahráním. Obsahuje velké množství akcí, ke kterým může uživatel přiřadit vlastní hlasový povel, ale také může veškeré akce různě kombinovat, takže v konečném důsledku má na výběr z velké škály možných akcí.

Velkou výhodou je určitě přizpůsobení mikrofonu okolnímu prostředí. Po spuštění programu je uživatel automaticky vyzván, aby provedl kalibraci mikrofonu. Tím dojde k tomu, že program si uloží do paměti okolní hluk a ten poté od hlasových povelů odečítá a zvyšuje tím efektivitu celého procesu. Další nespornou výhodou je nastavení tolerance přijímání povelů. Je tím myšleno to, že pokud snížíme toleranci, bude systém reagovat pouze na přesně vyslovené povely tak, jak byly nahrány a nijak jinak. Pokud naopak toleranci zvýšíme, může být daný povel být vysloven ne úplně přesně, ale s určitými odchylkami. Tím je částečně eliminován problém toho, že systém reaguje pouze na uživatele, který příkazy nahrál, ale příkazy mohou vyslovovat i jiní uživatelé a systém bude reagovat správně.

Z druhé výhody, která byla ukázána, plyne i jedna značná nevýhoda. Pokud bude tolerance nastavena příliš vysoko a budou nahrané dva podobné příkazy, tak se ty příkazy budou dost plést. Tento problém jde taky eliminovat a to tím, že se nastaví tolerance přibližně vyváženě. To zapříčiní menší chybovost, ale také, že uživatelé musí vyslovit

příkaz hodně podobně, avšak pokud uživatel ví, že má vyslovovat příkaz přirozeně bez nějakých emocí, tak je to taky celkem snadno překonatelný problém.



Obr.2 Prostředí programu JetVoice

### 5.3 Respondings Heads

Tento poměrně jednoduchý program umožňuje rozpoznávání a vykonávání příkazů, které si uživatel předvolí. Po spuštění aplikace se na obrazovce objeví animovaná hlava, která reaguje na uživatelské příkazy.

Program má v sobě zahrnutých pět základních akcí, kterými uživatel může ovládat příkazy. První akcí je akce *Script*. Tato akce otevře předvolený skript ve Visual Basic. Druhou akcí je akce *Open*, která jak již název napovídá, otevře předvolený soubor nebo program. Třetí akce je spíše taková zábavná, nazvaná *Speak*, kdy uživatel něco napíše do předvoleného okna a program mu na to něco odpoví. Čtvrtou akcí je akce *Type*, která pod sebou skrývá psaní textu, kdy uživatel hláskuje jednotlivá písmena. Poslední akcí je akce *Read*, která přečte vybraný text.

Celá aplikace funguje na rozpoznávacím enginu od MS Speech, který je již nainstalován v operačním systému Windows. Problém s tímto enginem je, že umí rozpoznávat pouze angličtinu a pro jiné lokalizace je tento engine z operačního systému vynechán [14]. Proto nemohu plnohodnotně hodnotit tento program, jelikož sice jsem ho nainstaloval a chtěl vyzkoušet, ale vůbec nereagoval na mé příkazy, i když jsem používal příkazy v angličtině.

Obecně bych tento program bral pouze jakýsi prostředek na hraní si s počítačem, a ne jako plnohodnotný program, kterým jde ovládat počítač hlasem. Jedinou výhodou

programu je to, že je freeware, ale osobně si nedokáží představit, kdyby tento program nebyl zdarma, kdo by si jej koupil.



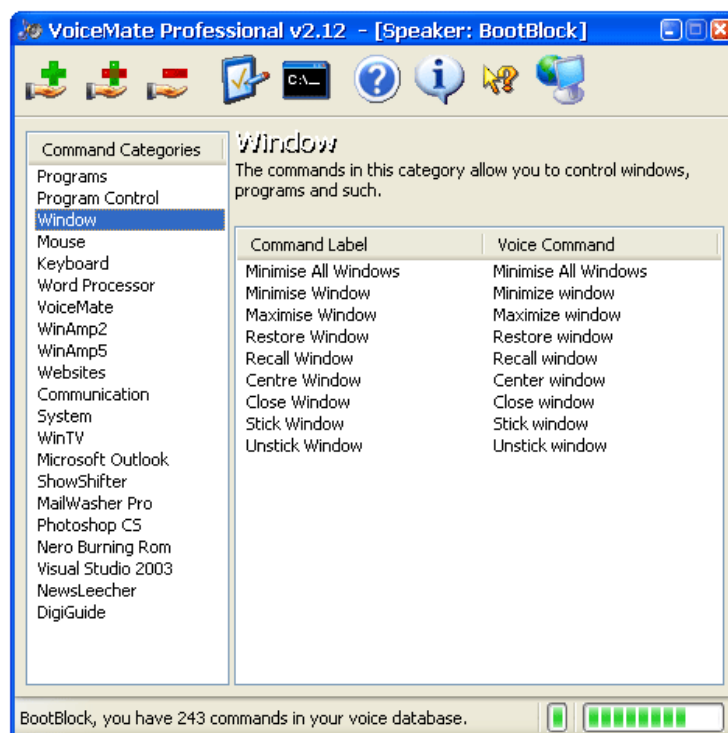
Obr.3 Ukázka programu Respondings Heads (zdroj: Návod k instalačnímu souboru)

## 5.4 VoiceMate Professional

Tato aplikace z roku 2006 umožňuje uživatelům také ovládat počítač hlasem. Na internetu je k dispozici pouze její shareware verze na 14 dní. Aplikace je napsána kompletně v programovacím jazyku Visual Basic [25].

Program umožňuje uživatelům používat již předvolené příkazy, kterých je asi kolem 70. Dále také umožňuje vytvořit si vlastní povely s přiřazenou operací, co má počítač udělat. Tento způsob je však poměrně dost krkolomný, jelikož tyto příkazy se dle návodu mají vytvářet přímo napsáním programovacího kódu VB. Program se mi sice podařilo nainstalovat, ale bohužel nefungoval, tak nemohu posoudit, jak je náročné vytvořit nějaký vlastní příkaz. Ale vzhledem k faktu, že daný program nemusí používat jen člověk, který umí programovat ve VB, ale může jej chtít ovládat i obyčejný člověk bez znalostí programování, tak se jedná o dost nešťastně zvolený způsob přidávání příkazů.

Program obecně působí na vlastních internetových stránkách dost profesionálně, především díky tomu, že v popisu programu je napsáno, že se jedná „o nejlepší program na světě, který umí ovládat počítač hlasem“, ale jeho použití je dle mého názoru jen pro malou skupinu lidí, kteří umí programovat ve VB. Navíc, když jsem program nainstaloval, tak se mi jej nepodařilo zprovoznit z toho důvodu, že mi psal, že již vypršela zkušební doba 14 dní. Moc profesionálně také nepůsobí fakt, že nikde není uvedeno, kolik stojí plná verze programu a jestli také obsahuje nějaké funkce navíc. Jelikož ale poslední aktualizace pochází z roku 2006, tak zřejmě nejde o program, který je masivně používán a proto je potřeba jeho aktualizace.



Obr. 4 Prostředí programu VoiceMate Professional

## 5.5 Pearl-box Voice

Tento malý program je určený pro operační systém Linux [10]. Opět zvládá plnohodnotně plnit příkazy, které si uživatel sám nadefinuje.

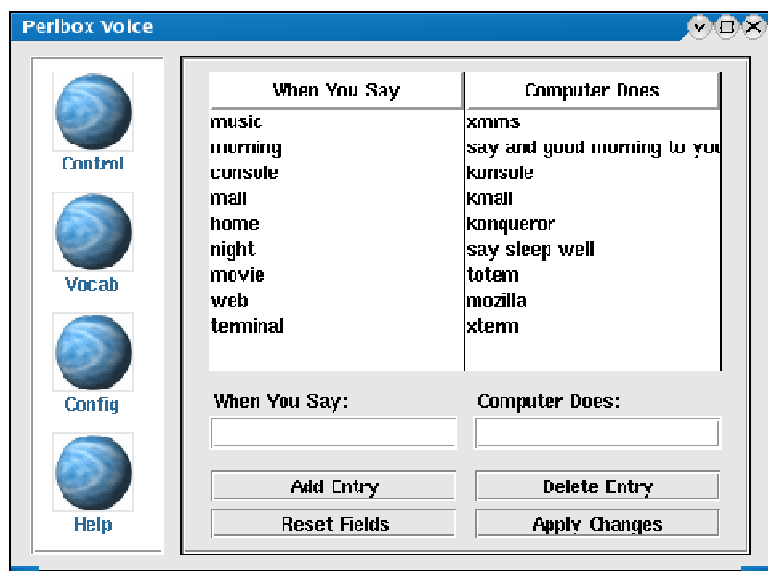
Samotné ovládání programu je poněkud jiné než u ostatních programů na ovládání počítače hlasem. Prvním důležitým krokem pro správné fungování programu je počáteční nastavení programu, bez něhož je program prakticky nepoužitelný. Nejdůležitější je nastavit tzv. klíčové slovo. Toto slovo slouží k ověřování příkazů uživatelem, neboli uživatel řekne příkaz, ale operace se neuskuteční do té doby, než je příkaz potvrzen klíčovým slovem. Toto opatření slouží především k tomu, že pokud máme příkaz, kterým je běžně používané slovo v mluvě člověka, tak by se daný příkaz spouštěl sám od sebe neustále.

Tento program má o dost rozdílné zadávání vlastních příkazů než je tomu u ostatních programů. U tohoto programu se musí daný příkaz nejdříve napsat do připraveného editoru a až potom tento příkaz vyslovit. To je poměrně neobvyklý přístup a dle mého názoru zdoluhavý proces a ve skutečnosti nechápu jeho význam. Dalším omezením je určitě to, že vyslovované slovo musí člověk vyslovovat s anglickou výslovností, jinak daný příkaz nebude správně fungovat.

Sám jsem program nezkoušel, ale dle různých komentářů k tomuto programu na internetu mohu konstatovat, že program funguje s poměrně velkou efektivitou a úspěšností vykonávání příkazů. Jedním z nedostatků programu je to, že pokud máme podobné příkazy a nevyslovujeme je dostatečně zřetelně, tak dochází ke špatné

interpretaci příkazu. Tento problém je však lehce eliminován tím, že jsou nastaveny různé příkazy, které nejsou nějak podobné. I nastavení je poměrně složité, ale pokud se uživatel tímto vším probere, tak má v ruce opravdu kvalitní nástroj k ovládní počítače hlasem.

Vývoj tohoto programu byl ukončen již v roce 2004 z neznámého důvodu, a tudíž program již nikam neroste a nerozvíjí se.



Obr. 5 Prostředí programu Pearlbox Voice (zdroj: <http://www.abclinuxu.cz/>)

## 5.6 Tyle

Tyle patří mezi novější aplikace v oblasti rozpoznávání počítače hlasem. Jedná se o malý program, kterým může uživatel spouštět různé programy a soubory v počítači. Z internetu lze stáhnout freeware verze programu, kde však uživatel může nahrát maximálně 15 příkazů. V placené verzi je počet příkazů neomezený [11].

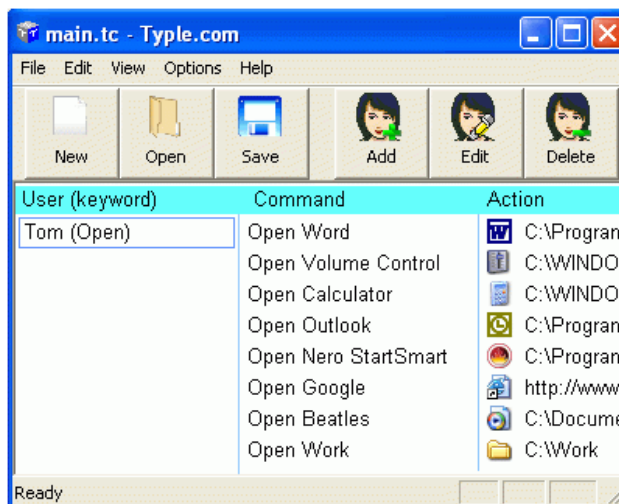
Program je dostupný v 10 světových jazycích, avšak čeština mezi nimi chybí. V aplikaci nejsou žádné dopředu předvolené příkazy, tzn., že uživatel si všechny musí vytvořit sám. Vytvoření je poměrně jednoduché, avšak princip rozpoznávání příkazů funguje trochu jinak než u podobných programů. Nejdříve se musí vytvořit tzv. klíčové slovo, které značí jaký příkaz se bude vykonávat, např. příkaz *Open* bude značit, že se bude jednat o skupinu příkazů, které budou něco otevírat. Tudíž uživatel musí nejdříve vyslovit příkaz *Open* a hned za ním příkaz, co konkrétně se má otevřít.

Pokud program spustíme, tak na pozadí běhu počítače program naslouchá a analyzuje slova, která byla vyslovena, a porovnává je s příkazy, které má nainstalované ve své databázi. Výhodou programu je to, že neustále analyzuje nově nainstalované programy, neboli pokud dojde k nainstalování nového programu v počítači, program uživatele sám upozorní, jestli nechce přidat k danému programu, který ho bude spouštět.

Program jsem sám nainstaloval a vyzkoušel a musím říct, že práce s programem je opravdu jednoduchá a intuitivní, avšak funkcionality celého programu je na hodně špatné

úrovni. Příkazy, které jsem vytvořil, ve většině případů nefungovaly, a když už se daný příkaz provedl, tak otevřel úplně jiný soubor nebo program než měl být spuštěn.

Oproti jiným vyzkoušeným programům na ovládání počítače hlasem program hodně pokulhává ve funkcionalitě, možná je to tím, že všechny algoritmy, které používá na rozpoznávání hlasu, nejsou vytvořené odborníky na rozpoznávání hlasu, ale pouze programátory, kteří do této problematiky příliš nevidí.



Obr. 6 Prostředí programu Typle

## 5.7 e-Speaking

e-Speaking je dalším z řady menších programů, které umějí ovládat počítač za pomoci hlasových příkazů. Program pracuje pod operačním systémem Windows XP, na novější operační systém dosud nebyl přizpůsoben. Poslední aktualizace programu je z poloviny roku 2007. Je sice licencován jako freeware, ale pokud chce uživatel program používat déle než 30 dní musí si jej již zaplatit.

Program je velice intuitivní a jednoduchý na ovládání prakticky pro všechny uživatele bez ohledu na jejich počítačové schopnosti. Program v sobě obsahuje více než 100 již předvolených příkazů, které se mohou libovolně kombinovat a upravovat. Těchto 100 předvolených příkazů obsahuje téměř všechny operace, které lze v počítači provést, např. příkaz zpět, odstranit, posunout dolů apod. Díky této široké škále příkazů lze provést téměř veškeré operace hlasem.

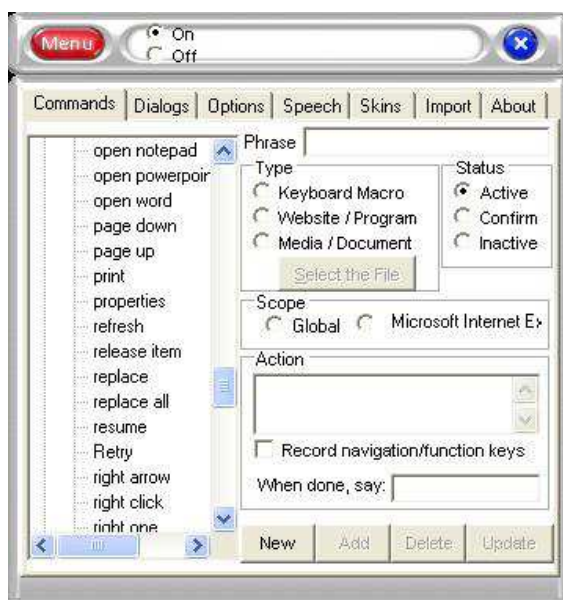
Nevýhodou může pro některé uživatele být vysoký počet těchto příkazů, jelikož při běžném používání si nemusejí veškeré tyto příkazy zapamatovat. To lze částečně eliminovat tím, že si pod daný příkaz zvolí svůj vlastní hlasový povel, který se jim bude dobře pamatovat.

Program je vystavěn na technologii rozpoznávání hlasu v Microsoft Windows Speech Accessibility, ale jelikož tato technologie je obsažena pouze pro anglickou lokalizaci Windows, tak se mi ji nepodařilo osobně vyzkoušet. Ale zhlédl jsem na internetu několik videí, na kterých je funkcionalita programu velice dobře ukázána. Uživatel prochází

internet i celý počítač pouze díky tomuto programu a vůbec k tomu nepoužije klávesnice nebo myš. Tento program se dokonce umí přihlásit na email, tzn., že umí rozpoznávat jednotlivá písmena či slova a přepisovat je do textového editoru.

Program také obsahuje jednu důležitou funkci, a tou je zapnutí tzv. avatara. Touto funkcí je myšleno to, že na obrazovce se objeví fotka člověka a ta, pokud je příkaz správně rozpoznán, tak uživateli sdělí jednoduchou odpověď, že příkaz byl rozpoznán. Tudíž uživatel vždy ví, jestli daný příkaz byl rozpoznán a jestli bude vykonán. Nemusí tedy čekat, jestli časová doba po vyslovení příkazu je jen prodleva počítače s provedením příkazu, nebo jestli program daný příkaz správně nerozpoznal. Tato výhodná funkce má jeden malý nedostatek, a tím je to, že vzhled lidské hlavy na obrazovce neodpovídá typu hlasu, kterým program odpovídá. Fotky lidí jsou vybrány tak, že jsou to mladí, krásní lidé, ale hlas, který je k nim přidán, vůbec nekoresponduje s lidmi na fotkách, jelikož je to hodně strojový hlas, který nezní vůbec jako lidský.

I přes některé malé nedostatky se program jeví jako poměrně sofistikovaný, díky kterému uživatel dokáže ovládat počítač hlasem.



Obr. 7 Prostředí programu e-Speaking (zdroj: [www.e-speaking.com/](http://www.e-speaking.com/))

## 5.8 Dragon Naturally-Speaking

Je program na rozpoznávání lidské řeči, který je v určitých etapách vyvíjen od roku 1982. Obrovským kladem tohoto programu je použití na více platformách, jak na Windows, tak i na Mac.

Program je použitelný pro tři základní typy úlohy. První úlohou je diktát, neboli uživatel může tento program využít k diktování souvislého textu do textového editoru. Druhou úlohou může být převod textu do zvykové nahrávky, tzv. text-to-speech. Úloha funguje tak, že do aplikace je nahrán dokument, který má být převeden do mluvené podoby a spustí se konverze. U této úlohy lze zvolit, jaký hlas celý text bude číst. Lze

vybrat buď mužský, ženský nebo dětský. Třetím typem úlohy, který program zvládne udělat je vykonat příkaz, který je přiřazen k určitému hlasovému povelu. Vývojoví pracovníci tento program dotáhli tak daleko, že program umí i nastavit např. velikost a barvu písma v Microsoft Wordu. Další zajímavou funkcí může být ovládání Microsoft Excel pomocí hlasu, vkládání různých funkcí, vytváření grafů apod.

Program je opravdu na profesionální úrovni, ale to vše je vykoupeno poměrně dost vysokou cenou, která je v *Home* verzi 100 \$. Využití programu je v současné době na naše poměry dosti rozsáhlé. Je používán v oborech jako je zdravotnictví, a to především pro diktování zdravotnických dokumentací, které v určitých případech mohou být velmi dlouhé a ušetří se tak především čas spojený s psaním zprávy. Další obor, který tento program využívá je státní sféra, také k diktování textu a vytváření mluvených nahrávek z textů. V neposlední řadě se program používá ve vojenství, ve vzdělávání, finančnictví, v oblasti práva apod.

Dragon NaturallySpeaking patří mezi špičku v oblasti ovládání počítače hlasem. Je to dáno především dlouhým vývojem celého programu a také tím, že na vývoji se podílí opravdu špičkoví vědci z oblasti rozpoznávání hlasu. Program ukazuje, že již lze plnohodnotně ovládat počítač za pomoci lidského hlasu, a že už to není jen idea vědců a autorů sci-fi.

## 5.9 Rozpoznávání hlasu ve Windows

První náznaky toho, že nejvyužívanější operační systém Windows by mohl být ovládan lidským hlasem, se objevili kolem roku 2000. Původně byl tento program přidán pouze k balíku Microsoft Office jako pomocník při diktování souvislého textu, ale od nástupu operačního systému Windows Vista se stal tento program standardním vybavením operačního systému pro anglickou lokalizaci [6].

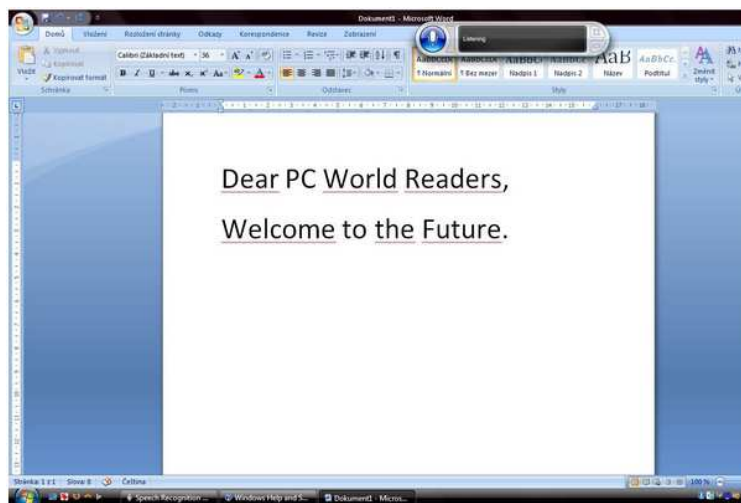
Největším problémem tohoto ne příliš využívaného doplňku je především to, že ho lze využívat jen na operačním systému, který má anglickou lokalizaci. V poslední verzi systému Windows 7 je doplněk i v lokalizaci francouzské, německé, španělské a čínské. Pro naši českou lokalizaci se doplněk do budoucna nepřipravuje kvůli náročnosti českého jazyka.

Program v sobě obsahuje již vytvořené základní příkazy pro otevírání, zavírání a další jednoduchou práci s programy [4]. Uživatel si sám může vytvořit své příkazy. Algoritmus, který je v programu využíván pracuje velice spolehlivě, tzn., že program nepotřebuje mít dopředu analyzovaný hlas mluvčího a přesto provádí příkazy s velkou úspěšností.

Pro diktování a úpravu textu v Microsoft Wordu je program opravdu velkou pomůckou, jelikož úspěšnost rozpoznání je téměř 100 %. Velkou výhodou programu je to, že je zdarma v operačním systému, jelikož pokud by byl distribuován jako samotně prodejný program zřejmě by jeho cena byla nemalá. Problém české lokalizace je velice



lehce odstranitelný tím, že se nainstaluje balík anglické lokalizace a může program využívat i český uživatel.



Obr. 8 Prostředí programu na rozpoznávání hlasu ve Windows

## 5.10 Zhodnocení programů

V této kapitole jsem vybral různé programy, které umějí ovládat počítač hlasem. Od dob kdy tyto programy teprve začínaly, uplynulo již poměrně hodně času a jejich vývoj hodně pokročil dopředu.

V dnešní době již existuje poměrně dost programů, které zvládají ovládat počítač hlasem. Některé jsou spíše na hraní, které ukazují, že vůbec lze ovládat počítač hlasem, ale jejich funkcionality není velká a spolehlivost rozpoznávání příkazů a mluveného slova je někdy hodně malá. Na druhou stranu již dnes existují aplikace, kterými opravdu lze plnohodnotně ovládat počítač hlasem a komunikovat s ním. Tyto programy jsou určitě velkým příslibem, že ovládání počítače hlasem se stane běžnou součástí práce s počítačem a ne jen doplňkem alternativního ovládání počítače.

## 6 OVLÁDÁNÍ MOBILNÍHO ZAŘÍZENÍ HLASEM

Mobilní telefon v současné době patří k osobnímu vybavení téměř každého člověka na světě. Mobilní telefon již není jen přístroj na volání a posílání SMS zpráv, ale stále častěji se z něj stává pomocník při práci, při trávení volného času a nemálo lidí považují svůj mobilní telefon jako svou nejosobnější věc.

Proto i vývojoví pracovníci firem, které produkují nové mobilní telefony, se snaží o to, aby práce uživatele s mobilním telefonem byla co nejjednodušší a také co nejefektivnější. V minulosti bylo největším hitem u mobilního telefonu hlasové vytáčení, které však v současné době nahradilo ovládání téměř celého telefonu pomocí hlasu. Další pokrokovou technologií v oblasti mobilních telefonů je určitě ovládání hlasem GPS navigace, která je zabudovaná přímo v mobilním telefonu.

### 6.1 Hlasové vytáčení

Již dlouhou dobu obsahují mobilní telefony funkci hlasového vytáčení. Princip této, dle mého nepříliš využívané funkce je ten, že uživatel mobilního telefonu si k jednotlivým položkám ve svém adresáři namluví povely, které po vyslovení povelu automaticky vytočí přiřazený kontakt k danému povelu. Tato funkce ze začátku uvedení do jednotlivých mobilních telefonů byla poměrně převratnou funkcí, ale po začátečním vyzkoušení toto nadšení trochu vyprchalo a stalo se tzv. funkcí, kterou využívá velmi malé procento lidí.

Důvod, proč se hlasové vytáčení nestalo tak převratnou funkcí, jak se předpokládalo, je ten, že hlasový modul, který byl v mobilních telefonech používán, nebyl příliš kvalitní a v mnoha případech se stávalo, že se uživateli vytočilo nesprávné telefonní číslo. Dalším důvodem bylo také dvojí spouštění funkce hlasového vytáčení. Bud uživatel musel nejdříve funkci ručně zapnout a potom teprve říci daný povel. To potom mohl sám vytočit ručně dané telefonní číslo, anebo měl uživatel tuto funkci zapnutou neustále, ale potom se velmi často stávalo, že telefon analyzoval zvukový podnět, který byl podobný předvolenému povelu a začal někomu volat a uživatel mobilního telefonu vůbec nevěděl, že někomu volal.

Hlasové vytáčení si však oblast použití našlo, a to především u řidičů automobilů, kdy řidiči, pokud musí v automobilu telefonovat je pro ně jednodušší vytočit telefonní číslo pomocí hlasu a soustředit se při tom na řízení automobilu. Celkově nemůžeme říci, že hlasové vytáčení bylo zlomovým bodem pro vývoj mobilních telefonů, ale určitým způsobem přispělo k lepšímu pochopení ovládání mobilního telefonu hlasem a tudíž i ke kvalitě dnešních funkcí ovládání mobilního telefonu hlasem, které jsou hojně používány.

### 6.2 Hlasové vyhledávání

V poslední době poměrně hojně využívanou funkcí je hlasové vyhledávání. Tato funkce přímo souvisí s rozvojem a rozšířením internetu v mobilních telefonech. Používá se především kvůli tomu, že mobilní telefony a celkově mobilní zařízení nedisponují tak

kvalitními a dobře ergonomicky řešenými vstupními možnostmi jako stolní počítače nebo notebooky. Další rozdíl je v tom, že mobilní telefony, tablety atd. jsou používány v jiných situacích (ve spěchu, v tramvaji, při chůzi městem apod.), kdy je použití hlasového vstupu více uživatelsky přívětivé [3].

Hlasové vyhledávání funguje pouze na mobilních telefonech s operačním systémem Android, tzn., že je z dílny společnosti Google. Název této aplikace je Google Search a je pro mobilní telefony distribuována zcela zdarma. Princip celé aplikace je velice jednoduchý, kdy k tomu, aby se začalo vyhledávat pomocí hlasu, se musí nejdříve říct klíčové slovo, které aplikaci upozorní, že se bude vyhledávat. Klíčové slovo je v tomto případě SEARCH, ale uživatel si jej podle svého uvážení může změnit. Poté již uživatel vysloví hledané slovo či skupinu slov a aplikace vyhledá na internetu odpovídající výsledky. Výhodou celé aplikace je v našich podmínkách to, že podporuje vyhledávání v českém jazyce, tzn., že umožňuje vyhledávat včetně veškerých interpunkčních znamének, které čeština používá.

Po technologické stránce má hlasové vyhledávání velice robustní řešení. To se projeví pokud narazíme na počet slov, které musí Google Search evidovat ve své databázi. Např. české hlasové vyhledávání poskytované googlem má databázi "v rozsahu více než milionu (...) slov"[6]. Takovou databázi je nejdříve nutné vybudovat. Google k vybudování této databáze tak trochu využívá samotné uživatele, kteří tuto aplikaci využívají. Google Search si do své databáze ukládá veškeré vyhledávání, které uživatelé provádějí a tím se neustále zvětšuje jeho databáze. Tím, že databáze je již poměrně rozsáhlá, tak se také zpřesňuje rozpoznávání jednotlivých slov a čímž se aplikace neustále vylepšuje.

Samotná aplikace Google Search umí také spouštět určité aplikace v mobilním telefonu, kdy se pouze zvolí jakékoliv jiné klíčové slovo místo slova *SEARCH*. Celé ovládání mobilního telefonu se tak poměrně dost zrychlí a zefektivní.

Mobilní telefony s operačním systémem Android, které již využívají poměrně kvalitní systém na rozpoznávání lidského hlasu, mohou také používat aplikace na diktování např. na diktování SMS zpráv nebo emailu. Princip je také poměrně jednoduchý, kdy uživatel diktuje text a ten se mu zobrazuje na displeji telefonu. Pokud aplikace špatně rozpozná dané slovo, uživatel může slovo sám opravit. Tato funkce opět uživateli pomůže zrychlit operace, které trvají v porovnání s použitím hlasového diktování, neúměrně dlouho.

Hlasové vyhledávání bylo vyzkoušeno s poměrně vysokou úspěšností a může teda být řečeno, že se jedná o plně funkční využití hlasového modulu.

Další oblastí kde v současné době mobilní telefony využívají hlasové technologie je tzv. konverzační mód pro Google Translator. Pod touto funkcí se skrývá skvělý pomocník pro mnoho lidí, kteří potřebují vyslovit slova a věty v cizím jazyce. Funkce pracuje tak, že uživatel spustí aplikaci a vybere si jazyk, ze kterého se bude překládat a do kterého se bude překládat. Poté již může vyslovit slovo či celou větu a aplikace automaticky rozpozná jednotlivá slova ve větě a ty přeloží do jazyka, který si uživatel zvolil. V poslední fázi je celá věta v cizí řeči přečtena danou aplikací včetně přízvuku a

dalších důležitých prvků řeči. Tato funkce se určitě stane neocenitelným pomocníkem lidí v cizí zemi, kdy už nebudou muset nosit sebou velké slovníky a lámat si hlavu nad výslovností, jelikož mobilní telefon vše, včetně vyslovení věty, řekne za ně.

### **6.3 Ovládání GPS navigací v mobilním telefonu pomocí hlasu**

Rozvoj a velký rozmach hlasových technologií v mobilních telefonech dostihl i oblast map v mobilních telefonech, konkrétně těch mobilních telefonů, jež v sobě mají zabudovaný GPS navigátor. Tato funkce je také omezená v tom, že uživatel s telefonem musí být připojen k internetu.

Celá funkce pracuje na principu, že uživatel spustí aplikaci Google Search a slovně zadá požadavek, že chce najít nějakou adresu. Spustí se aplikace Google Voice Navigate, ta adresu najde na internetu a uživatel si může zvolit, jestli chce na danou adresu začít navigovat pomocí GPS navigace. Poté již telefon funguje jako klasická GPS navigace, včetně hlasového navigování. Uživatel si také v aplikaci může předdefinovat určité příkazy na ovládání mapy. Tím je myšleno, např. příkaz se podívat o několik ulic dále, zmenšit si či zvětšit měřítko, posunout mapu doprava či doleva, ale také zapnout jiný podklad. To je další výhodou této funkce, že uživatel nemusí využívat jen mapy, které jsou vytvořeny pro GPS navigaci, ale může si zapnout jakýkoli podklad z Google Maps. Dále si také může pomocí určitého povelu zapnout různé doplňující funkce, např. počasí či semaforey.

Tato funkce nevyužívá mapy, které by byly uloženy přímo v mobilním telefonu, ale stahuje aktuální mapy z Google Maps, což může být pro některé uživatele problém v tom, že nemají na svém mobilním telefonu tak rychlé připojení k internetu, aby se dostatečně rychle načítala data. Pokud uživatel nemá rychlejší připojení k internetu, je tato funkce pro něj nevyužitelná, jelikož mapy by se mu načítaly později, než by bylo zapotřebí a jen by ho to např. při navigování v autě mátl.

Uživatel, který však má na svém mobilním telefonu rychlé připojení k internetu využívající technologie 3G či 4G sítí, dostává do ruky ještě o něco vylepšenou funkci. V aplikaci totiž nemusejí být využívány mapy z mapového serveru Google Maps, ale může uživatel využívat mapy z mapové aplikace Google Earth včetně všech doplňkových funkcí této aplikace. Tzn., že mohou být například zapnuty 3D budovy, takže uživatel na displeji svého mobilního telefonu uvidí nejen cestu, kam má jet, ale i budovy kolem. Dále si také může zvolit, pod jakým úhlem se bude na navigování dívat.

Celkově lze považovat hlasové vyhledávání v mapách, tak jak jej vytvořil Google, za průlomovou technologii v oblasti GPS navigace pomocí mobilního telefonu. Na jednu stranu je to, že aplikace využívá ke své práci mapy, které jsou na internetu, a ty neustále stahuje, výhodou, jelikož uživatel se nemusí starat o zdoluhavé a často i drahé aktualizace map a v každém okamžiku má k dispozici nejaktuálnější informace, jaké může mít. Na druhou stranu to, že aplikace využívá pro stahování, tak pro vyhledávání internet pro některé uživatele bez internetu v mobilním telefonu poměrně dost omezující a musí volit

alternativní řešení. Také pro uživatele, kteří mají internetové připojení v mobilním telefonu s omezeným datovým tarifem je aplikace nepříliš využitelná, jelikož množství dat, které je zapotřebí, aby aplikace fungovala na optimální úrovni, je v některých případech tak velké, že uživateli stačí jeho datový tarif jen na velice krátký čas navigování. Avšak v dnešní době poměrně rapidně roste uživatelů, kteří na svém mobilním telefonu využívají internet a také se zvyšují datové tarify uživatelů, v optimálním případě využívají neomezený datový tarif.

V rámci diplomové práce byla vyzkoušena práce s mobilním telefonem, který umí ovládat GPS navigaci pomocí hlasu. Funkčnost byla velice dobrá ve všech případech, dokonce i v autě, kde jsou daleko více rušivé vlivy. Na druhou stranu bylo z datového tarifu daného telefonu vyčerpáno téměř 80 GB, což je až příliš a je to opravdu velkým handicapem tohoto systému.

## 6.4 Hlasové ovládání GPS navigací

Jak Google Voice Navigate patří mezi poměrně nové aplikace v oblasti navigace pomocí mobilního telefonu, tak klasickou navigaci do auta může uživatel ovládat hlasem již nějakou dobu. Celý systém navigace s hlasovým ovládáním vznikla z potřeby, aby řidiči automobilů se plně soustředili na ovládání vozidla a nemuseli tak ručně nastavovat GPS navigaci [8].

Oproti aplikaci Google Voice Navigate jsou však navigace s hlasovým ovládáním poměrně dost pozadu, jelikož opět není použit tak kvalitní hlasový modul, aby bezchybně rozpoznával řidičovi povely a často tak musí zvolit klasické ruční nastavení GPS navigace. Tyto navigace také mají často problém s rozpoznáváním pokud hraje v automobilu rádio či zvuk motoru je příliš vysoký, což je u starších automobilů hodně časté. V těchto případech opět řidič musí sáhnout k ručnímu nastavení, čímž se snižuje výhoda, že v navigaci existuje hlasové ovládání.

V GPS navigacích pracuje hlasové ovládání na principu, že hlasový modul nejprve řidiči přečte celou nabídku, z čeho může vybírat, např. najít cíl či oblíbené cíle apod., a až poté uživatel může reagovat na, co si chce konkrétně vybrat. Pokud navigaci „skočí“ do řeči, tak se stejně nic nestane a uživatel musí počkat, až navigace dokončí nabídku.

Velmi často však navigace obsahuje v balení speciální tlačítko, které se připevní na volant a pomocí něj se spouští právě hlasové ovládání. Tento užitečný doplněk umožňuje používat hlasové ovládání jen v případech, kdy je opravdu potřeba a nedochází tak ke samospouštění, kdy např. zvukový podnět z radia aktivuje navigaci.

Hlasové ovládání navigace najdeme až ve vyšší cenové skupině GPS navigací a navíc, téměř žádná GPS navigace zatím neobsahuje hlasový modul pro český jazyk, tzn., že pokud máme nějaký název ulice či město, které je na vyslovení a rozpoznání těžké, obsahuje v sobě různé hlasové náročnosti, tak je rozpoznání většinou špatně interpretováno.

## 7 ARCSOUND

Pole geoinformatiky, je co do používání hlasového modulu, ještě stále příliš neorané. Důvodem proč tomu tak je, může být to, že geoinformatika je velice specifický obor a uplatňování hlasového modulu v praxi je pro geoinformatiku stále velmi složité. Je to také náročnost celého odvětví geoinformatiky, kdy i teoretické základy tohoto oboru jsou dosti obtížné a i proto musí být hlasový modul pro využití v geoinformaticce upraven tak, aby těmto základním pravidlům, na kterých geoinformatika stojí, plně vyhovoval a aby byl pro uživatele, kteří se s ním setkají v praxi, opravdovou výhodou a ne spíš přítěží.

Splnění všech těchto podmínek, tak aby byl opravdovým pomocníkem v geoinformaticce, je velice náročné a ne každému se to podaří. Částečně i proto, že se jedná o tak náročný úkol, se jej snažila vyřešit téměř nejznámější firma produkující software pro geoinformatiky, americká společnost ESRI. Výsledkem jejího snažení je nadstavba pro ArcGIS pojmenovaná klasicky pro tuto firmu ArcSound.

### 7.1 Popis aplikace ArcSound

Aplikace ArcSound vznikla v prosinci roku 2006 v dílně známého vývojáře, který vyvíjí nadstavby a extenze pro ArcGIS, Richie Carmicheala. Jedná se o vývojáře, který pracuje přímo pro společnost ESRI a tudíž je ArcSound licencován přímo na společnost ESRI.

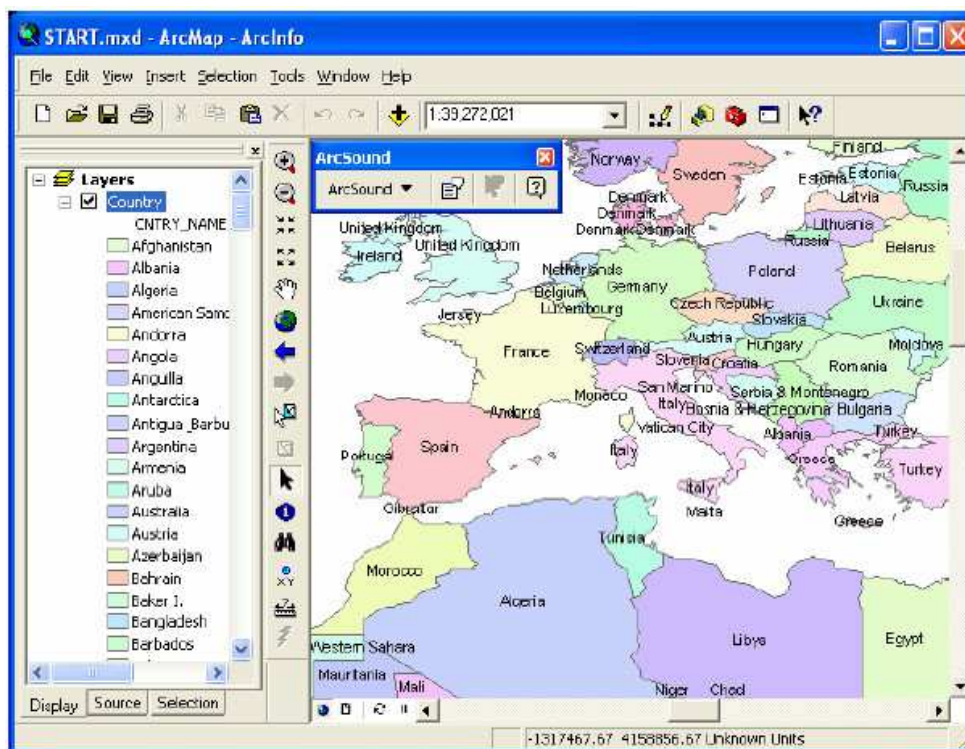
ArcSound je hodně specifický v tom, že se o něm, kromě stránek ESRI, nevede na internetu žádná diskuze. Částečně je to určitě kvůli tomu, že vývoj na této aplikaci byl zřejmě pozastaven, ale lze tak soudit jen podle toho, že poslední verze pochází právě z data, kdy byla aplikace publikována. Další zvláštností aplikace je to, že na internetových stránkách ESRI lze stáhnout pouze demo verzi celé aplikace, ale nikde, ani v manuálu, není vůbec zmíněno, čím je tato stáhnutelná demo verze omezená. Popřípadě není nikde uvedeno ani jak tuto demo verzi aktivovat na plnou verzi aplikace.

Samotná aplikace je velmi limitována v tom, na jakém operačním systému a na jaké verzi ArcGIS pracuje. Funkční verze operačního systému, na kterém je aplikace vyzkoušena, že funguje bez nějakého omezení, je Microsoft Windows XP SP1, Microsoft Windows XP SP2 a Microsoft Windows 2003. Verze ArcGIS, na kterých ArcSound pracuje, je ArcGIS 9.1 nebo ArcGIS 9.2. Tato omezení jsou částečně zapříčiněna tím, že vývoj aplikace dále nepokračuje a samotná aplikace není přizpůsobena novým verzím operačního systému a verzím programu ArcGIS. Dalšími důležitými komponentami, bez kterých ArcSound nefunguje, jsou Microsoft .NET Framework 2.0, ESRI .NET Assemblies a Microsoft DirectX, minimálně ve verzi 9.

Dále je také uváděno, že pro bezproblémový chod aplikace je důležité správně nastavit mikrofon. Tady je aplikace rovněž limitována tím, že v nastavení mikrofonu musí být nastaven Speech Recognizer, který slouží v operačním systému Windows k rozpoznávání hlasu. Bohužel však tento rozpoznávač hlasu lze nastavit pouze v anglické lokalizaci

operačního systému Windows. To znamená, že pro uživatele bez anglické lokalizace Windows nelze aplikace použít. Tady na tomto problému skončila i má osobní snaha vyzkoušet jak aplikaci pracuje. Celá instalace aplikace proběhla úspěšně, avšak po spuštění ArcMap a zapnutí ArcSound naskočila chyba programu, po které se celý program ukončil. Tudíž jsem nemohl osobně vyzkoušet funkčnost celé aplikace.

Aplikace se spouští jako klasický toolbar v ArcMap. Po zapnutí se nám zobrazí na obrazovce displeje toolbar ArcSound, který obsahuje 3 tlačítka. Tlačítka pro zapnutí jedné a druhé funkce a také tlačítko nápovědy.



Obr. 9 Zobrazení ArcSound v ArcMap

## 7.2 Funkce ArcSound

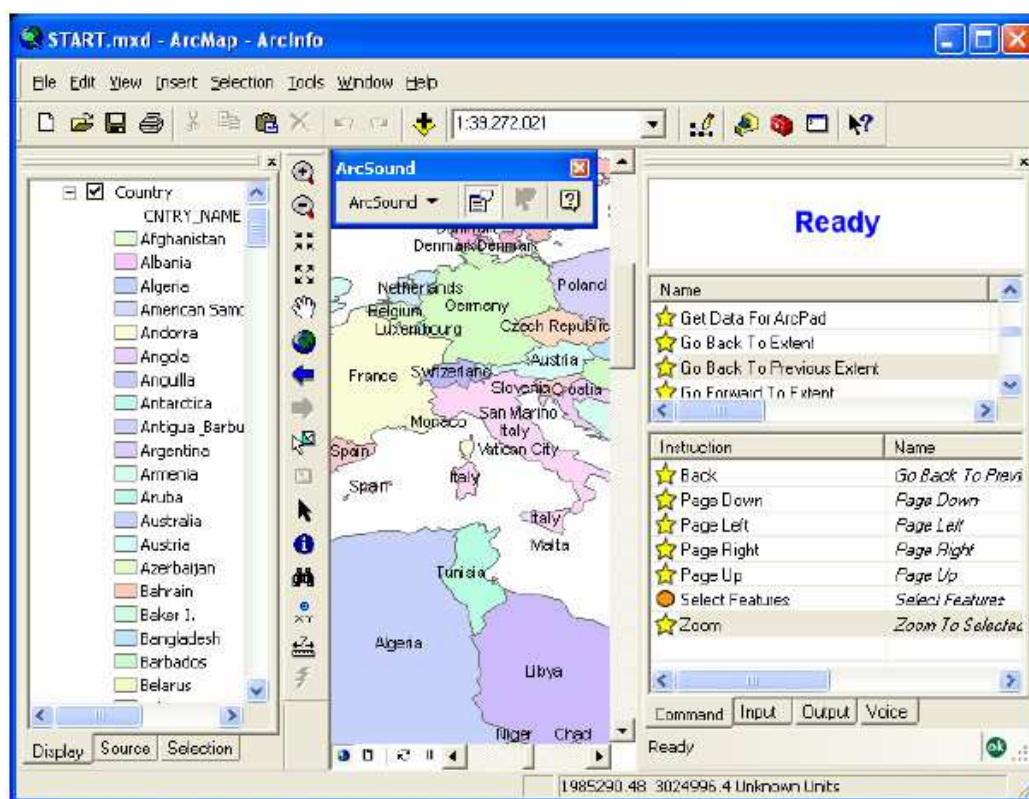
ArcSound má dvě základní funkce. První funkcí je spuštění libovolného příkazu v ArcMap pomocí hlasového povelu a druhou funkcí je přiřazení určitého zvuku k nějakému prvku v mapě.

### 7.2.1 Spouštění příkazu hlasem

Tato funkce se zapíná na toolbaru ArcSound jako první tlačítko zleva pod názvem Show Speech Windows. Po spuštění této funkce se nám okno hlavní mapy rozdělí na 4 podokna. První podokno nám zobrazuje samotnou mapu, kterou můžeme ovládat pomocí hlasových příkazů. Druhé okno nejvíce nahoře, nám zobrazuje, v jakém stavu se aplikace nachází, neboli jestli je připravena na příkazy, jestli něco načítá apod. Třetí podokno nám zobrazuje všechny příkazy z ArcMap, které se v naší verzi nacházejí a ke kterým můžeme přiřadit hlasový příkaz. Příkazy se nám do tohoto okna načítají po spuštění samotné

funkce Show Speech Windows, kdy aplikace najde veškeré příkazy, včetně příkazů od extenzí a nadstaveb, které máme v ArcMap nainstalovány, tudíž můžeme ovládat úplně vše. A poslední okno nám zobrazuje ty příkazy, kterým jsme již přiřadili určitý hlasový povel.

Nový příkaz se tvoří tak, že si jej najdeme v prostřední tabulce a poté jej jen přetáhneme do spodního okna. Poté, co již daný příkaz máme ve spodním okně, tak jej stačí označit a ve vrchním okně kliknout na příkaz READY a namluvit daný hlasový povel. Příkazy lze ve spodním okně různě pojmenovávat tak, jak to vyhovuje danému uživateli. Důležitou věcí při nahrávání jednotlivých povelů je dbát na to, aby jednotlivé příkazy byly od sebe co nejvíce různé, protože pokud jsou si trochu podobné tak se aplikace může jednotlivé příkazy lehce zaměnit, pokud dojde k nějaké změně v podmínkách, ve kterých byly povely namluveny a v jakých podmínkách jsou aplikovány.



Obr. 10 Funkce Show Speech Windows aplikace ArcSound

Určitou nevýhodou této funkce může být to, že příkazy, které si vytvoříme k jedné mapě, jdou použít jen právě k této mapě a ne k jiné. To znamená, že ke každé mapě musíme vytvářet příkazy nové.

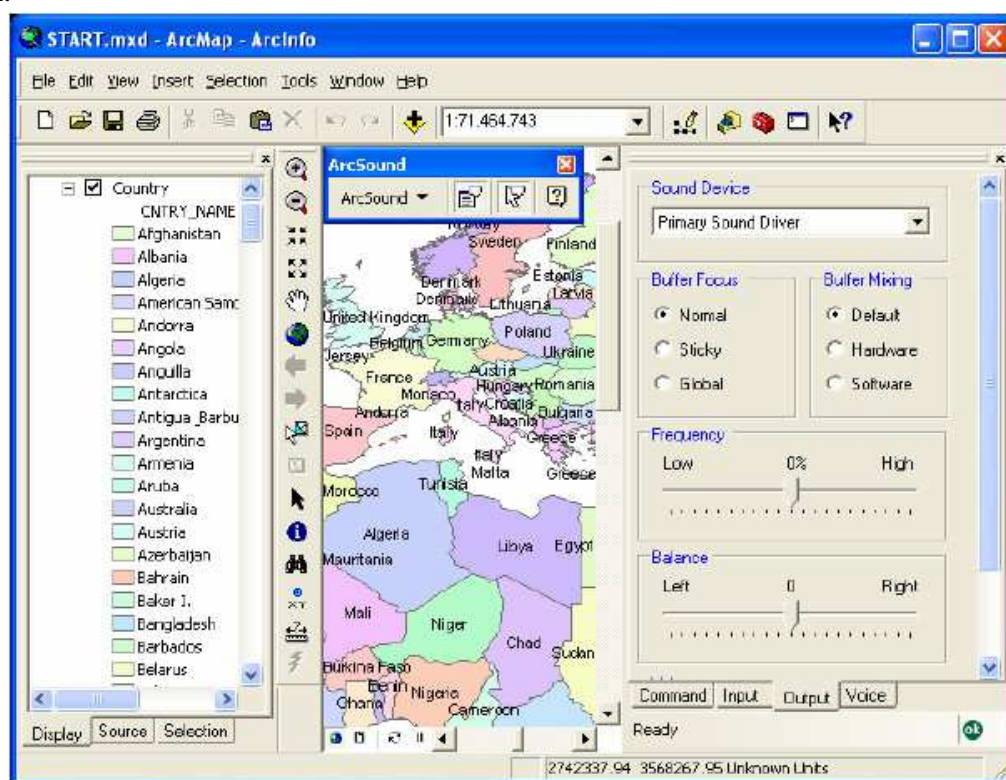
## 7.2.2 Zvuky v mapě

Aplikace ArcSound využívá právě Microsoft DirectX k nahrání a sdílení zvuků u nějakého prvku, který je uložen v geodatabázi. ArcSound ukládá tento prvek se zvukem jako binární objekt, stejně jako ukládá ESRI geometrii v databázové tabulce.



Funkce pracuje na principu, že pokud máme mapovou vrstvu a chceme k jednotlivým prvkům této vrstvy přiřadit nějaký zvuk, tak nejdříve vybereme vlastnosti celé vrstvy a tam zvolíme, o jaký jazyk se bude jednat. To se nastavuje kvůli tomu, že různý jazyk může být vyslovován s různým přízvukem a právě tato vlastnost nám zaručí výběr správného příkazu. Dalším krokem je vybrat v mapě ten konkrétní prvek, ke kterému chceme vybrat daný zvuk. Poté již vybereme na toolbaru ArcSound druhou funkci pojmenovanou Play and Record Sound. Tam si již můžeme vybrat, jestli potřebujeme zvuk nahrát nebo přidat zvukový soubor uložený v počítači (např. písničku). Pokud již máme přiřazené všechny zvuky tam, kde je potřebujeme mít, tak se nám nad každým prvkem v mapě zobrazí přehrávací tlačítko a po kliknutí na něj se nám přehraje zvuk, který jsme k němu přiřadili. Výhodou je určitě to, že u daného zvuku si uživatel může nastavit určité vlastnosti, jako jsou vyrovnávání zvuku nebo frekvence zvuku.

Tato funkce je výhodná v tom, že můžeme přiřadit určitý zvuk k nějakému prvku. Například můžeme přiřadit ke každému státu v mapě jeho státní hymnu, nebo ke státu přiřadit určité povídání o tomto státu. Tímto užitečným prvkem se nám mapa stane zajímavější a více virtuální, jelikož uživatel zapojí při prohlížení této mapy i další smysl, sluch.



Obr. 11 Nastavení zvukových vlastností jednotlivých prvků

### 7.3 Hodnocení ArcSound

Aplikace ArcSound dle mého názoru je velmi užitečný a silný nástroj pro tvorbu virtuálních map, kdy za pomoci zvuků a hudby vytvoříme pro uživatele daleko zajímavější produkt než jen statickou mapu. Dále také nabízí alternativu ke klasickému

ovládání ArcMapu za pomoci hlasu, což je určitě velmi perspektivní směr vývoje tohoto softwaru.

Na druhou stranu je spuštění aplikace daleko složitější než ovládání celé aplikace. Také to, že se již od roku 2006 neobjevila nikde novější verze aplikace, která by v sobě obsahovala jednak jednodušší instalaci, ale také podporu více verzí operačního systému a také novějších verzí samotného ArcGIS.

Sám bohužel nemohu říct, jaká je funkcionality programu, jelikož se mi aplikaci nepodařilo nastavit do takové hloubky, aby správně fungovala, i když jsem aplikaci testoval na různých operačních systémech i verzích ArcGIS. Ale funkce, které aplikace obsahuje, bezesporu přispěje do budoucna k daleko více virtuálnímu ovládání i prohlížení map, a to nejen v ArcGISu.

## 8 MOŽNOSTI UPLATNĚNÍ V GIS

Jak již vyplývá z cílů této diplomové práce, měl by být podán určitý přehled oblastí geoinformatiky, kde by byl hlasový modul využitelný. Budou uvedeny především teoretické oblasti, kde by hlasový modul byl zřejmým přínosem pro uživatele a kde by se stal užitečným nástrojem pro další vývoj. Mnoho oblastí geoinformatiky již využívá nějakým způsobem hlasový modul, ať už je to modul na syntézu či na rozpoznávání hlasu. Ve většině případů se však jedná jen o jakýsi pohled do budoucna, kdy program obsahuje hlasový modul, který má uživateli ukázat, kam by mohl směřovat další vývoj v oblasti jeho využití.

U každé oblasti využití budou také uvedeny technické problémy, které zapříčiňují, že tato oblast stále není realizována, popřípadě je realizována jen zčásti.

### 8.1 Ovládání mapy hlasem

Ovládání mapy hlasem patří mezi základní oblasti geoinformatiky, ve které by byl hlasový modul hodně využívaný a byl by určitě přínosem pro uživatele. Díky hlasovému ovládání by uživatel nemusel často složitě ovládat mapu v jeho počítači, ale pomocí jednoduchých hlasových povelů zvolených podle toho, jak by danému uživateli vyhovovaly, by lehce ovládal veškeré operace spojené s ovládáním mapy. Důvodem, proč stále ještě není hlasový modul ve standardní výbavě programů na prohlížení mapy je zřejmě fakt, že kvalitní hlasový modul, který by opravdu ulehčil práci uživatelům je stále ještě příliš drahý pro výrobce těchto programů a také to, že výrobci si nejsou stále jisti, jestli je vůbec v těchto programech využitelný.

Hlasový modul by však byl neocenitelnou pomůckou pro handicapované lidi, kteří z různých důvodů nemohou používat počítač běžným způsobem a díky němu by se pro ně otevřela další oblast využití map v počítači, ke které stále nemají plný přístup.

Příloha č. 2 obsahuje návrh možných hlasových povelů na ovládání mapy hlasem.

### 8.2 Vyhledávání v mapách

Tato oblast plně navazuje na předchozí využití, ovládání mapy hlasem. Vyhledávání v mapách pomocí hlasu by posunulo právě zmiňované ovládání hlasem na další úroveň vývoje. Vyhledávání pomocí hlasu je totiž velmi náročné na použitý hlasový modul, který musí být opravdu na špičkové úrovni. Vyhledávání je totiž pro hlasový modul zrádné v tom, že uživatel může pokaždé pro vyhledání jedné konkrétní věci použít různě položený dotaz, jednou může být jinak citově zabarvený, jindy zase třeba může být adresa vyslovena v opačném pořadí než ji hlasový modul má naučenou a poté již nedojde k nalezení požadovaného prvku. To znamená, že hlasový modul v sobě musí také obsahovat určitý stupeň umělé inteligence, která pochopí právě tyto drobnosti a vždy vyhledá požadovaný prvek.

Něco na tomto principu využívá právě zmíněný Google pro své vyhledávání v mapách, ale vyhledávání probíhá na internetu, který však umí vyhledávat pomocí různě položených dotazů. Tato oblast je právě díky aplikaci od Googlu na velmi dobré cestě, aby za nedlouho nebyl k vyhledávání používán internet a stejně kvalitní vyhledávání by bylo možné i bez něj.

### **8.3 Interaktivní mapy**

Pod spojením interaktivní mapy si každý uživatel představí něco jiného. V tomto případě je myšleno to, že mapa si bude s uživatelem povídat. Je to myšleno tak, že uživatel mapy vysloví na mapu nějaký dotaz a mapa mu na něj odpoví. Např. uživatel má před sebou politickou mapu a potřebuje zjistit nějaké informace o určitém státu bez toho, aby se musel připojit na internet a tyto informace tam hledat. Místo toho vysloví příkaz: „Najdi základní informace o České republice.“. Mapa buď ze své databáze vybere požadované informace, nebo se sama připojí na internet a informace najde tam. Následně mapa výsledek přečte uživateli. Celý tento systém má mnoho dalších uplatnění, např. při vyhledávání spojů apod.

Úskalí tohoto systému interaktivní mapy tkví především v tom, že musí obsahovat kromě kvalitního hlasového modulu na rozpoznávání hlasu i kvalitní syntetizátor, který bude umět přečíst přirozeně informace, které najde v databázi nebo na internetu. Celý systém by však byl velmi cenným pomocníkem pro všechny, kteří potřebují a používají mapu jako komplexní pomůcku pro získávání informací. Velké využití by našel ve vzdělávání, kde by měly děti takovou virtuální mapu jako komplexní učebnici se všemi potřebnými informacemi v jednom.

### **8.4 Použití hlasového modulu u GPS**

O této oblasti využití již bylo v této diplomové práci psáno. Bude jen zmíněno, že hlasové moduly v GPS navigacích stále ještě nedosahují takové kvality, aby pokaždé poznali vyslovené slovo, takže určitě budou také potřebovat v sobě obsahovat určitý stupeň umělé inteligence, aby rozpoznání bylo vždy stoprocentní. GPS navigace by v sobě měly obsahovat také modul na ovládání mapy hlasem, jelikož by se tím také zjednodušilo ovládání celé navigace, když řidič řídí automobil. Opět musí být zmíněno, že pokud chceme v GPS navigaci kvalitní hlasový modul, musíme si poměrně dost připlatit a dále platí, že společnosti raději používají méně kvalitní a tím pádem levnější hlasový modul, aby tyto GPS navigace byly prodejné.

### **8.5 Využití v DPZ**

Každému, kdo někdy upravoval snímky DPZ, je znám fakt, že nastavení různých parametrů u snímků je důležité, aby vyniklo, to co z těch snímků vyniknout má. Tudíž si

člověk, který tyto parametry nastavuje, musí s tím, aby je správně nastavil opravdu pohlát. A to může být v některých případech náročné, jelikož musí stále zkoušet nové nastavení parametrů, dokud není ideální. A právě v tom případě nastupuje na řadu hlasový modul. Díky hlasovému modul může nastavovat parametry pomocí komfortně pomocí hlasu a nemusí se namáhat s ručním nastavováním. Běžný člověk by tuto pomoc zřejmě nevyužil a neocenil ji, ale člověku, který musí neustále pomocí myši nastavovat různé parametry u snímku, taková pomoc určitě přijde vhod. Opět je důvod proč není dodáván software k úpravě snímků DPZ zřejmý- velká cena hlasového modulu a tím zvýšení ceny softwaru.

## **8.6 Využití ve 3D**

Tato oblast využití je shodná s první oblastí využití, ovládání mapy hlasem, s tím rozdílem, že přenáší ovládání do 3D prostoru v počítači. Hlasový modul by byl využitelný při ovládání různých průletů nad 3D povrchem jak jej známe např. z ArcScen, popř. Google Earth. Ovládání těchto průletů je často dost náročné, je dost náročné, jelikož průlet je velice citlivý na jakékoli větší či menší zmáčknutí klávesnice. Díky hlasovému modulu by se tento problém částečně odstranil, kdy by si uživatel sám vytvořil příkazy na ovládání průletu nad 3D a mohl by si i zvolit různou hodnotu citlivosti daného kroku. Hlasový modul by v tomto případě přispěl k lepší ovladatelnosti průletů nad 3D povrchem a uživatel by se tak nemusel zabývat tlačítky na klávesnici, ale plně se věnovat průletu na 3D povrchem a mít pocit, že opravdu letí nad daným povrchem, než aby se stále soustředil na správné klávesy pro ovládání.

## **8.7 Hlasové ovládání totální stanice**

Oblast geodézie je určitě další oblastí, kde by hlasový modul našel uplatnění. Při práci v terénu s totální stanicí musí člověk, který tuto stanicí ovládá neustále nastavovat různými souřadnicemi a parametry, aby totální stanice pracovala správně. Toto nastavování je vždy velmi zdlouhavé, a když to člověk dělá během dne několikrát je to poměrně dost jednotvárné a spíše zatěžující. Pomocí hlasovému modulu by celé ovládání totální stanice fungovalo pomocí hlasových příkazů a bylo tak velmi cenným pomocníkem v terénu. Další využití pro ovládání totální stanice je u stanic, které se ovládají na dálku, tzn., že ji ovládá člověk, který chodí s vyměřovací tyčí. Pomocí hlasových povelů by se práce s totální stanicí opět zjednodušila. Jelikož požadavky na přesnost totální stanice jsou poměrně velké, tudíž i hlasový modul, který by ovládal totální stanicí by musel být velmi kvalitní což by cenu, již tak dost drahých totálních stanic, ještě zvedlo.

## 8.8 Informační tabule

Informační tabule ve městech jsou pro cizí návštěvníky města velmi cenným zdrojem informací, často také jediným zdrojem informací o památkách a jiných zajímavých místech. V dnešní době jsou tyto informační tabule většinou dotykové. To s sebou nese mnoho potíží, například vyhledávání, kdy dotyková obrazovka nemusí správně fungovat a uživatel tak jednoduchý text píše delší čas, než je k tomu zapotřebí. Pokud by do těchto informačních tabulí byl zabudován hlasový modul, celá práce s tabulí se zjednodušila.

Informační tabule se samozřejmě většinou nachází na hlučných místech, tudíž celý proces rozpoznávání hlasových povelů by byl značně ztížen, ale tento problém by byl vyřešen tak, že by se neustále prováděla kalibrace na hluk okolního prostředí a tato hodnota by byla odečítána od přijímaného hlasového příkazu. Celý systém této informační tabule s hlasovým modulem by určitě přinesl jejím uživatelům, což jsou většinou turisté, značné zjednodušení.

## 8.9 Ovládání mobilního GISu hlasem

Mobilní GIS patří neodmyslitelně k efektivnímu sběru dat v terénu. Jeho ovládání je někdy však pro jeho uživatele příliš složité, jelikož uživatelé potřebují v terénu především jednoduchost, aby dokázali pracovat co možná nejefektivněji. Použitím hlasového modulu právě pro ovládání v terénu by mělo pro uživatele mnoho kladů. Zejména by to byla nenáročnost, jelikož by vše ovládal pomocí jednoduchých hlasových povelů. Např. zápis nového bodu by probíhal tak, že by uživatel vyslovil příkaz pro uložení nového bodu, ten by se uložil a pomocí hlasu by uživatel pojmenoval daný bod, což je daleko efektivnější než klasický postup. Samozřejmě, že mobilní GIS je používaný většinou ve venkovním prostředí, takže rušivá složka, která by mohla ovlivňovat přijímaný hlasový povel, je zde také zastoupena. Opět jako u předchozího případu, by to bylo řešeno pomocí kalibrace celého mobilního GISu na okolní prostředí.

## 8.10 Hlasové ovládání animací

Velice oblíbené a populární je dnes vytvářet animace z map, např. jak se dané území vyvíjelo v čase, jak se měnila určitá událost v čase apod. Uživatel, pokud sleduje určitou animaci, tak se chce soustředit pouze a jen na animaci. Pokud je animace příliš dlouhá, tak ovládání pomocí klávesnice či myši je pro uživatele značně rozptylující, jelikož se musí soustředit jak na animaci, tak na ovládání animace. Hlasové ovládání takových animací by uživateli značně zjednodušilo ovládání a celkově by tak přispělo k tomu, že uživatel si danou animaci lépe zapamatuje.

## **8.11 Hlasem plně ovládaný mapový program**

Vize programu na tvorbu map, který bude plně ovládaný pomocí lidského hlasu, není jen hudbou daleké budoucnosti. Aplikace ArcSound umí přiřadit hlasový povel ke každému příkazu a plně tak ovládat celý ArcGIS. Jenže tento program je jen nadstavbou jiného programu. Správně by program měl v sobě obsahovat hlasový modul, který veškeré ovládání bude řídit.

## **8.12 Virtuální mapy**

Pojem virtuální mapa značí, že v dané mapě budou nějakým způsobem zastoupeny virtuální prvky, jako jsou například zvuky. Zvuky v mapě mohou přispět k tomu, že uživatel si mapu lépe zapamatuje a bude pro něj mnohem zajímavější. Zvuky mohou být v mapě zastoupeny v několika podobách. První je, že k určitým prvkům v mapě jsou přiřazeny určité zvuky, například ke státům jejich hymna. Další podobou může být namluvený určitý text. Dále také může být použita určitá hudba pro navození atmosféry při prohlížení mapy. Všechny tyto podoby zvuků v mapě musí umět program na tvorbu map umět do mapy přidat, tak aby byly plně použitelné. Tzn., že zde musí být zastoupen určitý modul, který zajistí, aby vše správně fungovalo. Většinu těchto požadavků na existenci zvuků v mapě umí již zmíněný ArcSound, jenže vývoj tohoto programu byl již zastaven, tak by bylo zapotřebí vyvinout ještě další programové prostředky, které toto budou umět.

## 9 NÁVRH A REALIZACE APLIKACE NA OVLÁDÁNÍ MAPY HLASEM

Jako jedním z dílčích cílů této diplomové práce bylo vytvořit aplikaci, která by ukázala, že opravdu lze ovládat mapu hlasem. Vytvoření celého programu, který by uměl ovládat mapu hlasem je však nesmírně náročný úkol i pro zkušené programátory, jelikož se musí prokousat základními fyzikálními zákony o šíření a rozpoznávání zvukových signálů a všechny tyto poznatky poté aplikovat do programového kódu.

Z toho důvodu byl také zvolen způsob, který není programátorsky tak obtížný, ale ukáže vše, co by taková aplikace na ovládání mapy hlasem měla umět.

Vytvořená aplikace je složena ze dvou programů, JetVoice a ArcExplorer. Program JetVoice v tomto případě slouží jako prvek, který umí ovládat počítač pomocí hlasových příkazů. ArcExplorer zde funguje jako tzv. prohlížeč již hotových map.

### 9.1 Použití JetVoice

Program JetVoice byl vybrán pro tuto aplikaci z toho důvodu, že je velice jednoduchý na ovládání a při zkoušení různých programů, které umí ovládat počítač pomocí hlasových příkazů, vyšel ve funkčnosti a ve spolehlivosti jako nejlepší freeware program.

Jak již bylo napsáno výše, JetVoice umí ovládat počítač pomocí hlasových příkazů a má v sobě zabudované různé možnosti akcí, které může v počítači dělat. Tyto jednoduché akce se poté mohou také různě kombinovat, takže může vzniknout i příkaz, který je poměrně dost složitý.

Program se v počítači nemusí instalovat, takže spouštění programu probíhá tak, že se spustí pomocí ikony *JetVoice.exe* ve složce, ve které ji máme uloženou. Po spuštění programu musí uživatel zapnout tzv. *Příjem dat*, který slouží ke kalibraci mikrofону na okolní prostředí. Předem nachystané hlasové povely jsou automaticky do programu načteny.

Mnou nachystaný soubor hlasových příkazů obsahuje 20 základních hlasových povelů pro manipulaci s mapou. Těchto 20 povelů ukáže to, že lze ovládat mapu hlasem.

Nejdříve musí uživatel vyslovit příkaz pro spuštění předem nachystané mapy (jednotlivé konkrétní příkazy budou rozvedeny v kapitole 10). Mapa se automaticky spustí v programu ArcExplorer a uživatel může dalšími hlasovými příkazy ovládat mapu a zjišťovat z ní informace, které potřebuje.

### 9.2 Použití ArcExplorer

ArcExplorer je program od společnosti ESRI, který slouží pro prohlížení geografických dat. Kromě základních operací na ovládání mapy obsahuje také nástroje na

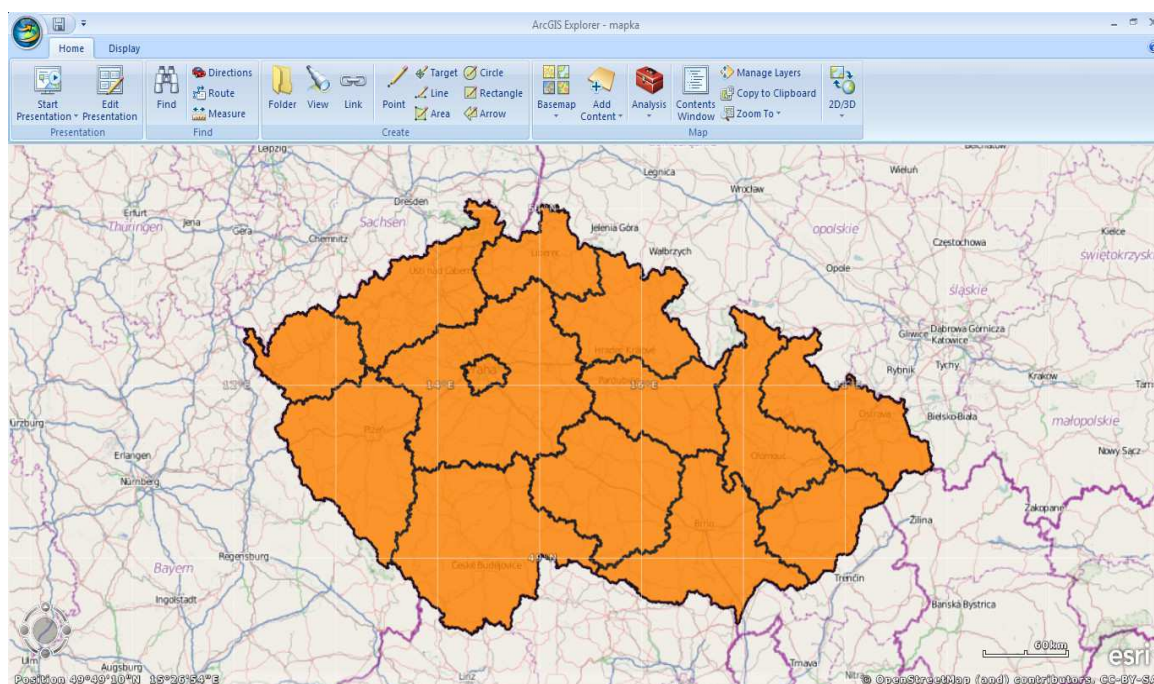


zjišťování informací z mapy. Program umí zobrazit různé rastrové formáty (TIFF, GIF,...), ale také vrstvy SHAPEFILE, ve kterých jsou obsažena i potřebná atributová data. Všechny tyto funkce, které v sobě tento program má, plně poslouží funkci, kterou pro naši aplikaci potřebujeme. Navíc program je plně zdarma, takže nemusíme mít starosti s libovolným použitím.

Jak již bylo napsáno výše, ArcExplorer slouží v našem ukázkovém příkladu jako prohlížeč již hotových map. Tzn., že máme dopředu připravené mapy, které se po vyslovení příslušného příkazu otevřou automaticky v ArcExplorer. Program však musí být nainstalován do počítače, kde chceme aplikaci vyzkoušet. Instalační soubor je na přiloženém DVD.

Připravené mapy jsou ve formátu NMF, což je nativní formát pro použití právě v ArcExplorer. Tento formát patří mezi XML formáty, který přečte každá XML čtečka.

Příkazy, které jsou ovládány pomocí hlasových povelů, vychází ze základního ovládání programu ArcExplorer. Mezi tyto příkazy patří ovládání pomocí směrových šipek, ovládání měřítka zobrazované mapy pomocí kolečka myši a také základní příkazy na ovládání mapových oken v programu.



Obr.12 Ukázka programu ArcExplorer s vytvořenou mapou

## 10 DEFINOVÁNÍ HLASOVÝCH POVELŮ

Definování správných hlasových povelů je velmi důležité pro správnou funkčnost celého systému na ovládání mapy pomocí hlasu. Při definování celého souboru hlasových povelů pro ovládání určité mapy je zapotřebí dbát především na to, aby uživatel neměl žádný problém s výslovností daných povelů a také, aby pro něj byly povely snadno zapamatovatelné.

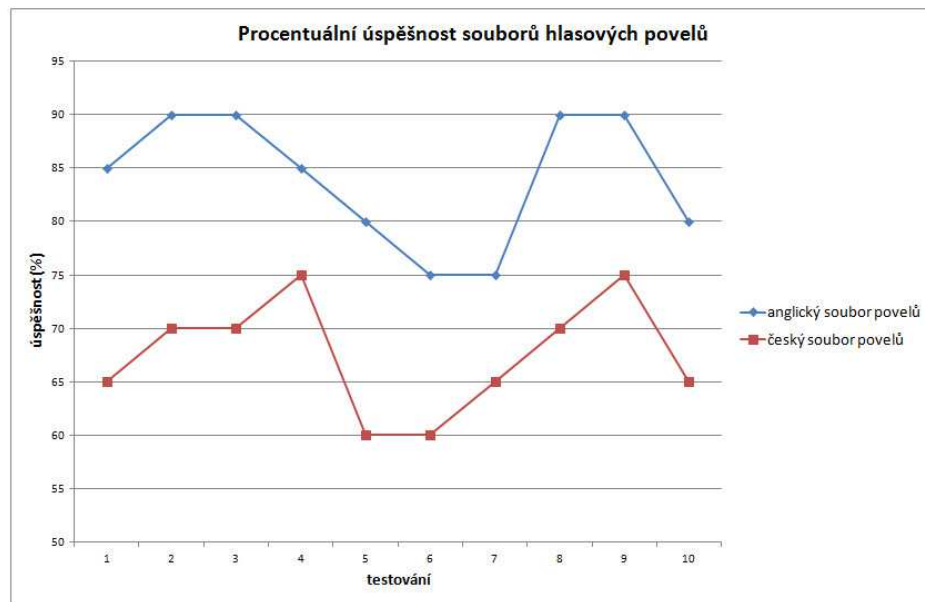
### 10.1 Snadná výslovnost

Podmínka snadné výslovnosti hlasových příkazů je splněna tehdy, když jsou povely vybírány tak, že povely nebudou v sobě obsahovat těžko vyslovitelné hlásky, jako jsou Ř, Ž a jiné. Je to důležité proto, že ne každý vyslovuje tyto hlásky stejně, a poměrně dost lidí má problém s výslovností těchto hlásek. Proto je důležité také dopředu zvážit zda-li je pro hlasové příkazy vhodné volit češtinu a ne třeba zvolit jiný jazyk, např. angličtinu, která v sobě obsahuje daleko méně těchto problémových hlásek. Použitím angličtiny je však také z určitého důvodu riskantní, a tím je to, že ne každý uživatel umí vyslovit anglická slova se správnou výslovností.

Při výběru jazyka, který bude použit u hlasových povelů ve vytvořené aplikaci, byly vyzkoušeny dva různé soubory hlasových povelů. První byl tvořen slovy pouze v češtině a druhý byl naopak tvořen slovy pouze v angličtině. Každý soubor obsahoval 20 hlasových povelů, kdy ke každému byl přiřazen příkaz na provedení určité akce v mapě.

Celkově bylo při zkoušení provedeno 10 hlasových zkoušek u obou souborů. Postupně bylo zaznamenáváno, jestli hlasový povel provedl opravdu to, co měl provést, a na konci každého testování byla vypočtena procentuální úspěšnost u daného souboru. V následujícím grafu je znázorněna právě úspěšnost obou souborů v jednotlivých testováních.

Graf 1 Procentuální úspěšnost souborů hlasových povelů



Z grafu můžeme lehce usoudit, že angličtina byla v tomto případě o dost úspěšnější než čeština. Průměrná úspěšnost angličtiny ve všech 10 testováních byla 84%, kdežto úspěšnost češtiny byla pouze 67,5%. Z toho můžeme usoudit, že angličtina je obecně vhodnější pro použití jako jazyka, pro hlasové povel na ovládnání mapy hlasem. Nemůžu přesně říct, co přesně ovlivňuje to, že daný povel není správně rozpoznán. Ve všech testování byly naprosto stejné okolní podmínky i povel vyřekoval stejný mluvčí.

Při testování jsem si také povšiml důležitého poznatku, a to toho, že úspěšnost byla větší u povelů, které jsou kratší, jelikož uživatel snadněji udělá chybu v povelu, který je delší než u toho, který je kratší.

## 10.2 Snadné zapamatování

Pod touto podmínkou je myšleno to, aby si uživatel pod daným povel představitel, co povel bude vykonávat. Tzn., aby povel nějakým způsobem byl spjatý s tím, co bude vykonávat. Pokud uživatel uvidí tady tuhle asociaci mezi vysloveným povel a tím co povel vykonává, lépe si jej zapamatuje pro další použití. Celkově se tedy uživateli zlepšila práce s celým souborem hlasových povelů.

Ve vytvořené aplikaci v rámci diplomové práce je tato podmínka splněna jen částečně. Je to z toho důvodu, aby bylo ukázáno, že opravdu lze ovládat počítač hlasem. Tzn., že u některých příkazů, je zvolen takový hlasový povel, který s příkazem nemá téměř nic společného. Je to z toho důvodu, že ty povel, které byli těžké na vyslovení a tudíž nesplňovali první podmínku snadné výslovnosti, jsou zvoleny takové povel, které jsou na vyslovení naopak jednoduché.

### 10.3 Zvolené povely ve vytvořené aplikaci

Jak již bylo napsáno výše, ve vytvořené aplikaci na ovládání mapy hlasem byla jako jazyk pro povely zvolena angličtina a také určité povely neodpovídají tomu, co přesně představují, kvůli lepší názornosti celého příkladu. Celkově soubor hlasových povelů obsahuje 20 povelů, mezi kterými jsou povely na otevírání předem připravených map, povely na ovládání mapy a povely na získávání informací z mapy. Většina povelů v sobě obsahuje právě jeden nebo více příkazů, které jsou již v programu JetVoice nachystány.

#### **Zvolené povely a vysvětlení jejich funkce:**

##### **Povel *MAPA***

Povel *MAPA* otevře předem připravenou mapu, která zobrazuje na podkladě obecně geografické mapě (z OpenStreet Map) vrstvu krajů České republiky. Povel využívá akce otevření souboru v asociativním programu, tzn., že soubor je otevřen v programu, pro který je soubor určen, v tomto případě ArcExplorer.

##### **Povel *OLOMOUC***

Povel *OLOMOUC* otevře předem připravenou mapu, která zobrazuje obecně geografickou mapu se zaměřením na město Olomouc. Tato mapa je volně k použití z OpenStreet Map. Stejně jako příkaz výše využívá akce otevření souboru v asociativním programu.

##### **Povel *RIGHT***

Povel *RIGHT* posune výřez mapy doprava. Povel využívá akci kliknutí na šipku doprava, která v programu ArcExplorer znamená právě posunutí doprava.

##### **Povel *LEFT***

Povel *LEFT* posune výřez mapy doleva. Povel využívá akci kliknutí na šipku doleva, která v programu ArcExplorer znamená právě posunutí doleva.

##### **Povel *UP***

Povel *UP* posune výřez mapy nahoru. Povel využívá akci kliknutí na šipku nahoru, která v programu ArcExplorer znamená právě posunutí nahoru.

##### **Povel *DOWN***

Povel *DOWN* posune výřez mapy dolů. Povel využívá akci kliknutí na šipku dolů, která v programu ArcExplorer znamená právě posunutí dolů.

### **Povel *IN***

Povel IN zvětší měřítko výřezu mapy. Povel využívá akci posunutí kolečka myši nahoru, kdy tento příkaz v programu ArcExplorer vyvolá zvětšení měřítka mapy.

### **Povel *OUT***

Povel OUT zmenší měřítko výřezu mapy. Povel využívá akci posunutí kolečka myši dolů, kdy tento příkaz v programu ArcExplorer vyvolá zmenšení měřítka mapy.

### **Povel *WINDOW***

Povel WINDOW zavře aktuální otevřené okno. Používá se v případě, že v programu ArcExplorer je otevřeno další okno, např. okno s informacemi o daném prvku. Povel využívá akci zavření aktuálního okna.

### **Povel *CLOSE***

Povel CLOSE zavře program ArcExplorer včetně otevřené mapy. Povel využívá akce zavření hlavního okna.

### **Povel *INFO***

Povel INFO zobrazí informace z atributové tabulky, pokud daná vrstva tabulku obsahuje. Povel využívá akce kliknutí levého tlačítka myši, která v programu ArcExplorer vyvolá otevření okna s danými informacemi.

### **Povel *CENTRO***

Povel CENTRO vycentruje kurzor myši přesně do středu obrazovky, aby manipulace a ovládání ArcExplorer fungovala jak má. Povel využívá akce posunu kurzoru myši na střed obrazovky.

### **Povel *APPLE***

Povel APPLE je podobný jak příkaz RIGHT s tím rozdílem, že posune výřez 3x doprava. Povel slouží pro rychlejší posunování v mapě, kdežto povel RIGHT slouží spíše pro jemné posunování. Povel využívá akci šipky doprava, která se opakuje 3krát. Povel APPLE neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *SUN***

Povel SUN je podobný jak příkaz LEFT s tím rozdílem, že posune výřez 3x doleva. Povel slouží pro rychlejší posunování v mapě, kdežto povel LEFT slouží spíše pro jemné

posunování. Povel využívá akci šipky doleva, která se opakuje 3krát. Povel SUN neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *FISH***

Povel FISH je podobný jak příkaz UP s tím rozdílem, že posune výřez 3x nahoru. Povel slouží pro rychlejší posunování v mapě, kdežto povel UP slouží spíše pro jemné posunování. Povel využívá akci šipky nahoru, která se opakuje 3krát. Povel FISH neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *MOK***

Povel MOK je podobný jak příkaz DOWN s tím rozdílem, že posune výřez 3x dolů. Povel slouží pro rychlejší posunování v mapě, kdežto povel DOWN slouží spíše pro jemné posunování. Povel využívá akci šipky doprava, která se opakuje 3krát. Povel MOK neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *PASTA***

Povel PASTA posune výřez mapy vlevo nahoru. Povel slouží pro ještě lepší pohyb v mapě. Povel využívá akci šipky doleva a nahoru. Povel PASTA neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *BOOK***

Povel BOOK posune výřez mapy vlevo dolů. Povel slouží pro ještě lepší pohyb v mapě. Povel využívá akci šipky doleva a dolů. Povel BOOK neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *CAR***

Povel CAR posune výřez mapy vpravo nahoru. Povel slouží pro ještě lepší pohyb v mapě. Povel využívá akci šipky doprava a nahoru. Povel CAR neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

### **Povel *PEN***

Povel *PEN* posune výřez mapy vpravo dolů. Povel slouží pro ještě lepší pohyb v mapě. Povel využívá akci šipky doprava a dolů. Povel *PEN* neodpovídá přesně tomu, co vyjadřuje, ale je to z toho důvodu, aby příkazy byly snadno rozeznatelné pro program.

## **10.4 Další povely**

V předcházejícím textu byly uvedeny hlasové povely pro základní ovládání mapy. Jednalo se o povely, díky nimž můžeme mapu jakkoli ovládat. Avšak tyto povely nejsou jediné, které by přišly v úvahu při vytvoření plně sofistikované, hlasem ovládané mapy.

Další oblastí povelů, které by bylo možno použít je vyhledávání v mapě. Díky možnosti vyhledávat pomocí hlasu by naše práce s mapou byla daleko efektivnější. K funkčnímu modulu na vyhledávání, by však bylo zapotřebí použít mnohem lepší rozpoznávač řeči, který bude mít v sobě i určité prvky umělé inteligence, jelikož vyhledávání nemusí vždy probíhat pomocí přesně definovaného výrazu, ale může být definován plynulou řečí a program si z toho musí vzít právě to podstatné pro správné vyhledávání.

V určitých případech uživatel potřebuje ovládat mapu, která je nějakým způsobem specifická. Jedná se např. o katastrální mapy, snímky DPZ a jiné druhy map. Tyto specifické mapy vyžadují speciální ovládání a k němu také speciální hlasové povely a příkaz, které umějí ovládat tyto mapy. Toto by bylo lehce vyřešitelné tak, že by uživatel mohl v programu definovat své vlastní povely a k nim si přiřadit ty příkazy, které on potřebuje ke správnému ovládání mapy. Takto by měl pracovat každý program. Měl by mít pro uživatele možnost definovat své vlastní příkazy, bez ohledu na prostředí ve kterém program pracuje.

## 11 VÝSLEDKY

V první části diplomové práce, která byla pojata spíše teoreticky, byla nejprve představena historie obecně ovládání počítače hlasem a také byly vymezeny oblasti, kde v současné době je hlasový modul plnohodnotně využíván. Také bylo stručně vysvětleno, jak lze počítač ovládat hlasem. Byly představeny základní metody a postupy, na kterých jsou hlasové moduly postaveny. Tato část byla napsána velmi stručně, tak aby bylo co možná nejučelněji popsáno, jak celá problematika ovládání počítače hlasem funguje.

V další části byly vybrány programy, které již v sobě obsahují modul na ovládání počítače hlasem. Každý z vybraných programů byl nainstalován, pokud tedy nebyl placený a pokud se instalace zdařila tak byl program vyzkoušen na celkovou funkčnost a celkově porovnáván se všemi vybranými programy. Z daného porovnávání vznikla srovnávací tabulka, která je uvedena jako příloha diplomové práce. Závěr testování jednotlivých programů je ten, že programy, které jsou zdarma, ne vždy pracují přesně, tak jak mají a hlasový modul v těchto programech není na takové úrovni, která zaručuje bezproblémovou práci s programem, především v oblasti přesného rozpoznávání hlasových příkazů. Naopak programy, které jsou placené, a které mají již za sebou nějaký vývoj, tak u těchto programů je hlasový modul již na takové úrovni, že úspěšnost rozpoznávání hlasových příkazů se velmi blíží k 100%.

Byl testován také program ArcSound, který je přímo vytvořen na ovládání mapy hlasem. Bohužel díky nefunkčnosti celého programu se testování nezdařilo a závěr tedy musel být vytvořen na základě videí a diskuzí o programu. Program a celá jeho koncepce je velice dobře zpracovaná, jenže zřejmě zastavený vývoj na program zapříčinil to, že program již není funkční. Program má velkou perspektivu na úspěch, a pokud by se podařilo obnovit práce na vývoji jistě by šlo o velký průlom v oblasti využití hlasového modulu v ovládání GIS softwaru.

Díky tomu, že oblast geoinformatiky je ještě ve velké míře nepokryta používáním hlasového modulu pro různé účely, tak mohl být představen velký počet aplikačních oblastí geoinformatiky, kde by právě hlasový modul byl přínosem. Geoinformatika jako jeden z velmi dynamicky se rozvíjejících vědních oborů skýtá velké možnosti pro využití hlasového modulu. Byly představeny téměř všechny oblasti geoinformatiky a ukázány aplikační možnosti právě na využití hlasového modulu v praxi.

V aplikační části diplomové práce byla udělána analýza vhodnosti hlasových povelů pro ovládání mapy hlasem. Z této analýzy vyplynulo, že lepší funkčnost vykazuje spíše anglický jazyk než náš rodný český jazyk. Výsledky této analýzy vyplývají z testování anglické a českého souboru hlasových povelů a procentuální úspěšnosti celého testování. Dále bylo také zjištěno, že hlasové povely musejí být vybírány se zřetelem na snadné vyslovování a také snadné zapamatování. Tzn., že povel v sobě nesmí obsahovat různé jazykové zvláštnosti (v češtině např. Ř apod.) a také, že povel by měl odpovídat tomu, co představuje v provedení v počítači. Byl vypracován základní seznam hlasových povelů



pro základní ovládání mapy, který sice neodpovídá výše uvedeným konvencím pro hlasové povely, ale splňuje požadavek na ukázkou ovládání mapy hlasem.

V poslední části byl vytvořen již zmíněný program na ovládání mapy hlasem. Tento program je složen ze dvou dílčích programů, ArcExplorer a JetVoice. JetVoice je program na ovládání počítače pomocí hlasových příkazů a ArcExplorer je prohlížeč pro mapy vytvořené v ArcGIS. Spojením těchto programů vznikl právě požadovaný program na ovládání hlasem, který v sobě obsahuje předdefinované příkazy pro manipulaci s mapou. Program je nastaven tak, aby nebyl závislý na mluvčím, tzn., že kdokoliv vysloví daný hlasový povel, příkaz, který je k tomuto povelu přiřazen se automaticky provede.

## 12 DISKUZE

Tato diplomová práce se skládá z několika dílčích cílů, které byly představeny na začátku práce. Jednotlivé cíle byly postupně splněny, avšak s různou úspěšností. Ne všechny cíle se podařilo splnit úplně, největším problémem se stala především nefunkčnost určitých programů, které měly být v práci využívány.

Jako nejvíce časově náročné se ukázalo studium potřebné literatury k pochopení problematiky ovládání počítače hlasem. Literatura je totiž psaná velmi složitě, kdy bylo nutné se prokousat fyzikální podstatou celé problematiky, ale i nastudovat literaturu k pochopení lingvistických zákonitostí zpracování lidského hlasu počítačem. Velkým problémem se stala dostupnost a aktuálnost literatury oborů, které s tématem diplomové práce souvisejí. Z toho důvodu byly do literatury zahrnuty i populárně-naučné články na internetu k dané tématice. Tyto články velkou měrou přispěly k aktuálnímu pohledu na oblast ovládání počítače hlasem, díky tomu, že byly oproti tištěné literatuře zaměřené na aktuální vývoj v této oblasti.

Původního dílčího cíle vyzkoušet funkcionalitu co nejvíce programů, které umějí ovládat počítač pomocí hlasu, se podařilo jen částečně. Bylo to zapříčiněné především tím, že některé programy již nelze na současných verzích operačních systémů provozovat, dále také faktem, že některé programy již mají ukončený vývoj a také nejdou nainstalovat. Dalším velkým problémem bylo, že ty programy, které jsou již na lepší úrovni, jsou většinou placené, a tudíž nemohly být vyzkoušeny a hodnoceny těchto programů bylo prováděno díky instruktážním videím na internetu a díky diskuzím k jednotlivým programům od uživatelů. Také nikde neexistuje žádný ucelený seznam programů, které využívají hlasový modul. Takže výběr programů byl prováděn na základě zmínek v různých článcích a odborné literatuře a je teoreticky možné, že nějaký existující program nebyl do hodnocení vybrán.

Cíl vytvoření aplikace, která bude umět ovládat mapu hlasem, se podařilo splnit také jen částečně. Hned na začátku práce bylo po prostudování jisté, že vytvoření samotného programu na ovládání mapy hlasem napsáním programového kódu, by bylo téma spíše pro zkušeného informatika. Obtížnost vytvoření celé této aplikace vedlo k závěru, že aplikace bude muset být založena na jiném způsobu vypracování. Byl zvolen způsob propojení již existujícího programu na ovládání počítače hlasem s prohlížečem již vytvořených map. Byl vybrán program, který je zdarma a který má oproti jiným programům vyšší funkčnost rozeznávání hlasových povelů. I vzniklá aplikace však má určité chyby. Jedná se především o občasně vypadávání funkcionality programu na ovládání počítače hlasem, kdy u tohoto problému není zjištěná povaha, proč k tomu dochází. Program sice byl zkoušen pro několik lidí a fungoval bez větších obtíží, ale i přesto občas některý hlasový povel není programem správně identifikován. Obecně však vytvořená aplikace splňuje požadavky na ukázkou ovládání mapy hlasem.

Dílčí cíl pro vytvořenou aplikaci ve smyslu vyhledávání v mapě pomocí hlasu se nepodařilo splnit vůbec. Důvodem je přílišná složitost této funkce. Funkce vyhledávání by mohla být řešena dvěma způsoby. První způsob je vytvoření hlasových povelů pro

všechny prvky v mapě, což je však díky obrovskému počtu prvků nereálné. Druhým způsobem by byla téměř umělá inteligence, která bude rozumět a chápat to, co uživatel vyslovuje a na základě toho vyhledávat v mapě, avšak tento způsob vyžaduje využívat daleko lepší hlasový modul včetně umělé inteligence.

## 13 ZÁVĚR

Cílem magisterské práce bylo popsat a navrhnout možnosti implementace hlasového ovládání digitálních map. Zaprvé popsat teoretické možnosti geoinformatiky, kde by mohl být hlasový modul prospěšný a poté aplikovat tyto poznatky na tvorbu vlastní aplikace na ovládání mapy hlasem a také vytvoření vhodných povelů pro práci s mapou.

První část diplomové práce se tedy zaměřila na teoretickou část, kde byly popsány možné oblasti použití hlasového modulu. Oblast geoinformatiky je stále ve využívání hlasových technologií pouze na začátku, a tudíž hlasový modul pro ulehčení práce ještě stále není implementován téměř nikde. Byly také otestovány již existující programy, které jsou zaměřeny na ovládání počítače hlasem. Ne vždy bylo možné z různých důvodů program spustit a otestovat jej. Také možnosti jednotlivých programů jsou značně rozdílné, kdy na jedné straně programy již umí rozpoznávat řeč téměř bezchybně a blíží se k vytvoření plně funkčního programu. Na druhé straně jsou programy, které se jen tváří, že umí ovládat počítač hlasem a jejich funkcionalita je na hodně mizivé úrovni.

Teoretické poznatky byly aplikované v druhé části diplomové práci, kdy byla vytvořena ukázka aplikace na ovládání mapy hlasem. Aplikace je složena ze dvou programů, ArcExplorer a JetVoice. ArcExplorer jako prohlížeč vytvořených map a JetVoice jako program, který umí na základě vlastních hlasových povelů ovládat mapu hlasem. Aplikace obsahuje 20 předvolených hlasových povelů pro základní práci s mapou. Hlasové povely byly vybrány na základě analýzy a porovnání vhodnosti českého jazyka a anglické jazyka, kdy anglický jazyk byl vybrán jako lepší varianta. Hlasové povely byly vybírány se zřetelem na snadné zapamatování a snadnou výslovnost, jakožto dvou základních pravidel pro tvorbu vhodných hlasových povelů.

## POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

1. Brněnští vědci vyvíjejí systém, který bezpečně pozná hlas určitého člověka. In: *IHNED.cz* [online]. 23.6.2010 [cit. 2012-02-13]. Dostupné z: <http://tech.ihned.cz/veda/c1-44432190-brnensti-vedci-vyvijeji-system-ktery-bezpecne-pozna-hlas-urciteho-cloveka>
2. HAJIČOVÁ, Eva, PANEVOVÁ, Jarmila a SGALL, Petr. Úvod do teoretické a počítačové lingvistiky. Svazek I, Teoretická lingvistika. 1. vyd. Praha: Karolinum, 2002. 156 s. Učební texty Univerzity Karlovy v Praze. ISBN 80-246-0470-1.
3. Hlasové vyhledávání. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2011 [cit. 2012-02-13]. Dostupné z: [http://cs.wikipedia.org/wiki/Hlasov%C3%A9\\_vyhled%C3%A1v%C3%A1n%C3%AD](http://cs.wikipedia.org/wiki/Hlasov%C3%A9_vyhled%C3%A1v%C3%A1n%C3%AD)
4. Jak nainstalovat a konfigurovat funkci rozpoznávání řeči v systému Windows XP. In: *Jak nainstalovat a konfigurovat funkci rozpoznávání řeči v systému Windows XP* [online]. [cit. 2012-02-13]. Dostupné z: <http://support.microsoft.com/kb/306537/cs>
5. KINCL, Jiří. Rozpoznávání a syntéza řeči. Praha, 1985.
6. Microsoft oznámil průlom v oblasti rozpoznávání hlasu. In: *Živě.cz* [online]. 2.11.2011 [cit. 2012-02-13]. Dostupné z: <http://www.zive.cz/clanky/microsoft-oznamil-prulom-v-oblasti-rozpoznavani-hlasu/sc-3-a-158586/default.aspx>
7. Nádraží - hlasový dialogový systém. In: *Youtube* [online]. 2010 [cit. 2012-02-13]. Dostupné z: [http://www.youtube.com/watch?v=VZ8I5M\\_mu2M](http://www.youtube.com/watch?v=VZ8I5M_mu2M)
8. Navigace, kterou můžete ovládat hlasem. In: *IDNES.cz* [online]. 3.10.2009 [cit. 2012-02-13]. Dostupné z: [http://mobil.idnes.cz/navigace-kerou-muzete-ovladat-hlasem-garmin-nuvi-860t-video-p4m-navigace.aspx?c=A080929\\_122414\\_navigace\\_kor](http://mobil.idnes.cz/navigace-kerou-muzete-ovladat-hlasem-garmin-nuvi-860t-video-p4m-navigace.aspx?c=A080929_122414_navigace_kor)
9. NOUZA, Jan, ed., KOLDOVSKÝ, Zbyněk, ed. a VÍCH, Robert, ed. Řeč a počítač: principy hlasové komunikace, úlohy, metody a aplikace: sborník článků. Vyd. 1. Liberec: Technická univerzita v Liberci, 2009. 235 s. ISBN 978-80-7372-548-8.
10. Ovládání PC hlasem. In: *ABC linuxu* [online]. 14.9.2005 [cit. 2012-02-13]. Dostupné z: <http://www.abclinuxu.cz/blog/Saljack/2005/9/ovladani-pc-hlasem>
11. Ovládání počítače hlasem s TYPLE. In: *EXTRA Windows.cz* [online]. 2.7.2009 [cit. 2012-02-13]. Dostupné z: <http://extrawindows.cnews.cz/ovladani-pocitace-hlasem-s-typole>
12. Počítač Vás poslechne na slovo. In: *IDNES.cz* [online]. 5.4.2009 [cit. 2012-02-13]. Dostupné z: [http://technet.idnes.cz/pocitac-vas-poslechne-na-slovo-a-to-doslova-f2x-software.aspx?c=A090403\\_153215\\_software\\_vse](http://technet.idnes.cz/pocitac-vas-poslechne-na-slovo-a-to-doslova-f2x-software.aspx?c=A090403_153215_software_vse)
13. Poznejte alternativní způsoby ovládání PC. In: *PCWorld* [online]. 9.3.2009 [cit. 2012-02-13]. Dostupné z: <http://pcworld.cz/hardware/poznejte-alternativni-zpusoby-ovladani-pocitace-6830>

14. Programy pro ovládání PC. In: *PETIT* [online]. 2009 [cit. 2012-02-13]. Dostupné z: [http://www.petit-os.cz/progr\\_pro\\_ovl\\_PC.php](http://www.petit-os.cz/progr_pro_ovl_PC.php)
15. Projekty. In: *Laboratoř počítačového zpracování řeči* [online]. 2011 [cit. 2012-02-13]. Dostupné z: <https://www.ite.tul.cz/speechlab/index.php/vyzkum/projekty.html>
16. PSUTKA, Josef. *Komunikace s počítačem mluvenou řečí*. 1. vyd. Praha: Academia, 1995. 287 s. ISBN 80-200-0203-0.
17. PSUTKA, Josef et al. *Mluvíme s počítačem česky*. Vyd. 1. Praha: Academia, 2006. 746 s. Česká matice technická; roč. 111, č. spisu 502. ISBN 80-200-1309-1.
18. Rozpoznávání řeči. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 7.2.2011 [cit. 2012-02-13]. Dostupné z: [http://cs.wikipedia.org/wiki/Rozpozn%C3%A1v%C3%A1n%C3%AD\\_%C5%99e%C4%8Di](http://cs.wikipedia.org/wiki/Rozpozn%C3%A1v%C3%A1n%C3%AD_%C5%99e%C4%8Di)
19. SIGMUND, Milan. *Rozpoznávání řečových signálů: přednášky*. Vyd. 1. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky, 2007. 122 s. ISBN 978-80-214-3526-1.
20. SIGMUND, Milan. *Speaker recognition: identifying People by their Voices: habilitation thesis [disertace]*. Brno: Vutium, 2000. 21 s. Vědecké spisy Vysokého učení technického. sv. 25. ISBN 80-214-1590-8.
21. Speech recognition. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2010 [cit. 2012-02-13]. Dostupné z: [http://en.wikipedia.org/wiki/Speech\\_recognition](http://en.wikipedia.org/wiki/Speech_recognition)
22. Strojový překlad. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 1.12.1008 [cit. 2012-02-13]. Dostupné z: [http://cs.wikipedia.org/wiki/Strojov%C3%BD\\_p%C5%99eklad](http://cs.wikipedia.org/wiki/Strojov%C3%BD_p%C5%99eklad)
23. ŠÍCHA, Miloš. *Elektronické zpracování signálů: základy analogové, digitální techniky*. 2. přeprac. vyd. Praha: Karolinum, 1998. 136 s. ISBN 80-7184-455-1.
24. UHLÍŘ, Jan a kol. *Technologie hlasových komunikací*. Vyd. 1. Praha: Nakladatelství ČVUT, 2007. vii, 276 s. ISBN 978-80-01-03888-8.
25. VoiceMate Professional. In: *Slunečnice.cz* [online]. 2009 [cit. 2012-02-13]. Dostupné z: <http://www.slunecnice.cz/sw/voicemate-professional/>
26. VoiceXML. In: *Hlasová komunikace pod kontrolou* [online]. 15.9.2010 [cit. 2012-02-13]. Dostupné z: <http://www.optimsys.cz/cs/technologie/voicexml>
27. Wolfgang von Kempelen. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 15.11.2006 [cit. 2012-02-13]. Dostupné z: [http://cs.wikipedia.org/wiki/Wolfgang\\_von\\_Kempelen](http://cs.wikipedia.org/wiki/Wolfgang_von_Kempelen)

28. Zpracování přirozeného jazyka. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 8.3.2009 [cit. 2012-02-13]. Dostupné z: [http://cs.wikipedia.org/wiki/Zpracov%C3%A1n%C3%AD\\_p%C5%99irozen%C3%A9ho\\_jazyka](http://cs.wikipedia.org/wiki/Zpracov%C3%A1n%C3%AD_p%C5%99irozen%C3%A9ho_jazyka)

## **SUMMARY**

The aim of the thesis was to describe the design and implementation options of voice control digital maps. First, describe the theoretical possibility of geoinformatics, which could be beneficial, and voice module then apply this knowledge to create their own applications to control the maps and voice commands for creating appropriate working with the map.

The first part of the thesis is focused on the theoretical part, which can be described using the voice module. Geoinformatics area is still in use voice technology only in the beginning, and therefore the voice module to facilitate the work still is not implemented almost anywhere. They were also tested existing programs that are aimed at controlling the computer voice. It is not always possible for various reasons, to run the program and test it. Also the possibility of individual programs vary considerably on the one hand, programs are already able to recognize speech nearly flawlessly and is close to a fully functional program. On the other hand, programs that only pretend to be able to control a computer voice, and their functionality is very negligible level.

Theoretical knowledge was applied in the second part of the thesis, which was created example application to map control voice. The application is composed of two programs, and ArcExplorer JetVoice. ArcExplorer viewer as maps and JetVoice a program that can on its own voice commands to control the map voice. The application includes 20 preset voice commands for working with a basic map. Voice commands were selected based on the analysis and comparison of the suitability of the Czech language and English language, when English was selected as the better option. Voice commands were selected with regard to ease of recall and ease of pronunciation, as the two basic rules for the creation of appropriate voice commands.



## **PŘÍLOHY**

# SEZNAM PŘÍLOH

## **Vázané přílohy:**

Příloha 1 Vyhodnocení testování programů na ovládání počítače hlasem

Příloha 2 Seznam vytvořených hlasových povelů

## **Volné přílohy**

Příloha 3 DVD

## **Popis struktury DVD**

Adresáře:

*Diplomova práce*- text diplomové práce ve formátu PDF

*Instalacni\_soubory*- instalační soubor programu JetVoice a ArcExplorer

*www\_stranky*-vytvořené www stránky k diplomové práci

*Vytvorena\_aplikace*- vytvořená aplikace na ovládání mapy hlasem

## Příloha 1

### Vyhodnocení testování programů na ovládání počítače hlasem

|                                 | Freeware              | Čeština | Složitost | Počet příkazů | Předvolené příkazy | Vlastní příkazy | Hodnocení |
|---------------------------------|-----------------------|---------|-----------|---------------|--------------------|-----------------|-----------|
| <b>MyVOICE</b>                  | NE (licence 6500 Kč)  | ANO     | MALÁ      | VELKÝ         | ANO                | ANO             | 90%       |
| <b>JetVOICE</b>                 | ANO                   | ANO     | MALÁ      | VELKÝ         | NE                 | ANO             | 75%       |
| <b>Respondigs Heads</b>         | ANO                   | NE      | STŘEDNÍ   | MALÝ          | ANO                | NE              | 20%       |
| <b>VoiceMate Professional</b>   | NE(cena není uvedena) | NE      | VELKÁ     | VELKÝ         | ANO                | ANO             | 45%       |
| <b>Perl-box Voice</b>           | ANO                   | NE      | STŘEDNÍ   | VELKÝ         | NE                 | ANO             | 60%       |
| <b>Typle</b>                    | ANO                   | NE      | STŘEDNÍ   | MALÝ          | NE                 | ANO             | 20%       |
| <b>e-Speaking</b>               | NE (14 \$)            | NE      | MALÁ      | VELKÝ         | ANO                | ANO             | 80%       |
| <b>Drafon NaturallySpeaking</b> | NE (100 popř 200 \$)  | NE      | MALÁ      | VELKÝ         | ANO                | ANO             | 90%       |
| <b>Software od Windows</b>      | ANO                   | NE      | MALÁ      | VELKÝ         | ANO                | ANO             | 90%       |

## Příloha 2

### Seznam vytvořených hlasových povelů

| Hlasový povel | Akce v mapě                 |
|---------------|-----------------------------|
| MAPA          | Otevře mapu České republiky |
| OLOMOUC       | Otevře mapu Olomouce        |
| RIGHT         | Vpravo                      |
| LEFT          | Vlevo                       |
| UP            | Nahoru                      |
| DOWN          | Dolů                        |
| IN            | Přiblížení                  |
| OUT           | Oddálení                    |
| WINDOW        | Uzavření podokna            |
| CLOSE         | Zavření celé mapy           |
| INFO          | Informace z mapy            |
| CENTRO        | Vycentrování ukazatele      |
| APPLE         | Vpravo 3krát                |
| SUN           | Vlevo 3krát                 |
| FISH          | Nahoru 3krát                |
| MOK           | Dolů 3krát                  |
| PASTA         | Vlevo nahoru                |
| BOOK          | Vlevo dolů                  |
| CAR           | Vpravo nahoru               |
| PEN           | Vpravo dolů                 |