



**Laboratoř počítačového
zpracování řeči
Fakulta mechatroniky
TU v Liberci**



Hlas a počítač

**Možnosti nasazení nových technologií v resortu školství
a jako pomůcka pro zdravotně postižené studenty**

Jan Nouza

1. Co umožňují (a co ještě neumí) hlasové technologie?
2. Proč je čeština pro počítač tak složitá?
3. Výzkumný tým na Technické univerzitě v Liberci a jeho produkty
4. Ovládání počítače hlasem
5. Diktování textu do počítače
6. Další možnosti využití hlasových technologií
7. Živé ukázky, video, diskuse

K čemu slouží hlasové technologie?




Podpora komunikace mezi člověkem a počítačem

- **Počítačová syntéza řeči** – „počítač mluví“
autom. hlášení na nádražích, čtení textových dokumentů, využití zejména pro nevidomé, v mobilních aplikacích,
- **Počítačové rozpoznávání řeči** – „počítač poslouchá a koná“
hlasové vytáčení, ovládání PC, diktát do počítače, přepis zpráv, rozhovorů, přednášek.....
- **Dialog s počítačem** – „počítač naslouchá a mluví“
automatické informační systémy, rezervační služby po telefonu, ...

Co umí a neumí hlasové technologie?

Hlasová syntéza

- **umí:** přečíst libovolný text, včetně zkratk a číslic
- **neumí:** mluvit tak přirozeně jako člověk 

Rozpoznávání řeči

- **umí:** a) poměrně spolehlivě rozpoznávat slovní povely
b) slušně zvládat úlohu diktování po slovech
c) s určitým omezením rozpoznávat plynulou řeč
d) obojí už dobře zvládá v tichém prostředí
e) nejdále dotaženo v jazycích jako je angličtina
- **neumí:** a) spolehlivě pracovat v hlučných podmínkách
b) rozpoznávat řeč osob mluvících současně
c) zvládat „ohebné“ jazyky jako je čeština
d) pochopit obsah mluvené řeči

Proč je rozpoznávání těžší než syntéza?

Hlasová syntéza

Nedokonalou řeč počítače vnímá **člověk vybavený intelektem** a schopností domýšlet si souvislosti.

Rozpoznávání řeči

Lidskou řeč, která je velmi složitá a různorodá, analyzuje **nedokonalý stroj** bez vlastní inteligence.

Klíčové problémy rozpoznávání řeči:

- ❑ **rozsáhlý slovník** (v češtině více než 2 miliony slov a tvarů)
- ❑ **přirozená řeč je plynulá** (mezi slovy nejsou pauzy, koartikulace, ...)
Text: Zavolej mi prosím v pátek z domova
Řeč: zavolejmiprosímfpátegzdomova
- ❑ **každý člověk mluví jinak** (jiná výslovnost, výška a barva hlasu, intonace, volba slov, momentální stav,)
- ❑ **řeč se nikdy neodehrává v úplném tichu** (mikrofon vždy snímá i okolní ruch, hluky, řeč jiné osoby, atd.)

Čeština jakožto ohebný jazyk

Porovnání slovní zásoby angličtiny a češtiny

(AJ ~ 50 tis. slovních tvarů, ČJ > 2 miliony)

Angličtina

driver, drivers, driver's

Čeština

řidič, řidiče, řidiči, řidičem,
řidička, řidičky, ...
řidičův, řidiččin,
řidičský, řidičská,

important

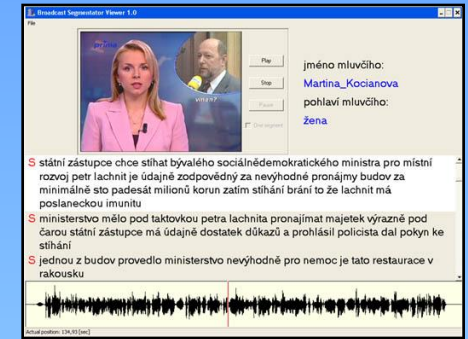
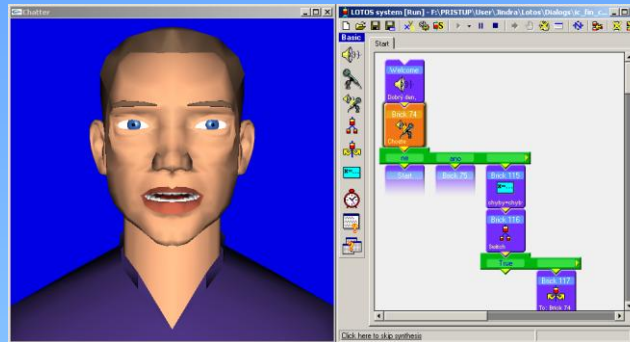
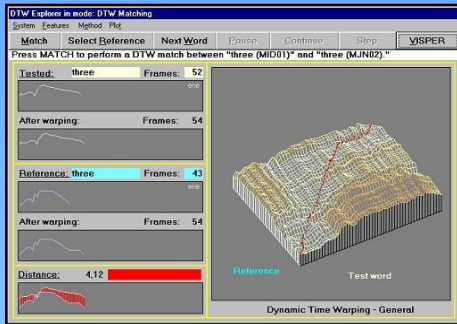
důležitý, důležitého, důležitá, ...
důležitější, nejdůležitější,

build, builds, built, building

stavit, stavím, stavíš, ...
stavil, stavili, stavily,
stav, stavící,
vystavit, postavit, zastavit, nestavit, ...

Důležité: Slovní tvar, který není ve slovníku rozpoznávacího systému, program nikdy nemůže rozpoznat.
Systémy pro ČJ nikdy nemohou mít kompletní slovní zásobu.

Co řešíme na TU v Liberci?



Výzkumný tým (12 osob, od roku 1993, národní i mezinárodní projekty, ocenění: Cena Ministrů za výzkum 2003, nominace na cenu MOSTY)

- **automatické rozpoznávání řeči se zaměřením na češtinu**
- **diktování do počítače**
- **automatický přepis** mluvené řeči a zvukových záznamů
- **hlasový dialog** s počítačem, zejména po telefonu
- **rozpoznávání osob** podle hlasu
- **audiovizuální komunikace** (rozpoznávání i syntéza s podporou vizuální informace)
- **hlasové technologie na pomoc postiženým**

Příklady aplikace výzkumu (1)

Program MyVoice – program na ovládání PC hlasem
Představen poprvé na INSPO 2005, nominace MOSTY 2006



Dita – kvadruplegička, první uživatelka systému, díky programu vystudovala SŠ a píše svůj blog <http://horochovska.blog.idnes.cz/>

MyVoice – princip činnosti

hlasové povely nahrazují klávesnici a myš
umožňují ovládat jakýkoliv program na PC

The screenshot illustrates the MyVoice application in use on a Windows XP desktop. The desktop environment includes several open windows and icons:

- DAEMON Tools**: A utility for mounting virtual drives.
- WinFast PVR**: A software interface for live TV recording and playback.
- Microsoft Word**: A document titled "Dokument1" containing the text: "Tento text byl nadiktován programem MyVoice." and "This text was input via software".
- Classic PhoneTools**: A virtual phone interface with a keypad and a list of contacts.
- MyVoice**: The main application window, showing a menu with options like "Seznam", "Nahoru", "Dolů", "Zavolej", "Polož", "Konec telefonu", "Základní skupina", and "Myš".

The MyVoice application interface includes a virtual phone keypad and a list of contacts:

Ime	Číslo
Doc. Doležal	3533
Jan Nouza	3208
Jindra Drábková	3182
Kancelář AP9	3484

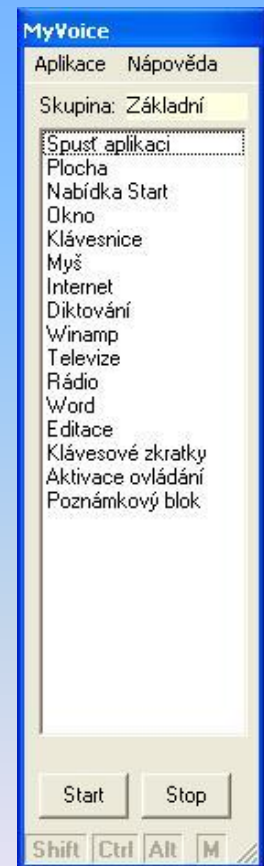
The application also features a "Speaker Phone" option and a "Konfigur." button. The desktop taskbar shows the Start button and several open applications, including Internet Explorer, Cool Edit 2000, and Zástupce - Dictate.

Jak funguje program MyVoice?

Ovládání PC bez použití rukou - pouze hlasem

- **Spouštění programů** z plochy, nabídky Start, z disku
- **Vkládání textu** po písmenech, případně po slovech nebo frázích (pokud je má uživatel předem připravené).
- Spouštění a synchronizace programu **MyDictate**
- **Ovládání elektronické pošty**, psaní i přijímání dopisů
- **Práce s internetem**
- **Programy pro zábavu i pro vážnou práci**, např. hry, programy pro kreslení, prohlížeče obrázků a videonahrávek, přehrávače hudby, apod.
- **Ovládání televize i rádia**, pokud je v počítači karta
- **Práce s prakticky libovolným programem**
- **Nejen pro hendikepované uživatele**

Cca 100 uživatelů v ČR (několik též v SR)



Příklady aplikace výzkumu (2)

SmartRoom – místnost s hlasovým ovládáním všech běžných spotřebičů (světlo, topení, větrák, rolety, TV, hifi, telefon, zámek)



V roce 2007 vybudována na TU v Liberci (ve spolupráci s firmou Jablotron)

Příklady aplikace výzkumu (3)

MyDictate - Diktování i editace textů hlasem (bez použití rukou).

The screenshot shows a Microsoft Word window titled "Diktát - Microsoft Word" with a menu bar (Soubor, Úpravy, Zobrazit, Vložit, Formát, Nástroje, Tabulka, Okno, Nápověda) and a toolbar. The main text area contains the following text:

Dobrý den,
tento text byl nadiktován hlasem bez použití rukou v programu MyDictate. Tento program má několik zajímavých funkcí

Below the text is a horizontal ruler and a status bar at the bottom of the window showing "Stránka 1 odd 1 1/1 na 6,3 cm ř. 5 sl. 29 ZÁZN REV ROZ PŘES Čeština".

Overlaid on the right side of the Word window is the MyDictate application window. It has a title bar "MyDictate" and a menu bar "Applikace Nápověda". The main area of the MyDictate window contains a text input field with the word "funkcí" in red. Below this is a section titled "VEZMI" followed by a list of words in a table:

první	funkci
druhý	funkcích
třetí	foukací
čtvrtý	funkcím
pátý	chodci
šestý	funkční
sedmý	soudci
osmý	Fuksy
devátý	skutcích
desátý	hudci

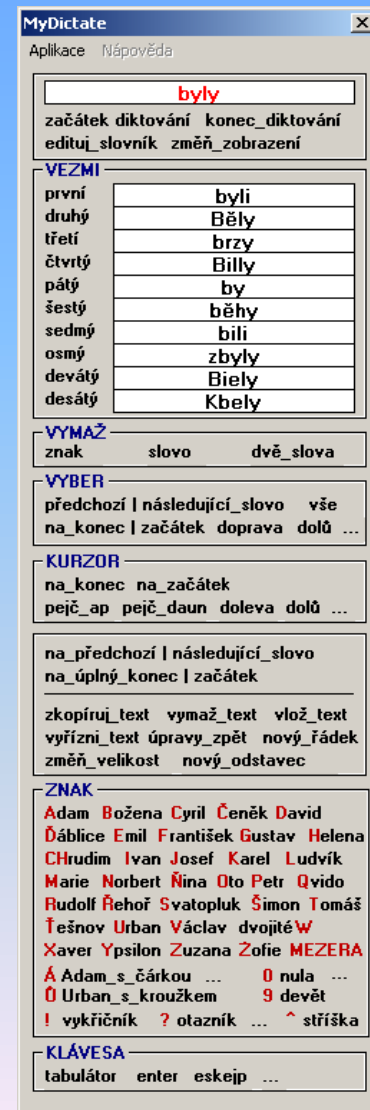
The Windows taskbar at the bottom shows several open applications: "Diktát - Microsoft ...", "handicap2007 - Mi...", "Bez názvu - Malování", and "MyDictate". The system tray on the right shows the date and time as "CS 16:06".

Jak funguje program MyDictate

Umožňuje diktovat text slovo po slovu:

- schopnost zvládnout několik set tisíc slovních tvarů – v poslední verzi 550 000 slov
- rozlišení mezi stejně znějícími slovy,
- možnost snadné opravy chyb,
- editace, formátování, apod.,
- možnost nadiktovat i slova, která nejsou ve slovníku (hláskováním),
- možnost přidávat slova do slovníku přímo při diktování,
- možnost průběžně zapínat i vypínat hlasové ovládání

První uživatel Jiří Zeman – v roce 2008 nadiktoval svou diplomovou práci (150 stran) na MU v Brně, dnes cca 60 uživatelů

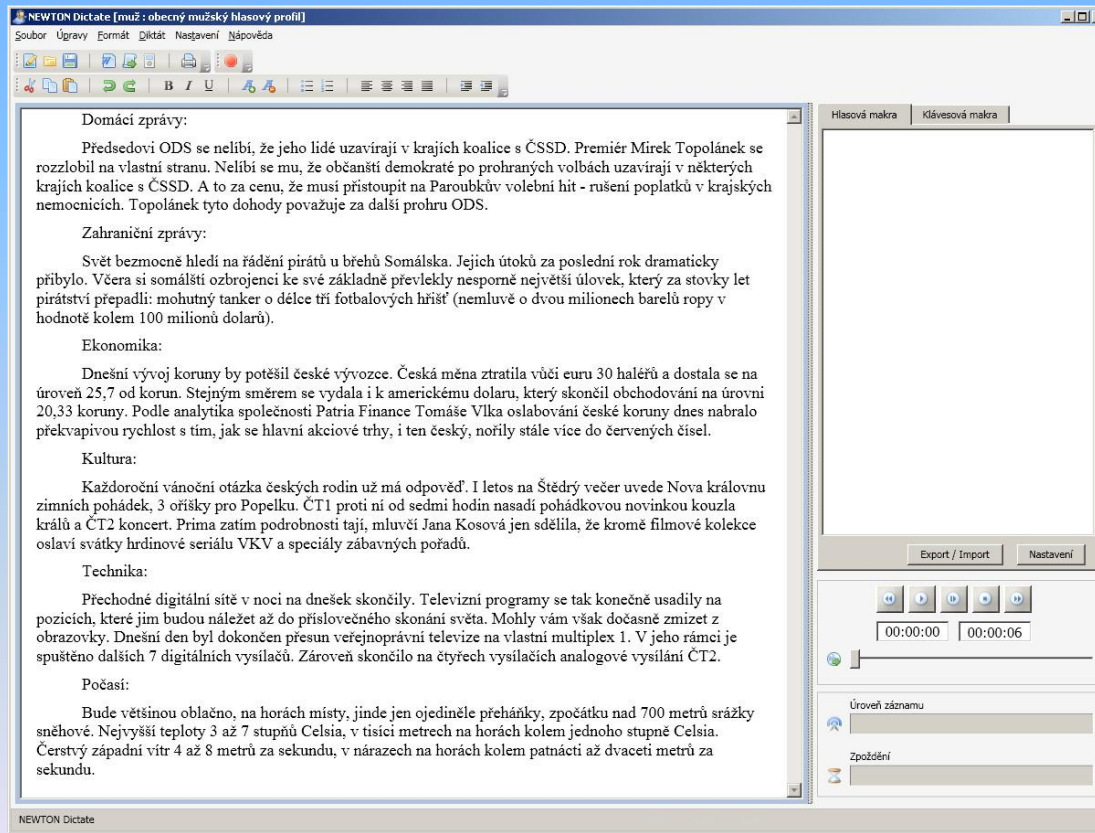


Používání programů MyVoice a MyDictate

- Program MyDictate spolupracuje s programem MyVoice – hendikepovaný člověk tak může hlasem ovládat PC i diktovat.
- Dita se naučila pracovat s programem MyDictate během jediného dne. Její zkušenosti jsou popsány na jejím blogu.
- Jiří Zeman nyní pracuje na Nejvyšším soudu v Brně a již rok hlasem diktuje jak běžné tak i velmi odborné texty, či SMS.
- Praktická úspěšnost diktování po zaučení je vyšší než 94 %, rychlost diktování je srovnatelná s psaním na klávesnici. Navíc neexistují překlepy.
- Distributorem programu je liberecká firma Fugasoft.
- Video, která ukazují, jak program pracuje, lze nalézt na stránce <http://www.fugasoft.cz/index.php?cont=mydictate>

Příklady aplikace výzkumu (4)

Newton Dictate – plynulé diktování (vyvinuto s firmou Newton Technologies) (350.000 slov), diktování všeobecných, soudních a lékařských textů



Možnosti využití ve školství – rychlá a efektivní tvorba textových dokumentů

Příklady aplikace výzkumu (5)

Automatické titulkování mluvených pořadů v reálném čase

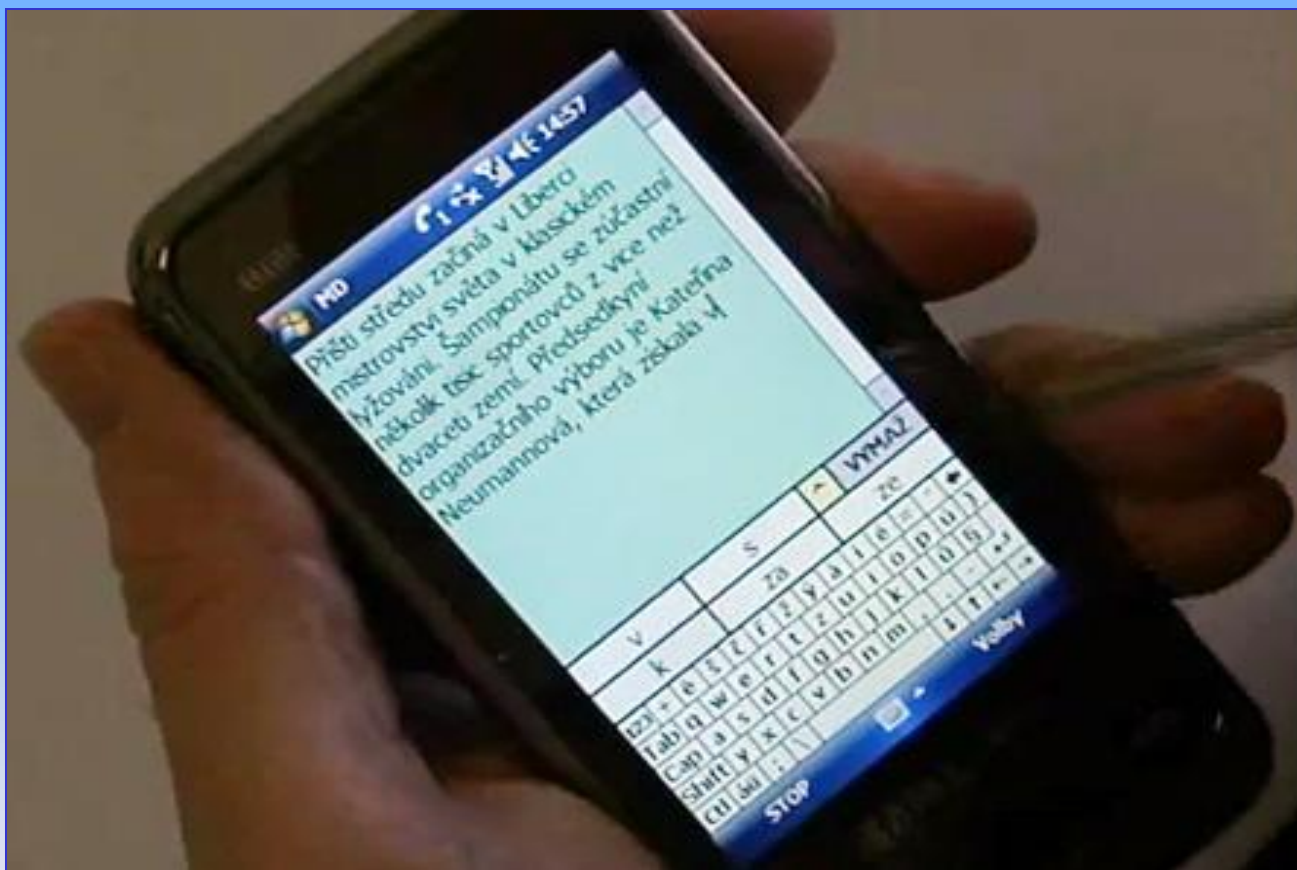


System snímá zvuk a okamžitě na obrazovku píše titulky. Přesnost přepisu závisí na kvalitě zvuku a na kvalitě mluvy (zatím cca 85 %).

Další možná aplikace pro školství – titulkování přednášek na VŠ

Příklady aplikace výzkumu (6)

Diktování do malých mobilních přístrojů



Videoukázky

1. *Skutečný uživatel - Dita Horochovská*
2. *Titulkování*
3. *Diktování do mobilu*
4. *Diktování do mobilu s překladem*