

Úvod do praxe stínového řečníka

Úvod

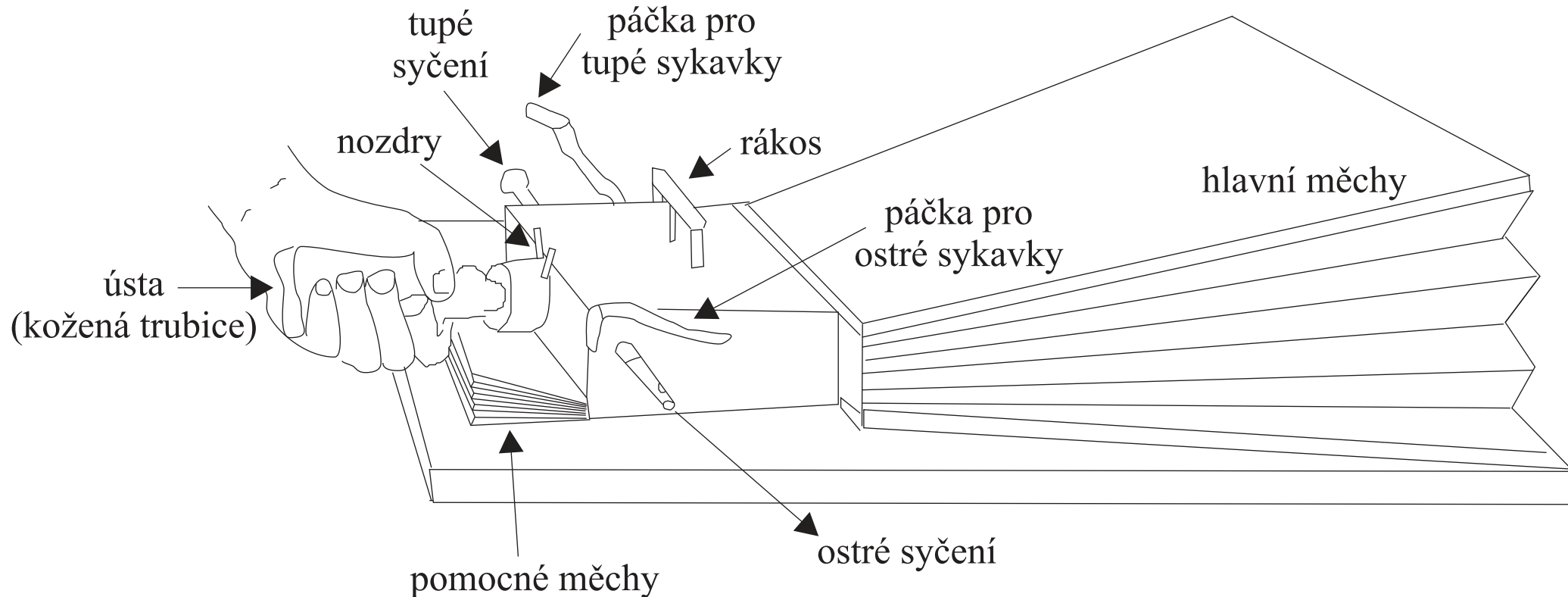
Kdo je to stínový řečník?

- Ukázka
- Simultánní tlumočnick z daného jazyka do téhož jazyka
- Jeho úkolem je přemlouvat televizní pořady tak, aby výsledná promluva byla vhodná pro rozpoznávací systém, který vyslovenou promluvu automaticky převede na titulky

Historie řečových technologií

- 1769 – Wolfgang Ritter von Kempelen (maďarský šlechtic narozený v Bratislavě a žijící ve Vídni) začal pracovat na mechanickém mluvícím stroji
- 1779 – německý profesor Christian Kratzenstein sestrojil akustické rezonátory napodobující hlasový trakt člověka, bylo jimi možné vytvářet samohlásky
- 1791 – von Kempelen představil první „syntetizér řeči“ schopný produkovat zvuky odpovídající hláskám

Von Kempelenův mluvicí stroj



(.wav)

Historie řečových technologií

- Rozvoj řečových technologií v současném pojetí nastal ve 20. století s rozvojem elektrotechniky
- 30. léta 20. století – 1. syntéza souvislé řeči (elektronický syntetizér VODER vyvinutý v Bell Labs)

VODER (Voice Operation DEMonstrator)

The image is a composite illustrating the Voder machine. On the left, a black and white photograph shows a woman sitting at a desk, operating the machine. A large screen behind her displays the word 'VODER' and a stylized graphic. The year '1936' is printed in the bottom left corner of this section. On the right, a close-up photograph shows a hand typing on the machine's keyboard. Below these is a detailed block diagram of the Voder's internal circuitry. The diagram includes a 'Generátor náhodného šumu' (Random noise generator) and an 'Oscilátor' (Oscillator). The noise generator outputs 'neznělý' (unvoiced) signals, while the oscillator outputs 'znělý' (voiced) signals. These signals pass through '10 rezonančních filtrů' (10 resonant filters). The filters are connected to a 'klávesnice' (keyboard) with keys numbered 1 through 10. A 'zápěstový přepínač' (Hand switch) is also connected to the filter bank. The output of the filters goes to a 'Zesilovač' (Amplifier), which is connected to a 'reproduktor' (Speaker). The keyboard also includes 't-d', 'p-b', and 'k-g' keys labeled as 'exploziv' (explosive), and a 'klávesa pro "ticho"' (Silence key). A 'nožní pedál pro kontrolu výšky hlasu' (Foot pedal for voice height control) is also shown. At the bottom of the diagram, two small icons represent audio files in '.wav' format.

Historie řečových technologií

- Největší pokrok učiněn s nástupem číslicových počítačů (2. polovina 20. století)
- Frederick Jelinek (Bedřich Jelínek) (1932 – 2010) – český vědec žijící v USA

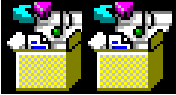




- je považován za hlavního autora teorie, podle které je možné mluvenou řeč na text převádět pomocí statistických metod (do té doby se strojové rozpoznávání mluvené řeči snažilo řídit lidským modelem, stroje měly rozpoznávat syntaxi, gramatiku a významy slov).

Oblasti řečových technologií

- Počítačová syntéza řeči
- Automatické rozpoznávání řeči
- Hlasové dialogové systémy
- Rozpoznávání řečníka (rozpoznávání lidí podle hlasu)
- Rozpoznávání emocí z hlasu

Počítačová syntéza řeči

- Syntéza řeči = proces umělého vytváření řeči (počítačem), převod textu na řeč
- Počáteční snaha – dosažení co největší srozumitelnosti syntetizované řeči (70. a 80. léta)  
(.wav)(.wav) (.wav) (.wav)
- Současná snaha – dosažení co největší přirozenosti syntetizované řeči 
(.wav) (.wav)
- Konečný cíl – vytvářet řeč v takové formě a kvalitě, aby nebyla rozpoznatelná od řeči člověka (opravdu to ale chceme?! – etický problém)

Využití syntézy řeči

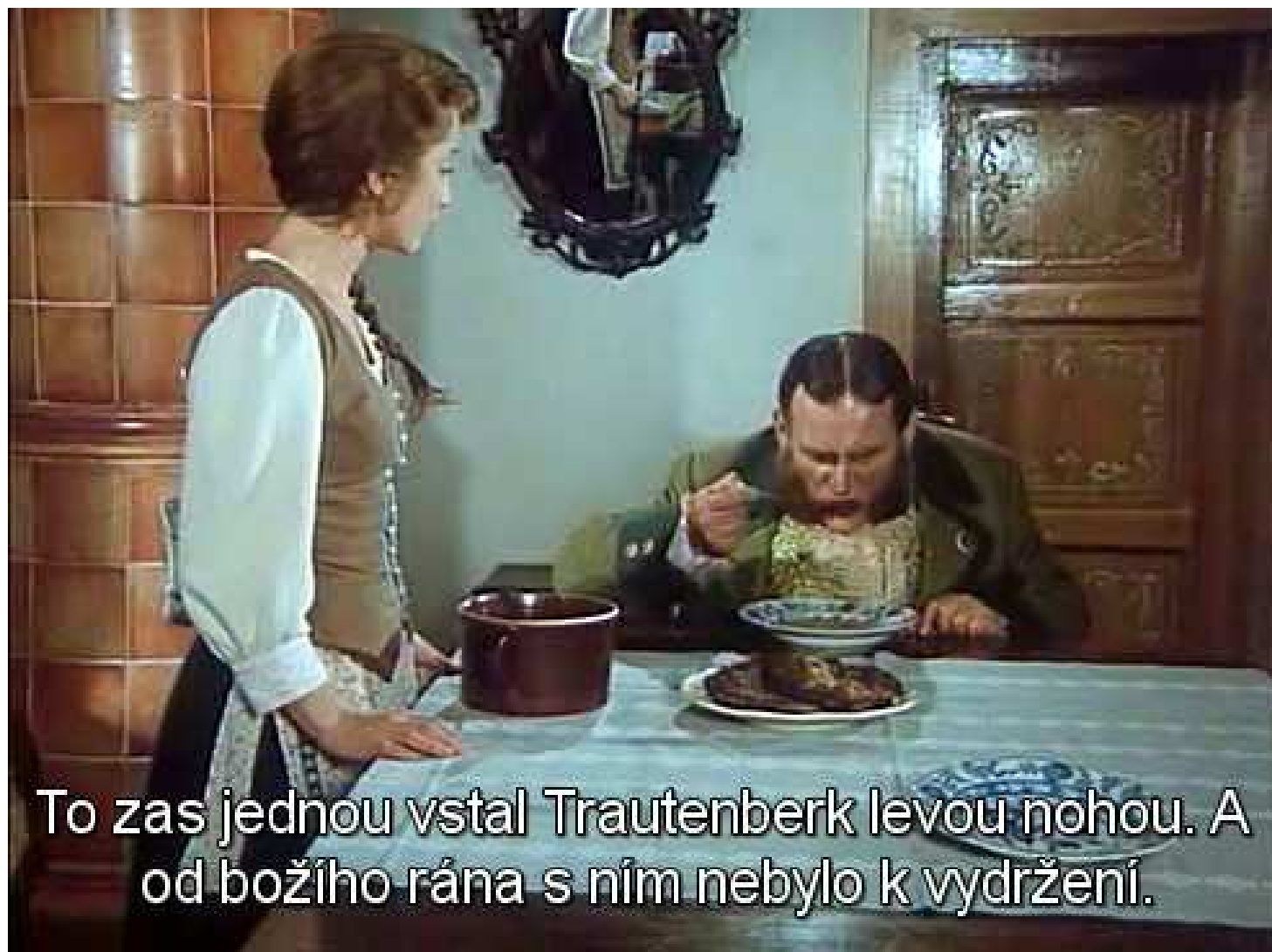
Všude tam, kde

- není jiná možnost, jak přijímat informace
 - aplikace pro nevidomé
 - telefonní aplikace
 - ...
- člověk-uživatel má zaměstnán zrak jinými činnostmi
 - operátor složitého přístroje
 - řidič auta
 - pilot letadla
 - ...

Konkrétní příklady aplikace syntézy řeči

- pomoc handicapovaným lidem (nevidomí, lidé s poruchami hlasu)
- telekomunikační služby (call centra . . .)
- informační služby (hlasová navigace v automobilech)
- dialogové systémy
- automatické čtení (SMS, e-maily, e-knihy, . . .)
- multimediální systémy (automatický dabing)
- zábavní průmysl (hračky, hry)
- výzkum (lingvistika, fonetika)
- výuka jazyků

Počítačová syntéza řeči - ukázka



To zas jednou vstal Trautenberk levou nohou. A od božího rána s ním nebylo k vydržení.

Ukázky současných systémů syntézy řeči

- Komerční sféra

- NUANCE VOCALIZER
- AT&T NATURAL VOICES
- ACAPELA GROUP
- IVONA TTS



- Akademická sféra

- Festival (Edinburgh)
- HTS (Nagoya)
- Mary TTS (DFKI, Saarbrücken)
- Epos TTS (ÚFE AV ČR Praha)
- ARTIC (KKY FAV ZČU Plzeň)



Automatické rozpoznávání řeči

- Rozpoznávání řeči = proces převodu řeči na text
 - Rozpoznávání izolovaně vyslovených slov
 - Rozpoznávání souvislé řeči

 - systémy s malým slovníkem (hlasové povely)
 - systémy se středně velkým slovníkem (omezené úlohy – předpověď počasí)
 - systémy s velkým slovníkem (lze hovořit přirozeně na jakékoli téma)

 - systémy závislé na řečnících (1 uživatel systému)
 - systémy nezávislé na řečnících (více uživatelů)

Ukázka rozpoznávání řeči – projekt ELJABR On-line titulkování sportovních přenosů



Ukázka rozpoznávání řeči – projekt ELJABR



On-line titulkování zasedání parlamentu



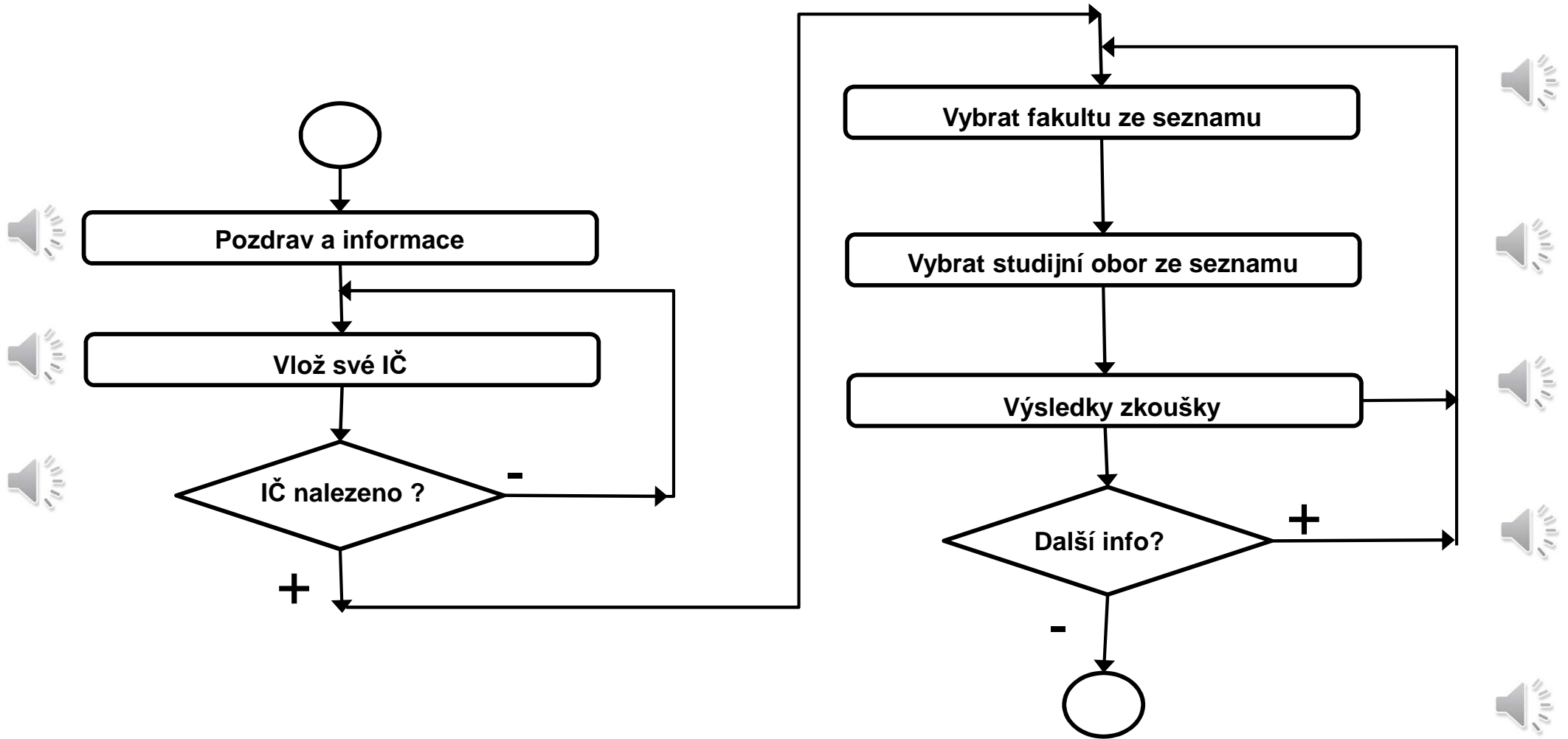
Hlasové dialogové systémy

- Umožňují uživatelům komunikovat prostřednictvím hlasu s počítačovými aplikacemi
- Kombinují systémy rozpoznávání a syntézy řeči
- Často se využívá telefonní rozhraní
- V současnosti pouze pro omezenou aplikační oblast

Využití hlasové dialogových systémů

- přístup k databázovým, informačním nebo monitorovacím systémům, zákaznická podpora
 - informace o odjezdech a příjezdech vlaků  
 - informace o přijímacím řízení
 - zákaznická samoobsluha (telefonní operátoři)
 - ...

Dialogový systém „Přijímací řízení“



Rozpoznávání řečníka a emocí

- Rozpoznávání řeči – cílem je určit, **co** se říká
- Rozpoznávání řečníka – cílem je určit, **kdo** to říká
 - Identifikace řečníka
 - Verifikace řečníka
- Rozpoznávání emocí – cílem je určit, **jak** se to říká

Využití systémů rozpoznávání řečníka

- bezpečnostní systémy
 - přístup do budov
 - přístup k databázím
 - telefonní transakce
- kriminalistika
 - anonymní telefonáty
 - analýza odposlechů

Využití systémů rozpoznávání emocí

- Kontrola psychického stavu uživatele, tj.
 - operátorů na velínech (např. v jaderných elektrárnách)
 - řídicích letového provozu
 - řidičů
 - ...

Kde se řečové technologie vyvíjejí?

- Akademická pracoviště (ČR)
 - ČVUT Praha - rozpoznávání řeči (zejména v zašuměném prostředí), syntéza řeči (hlavně prozodie)
 - VUT Brno – rozpoznávání řeči, rozpoznávání řečníka
 - Technická univerzita Liberec – rozpoznávání řeči
 - ZČU v Plzni – rozpoznávání řeči, syntéza řeči, rozpoznávání řečníka, dialogové systémy

Projekty ZČU, katedry kybernetiky

- Projekty [ELJABR](#) a [ELJABR II](#) (ELiminace JAzykových BaRiér handicapovaných diváků České televize) – pro Českou televizi
- Projekt [MALACH](#) (Multilingual Access to Large ArCHives) – vyhledávání ve zvukových archivech výpovědí z období holocaustu
- Projekt [MEGAWORD](#) – diktovací systém pro převod řeči na text s velmi velkým slovníkem
- Projekt IT-BLP (Intelligentní Technologie pro zvýšení Bezpečnosti Letového Provozu) – vývoj systému inteligentní komunikace mezi operátory řízení letového provozu a automatickým „počítačovým“ pseudopilotem

Kde se řečové technologie vyvíjejí?

- Komerční sféra – svět
 - Nuance, Acapela, Amazon – syntéza řeči
 - Google, Nuance, Microsoft, Apple – rozpoznávání řeči, dialogové systémy
- Komerční sféra – ČR (zejména rozpoznávání řeči)
 - NewtonTechnologies – systém NewtonDicate (TU Liberec)
 - NovaSoft – systém [NovaVoice](#) (ZČU Plzeň)
 - SpeechTech, s.r.o. – spin-off firma ZČU (řešení „na míru“ nejen pro malé zákazníky - např. Škoda Auto a.s., Česká televize)