Aktuální výsledky výzkumu ZČU v Plzni v oblasti podtitulkování pořadů pro neslyšící

Obsah:

1. Řešené projekty na podporu diváků ČT se specifickými potřebami
2. Automatické podtitulkování „živých" pořadů
3. Vytváření alternativní zvukové stopy z titulků (automatický dabing)
4. Generování znakového jazyka systémem Avatar
5. Závěr

Výzkumné projekty:

Eliminace jazykových bariér handicapovaných diváků České televize (ELJABR)

soutěž: Národní program výzkumu II (poskytovatel MŠMT)

období: r. 7/2006 až 6/2011

Eliminace jazykových bariér handicapovaných diváků České televize II (ELJABR_II)

soutěž: Podpora aplikovaného výzkumu ALFA (poskytovatel TAČR)

období: r. 7/2011 až 6/2016

Řešitelé projektů: Katedra kybernetiky FAV ZČU v Plzni, SpeechTech, s.r.o.

Finanční spoluúčast: Česká televize

CÍLE PROJEKTŮ ELJABR

Projekty ELJABR usilují o eliminaci jazykových bariér tří skupin sluchově handicapovaných televizních diváků:

1) sluchově postižených diváků

Cílem je opatřovat zvláště „živé" programy skrytými podtitulky, a to technologií automatického rozpoznávání řeči

2) starších, sluchově nebo mentálně postižených diváků, kteří nejsou schopni vnímat doprovodný zvuk, vadí jim snížená srozumitelnost reálných dialogů a doprovodné efektové složky

Cílem je vytvářet automaticky alternativní „klidnou" zvukovou stopu, a to technologií počítačové syntézy řeči

3) diváků, kteří komunikují znakovým jazykem (pouze v projektu ELJABR_II)

Cílem je hledání možností automatizovaného generování znakového jazyka systémem Avatar

Řešení cíle 1: Podtitulkování „živých" pořadů

- **Stenografem**, který využívá speciální stenografickou klávesnici
  - *Výhody:* trénovaní stenografové umí až 200 slov/min
  - *Nevýhody:* stenografů je absolutní nedostatek, jejich trénink trvá 2 až 3 roky, jsou velmi drazí a nevydrží pracovat dlouho
- **Písařem, který využívá velotypu** (speciální slabikový přepis)
  - *Výhody:* trénovaní velotypisté umí až 120 slov/min
  - *Nevýhody:* velotypistů je absolutní nedostatek, jejich trénink trvá nejméně 1 rok, jsou velmi drazí a nevydrží pracovat dlouho
- **Písařem, který využívá QWERTZ klávesnici**
  - *Výhody:* QWERTZ písařů je větší množství (umí až 90 slov/min)
  - *Nevýhody:* při přepisu často chybují, nevydrží pracovat dlouho
- **Automatickým rozpoznáváním řeči**
  - *Výhody:* moderní technologie, která zcela nebo zčásti automatizuje namáhavou lidskou činnost
  - *Nevýhody:* skutečné nasazení technologie vyžaduje „špičkové" zvládnutí náročných teoretických postupů a využití technických prostředků

Dva přístupy k automatickému vytváření podtitulků

- Počítačovým rozpoznáváním řeči **z doprovodné zvukové stopy**

  Výhody: levný provoz; „snaha" o doslovný přepis dialogu

  Nevýhody: relativně nízká přesnost rozpoznávání (60 až 85%), která je způsobena zejména:
  - potřebou rozpoznávat převážně spontánní řeč
  - častou přítomností hudby a hluku na pozadí dialogu
  - častým současným mluvením více řečníků

  Využití: při zpracování kvalitní zvukové stopy, kdy v danou chvíli mluví vždy jen 1 řečník; je známa tematická oblast dialogu

- Počítačovým rozpoznáváním řeči **s využitím „stínového řečníka"**

  Výhody: možnost dosažení vysoké přesnosti vytvářených titulků (>95%)

  relativní dostupnost „stínových řečníků" (intenzivní trénink „stínového řečníka" by neměl trvat déle než 3 až 6 měsíců)

  systém může být adaptován na hlas stínového řečníka

  Nevýhody: vytváření titulků není plně automatické (stínový řečník); nejde o doslovný přepis dialogu

  Využití: TV pořady diskusního charakteru anebo pořady s rušivým pozadím; vhodné je znát předem tematickou oblast zpracovávaného dialogu

Automatické vytváření podtitulků

- **Systém automatického rozpoznávání řeči**
  - systém pracuje na principu statistické indukce
  - strojovým učením jsou trénovány parametry tzv. **akustického** a **jazykového modelu**
  - akustický model má více než 20M parametrů a je trénován ze stovek hodin řeči namluvené více než 1 tisícem řečníků;
  - u jazykového modelu se trénují s využitím rozsáhlých textových zdrojů statistiky řazení slov a je vytvářen výslovnostní slovník (slovník má více než 1 milion slov);
  - statistiky řazení slov i slovníky jsou silně problémově závislé;
  - systém pracuje v reálném čase na běžném 4-jádrovém počítači (i notebooku).
  - analýza pořadů ČT, výběr pilotních pořadů pro podtitulkování
    - podtitulkování z **doprovodné zvukové stopy** – přenosy jednání PS Parlamentu ČR
    - podtitulkování s využitím **stínového řečníka** – diskusní pořad „Otázky Václava Moravce", Hyde park, sportovní přenosy (hokej, fotbal)

Automatické vytváření podtitulků

- **Výsledky řešení – podtitulkování z doprovodné zvukové stopy**
  - 26.11.2008 proběhl první test on-line podtitulkování přenosu jednání PS Parlamentu ČR (vysíláno na ČT24, skryté podtitulky na teletextové stránce 888)
  - od 12/2008 do 4/2010 probíhal souvisle experimentální provoz
  - od 5/2010 je podtitulkování pořadu řešeno jako služba (pro ČT poskytují řešitelé projektu, dosud vybaveno podtitulky více než 700 odvysílaných hodin)
  - technické řešení:
    - systém rozpoznávání řeči a SW modul vytvářející titulky jsou umístěny na ZČU a komunikují s ČT po telefonní lince
    - modul vytváření podtitulků „láme" souvislý proud slov (výstup ze systému rozpoznávání) do podtitulků a „gramatický korektor" doplňuje do podtitulků interpunkci a opravuje některé typy chyb
    - přesnost vytvářených podtitulků se pohybuje od 87 do 93%
    - bylo navrženo a realizováno poměrně unikátní zařízení umožňující ještě před odesláním každého podtitulku do ČT provádět jeho „okamžitou" manuální korekci (při opravách může pracovat současně i několik korektorů, a to po Internetu z různých lokalit) – lze využít v případě snížené přesnosti automaticky vytvářených podtitulků
    - (ukázka titulkování z doprovodné zvukové stopy)

Automatické vytváření podtitulků

- **Výsledky řešení projektu – podtitulkování s využitím stínového řečníka (1):**
  - první experimenty s přípravou stínových řečníků ukázaly nutnost jejich systematické přípravy
  - byl navržen a realizován **trenažér pro výuku stínových řečníků** (chráněno patentem) a navržena metodika jejich tréninku
  - originalita trenažéru spočívá v tom, že je stanoven přesný postup výuky při němž lze přesně měřit vzrůstající schopnosti stínového řečníka => k reálnému nasazení stínového řečníka pro titulkování dojde až po dosažení předem stanovené přesnosti vytvářených podtitulků
  - co musí stínový řečník při vytváření podtitulků zajistit:
    - přemlouvat TV dialog s cílem dosáhnout max. přesnosti rozpoznání své promluvy (při zajištění sémantického obsahu dialogu)
    - vkládat do systému (on-line) nová slova (většinou nová jména či místní názvy), je-li to potřeba
    - doplňovat interpunkci; označovat změnu řečníka
    - provádět případnou opravu podtitulku před jeho odesláním
    - pokud je potřeba odeslat podtitulek

Automatické vytváření podtitulků

- **Výsledky řešení projektu – podtitulkování s využitím stínového řečníka (2):**
  - v květnu 2011 byl spuštěn experimentální provoz podtitulkování diskusního pořadu „Otázky Václava Moravce"
  - od prosince 2011 je tento pořad pravidelně podtitulkován formou stálé služby (zajišťují řešitelé projektu) – zatím titulkováno více než 70 hod přenosů
  - v únoru 2012 bylo spuštěno pravidelné podtitulkování pořadu Hyde park, opatřeno podtitulky cca 85 pořadů (více než 80 hod)

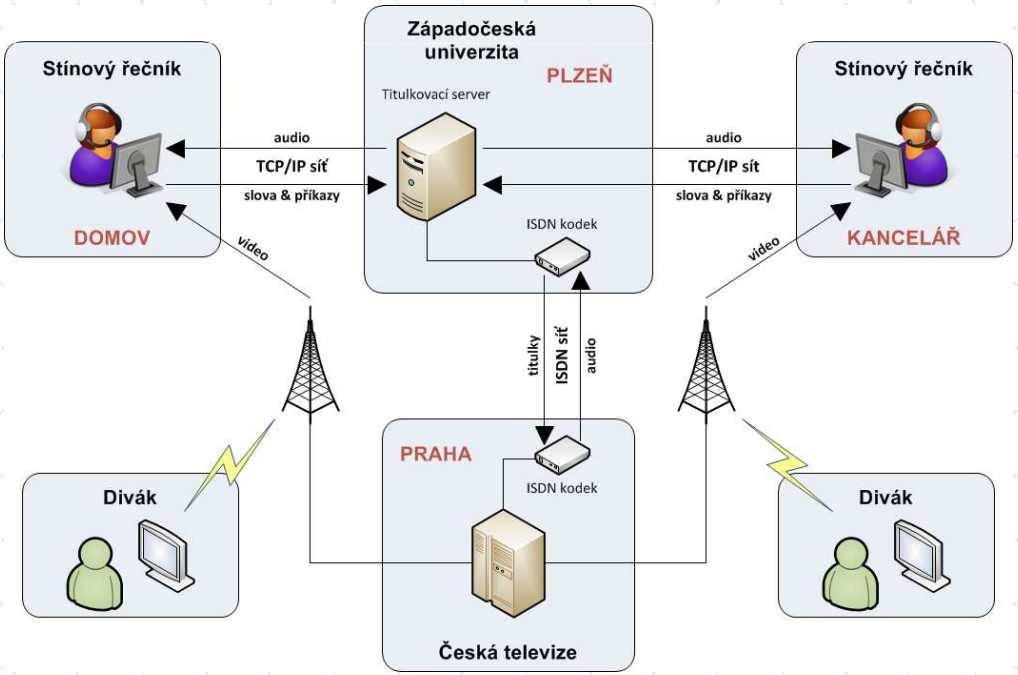
    (ukázka titulkování s využitím stínového řečníka)

  - v květnu 2012 byly v experimentálním provozu podtitulkovány 3 zápasy českých hokejistů na MS v hokeji (podtitulkován vlastní přenos + studia)
  - v červnu 2012 probíhá experimentální provoz podtitulkování 2 zápasů českých fotbalistů na ME v kopané
  - podtitulkování všech „živých" pořadů (včetně sportovních) má velmi pozitivní ohlas v komunitě neslyšících

    (ukázka titulkování s využitím stínového řečníka)

Automatické vytváření podtitulků

- **Výsledky řešení – titulkování s využitím stínového řečníka (2)**
  - systém rozpoznávání pracuje se slovníkem 1 mil. slov, akustický model je adaptován na hlas konkrétního stínového řečníka
  - přesnost podtitulků u pěti vyškolených stínových řečníků je vyšší než 98%
  - poměrně originální řešení se týká možnosti lokalizace stínového řečníka při pod titulkování „živého" pořadu – je vázáno pouze dostupností Internetu => lze pracovat např. z domova !!



Řešení cíle 2: Vytváření doprovodné zvukové stopy TV vysílání

- **Motivace:**
  - existuje poměrně velká skupina TV diváků, kteří nejsou schopni sledovat víceplánovou zvukovou stopu moderních TV programů
    - snížená srozumitelnost reálných dialogů, emoce, změny hlasu a tempa řeči
    - podkresová hudební složka
    - efekty na pozadí

    (ukázka dynamické stopy současných TV pořadů)

- **Řešení:** vytvářet alternativní doprovodnou „klidnou" zvukovou stopu
  - **vytvořením nového zvukového mixu**

    Výhody: profesionální kvalita

    Nevýhody: navýšení financí, porušení licenčních práv → **nerealizovatelné**
  - **počítačovou syntézou řeči z textu**

    Výhody: moderní technologie, automatizace dabingu
    (hlas vytvářen ze skrytých podtitulků, vysílaných na teletextové s. 888)

    Nevýhody: skutečné nasazení technologie vyžaduje „špičkové" zvládnutí náročných teoretických postupů syntézy řeči

Počítačová syntéza řeči

- IT technologie, která umožňuje převádět psaný text na mluvenou řeč
- Cíl: generovat řeč z **libovolného** textu
- **Není možné uložit všechna slova (věty) do počítače, a pak je jen přehrávat!**

- Na ZČU v Plzni je systém počítačové syntézy řeči vyvíjen od roku 1997
- Konkatenační korpusově orientovaný systém
- Komerční verze distribuována firmou SpeechTech

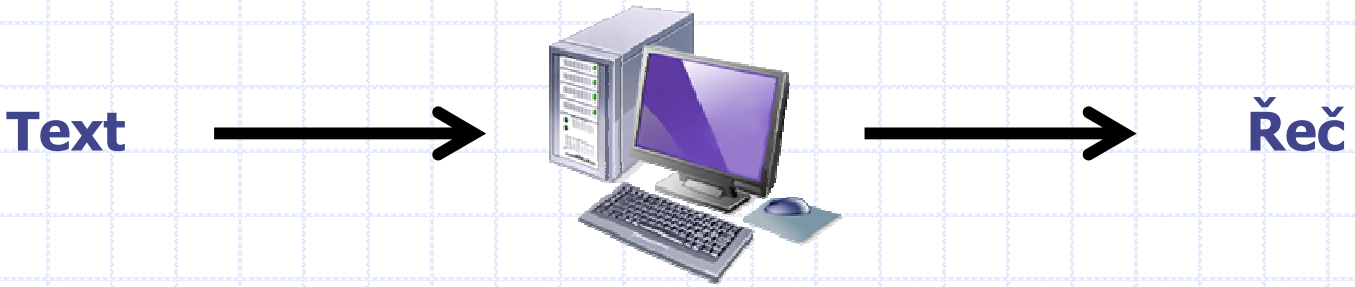


Schéma procesu TTS

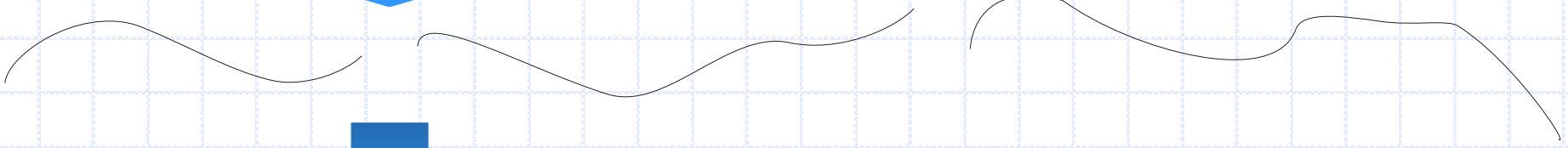
Dnes bude zataženo, v některých oblastech přeháňky, po 6. hod. očekáváme sněžení.

dnez bude zataženo vněkterích oblastech přeháňki pošesté hoďiňe očekáváme sněžeňí

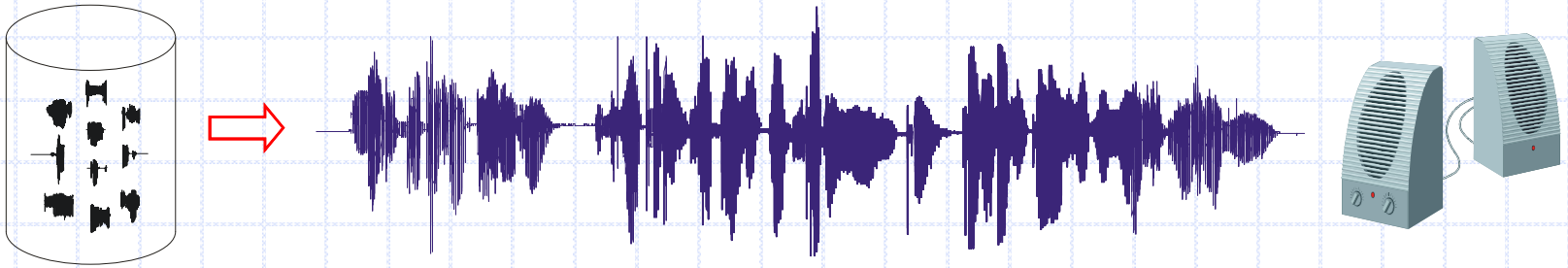
textová analýza, fonetická transkripce, prozodická slova

nádech pauza pauza pauza

prozodická analýza, intonační a rytmický průběh



výběr, spojování a úprava základních řečových jednotek



Specifické problémy při automatickém vytváření doprovodné zvukové stopy

- Vytváření řeči ze skrytých podtitulků TV vysílání (formát EBU STL)
- Synchronizace původní video a vytvářené audio stopy
- Adaptace systému pro potřeby vytváření doprovodné zvukové stopy
  - implementace algoritmu nelineárního zrychlování řeči
  - urychlení algoritmu syntézy řeči (možnost efektivní práce s rozsáhlými hlasovými inventáři v reálném čase)
- Vývoj nástroje pro automatickou detekci problematických podtitulků
  - umožňuje automaticky najít a označit podtitulky způsobujících příliš rychlé tempo syntetické řeči
  - možnost iterativní opravy (zjednodušení) textu podtitulků
  - využití ve fázi přípravy skrytých podtitulků nových TV pořadů ČT
- Vytvoření 4 nových vysoce kvalitních syntetických hlasů (2M a 2Ž)

Ukázky „monohlasový" pořad

míchání 2 vv nebo 4 hlasů

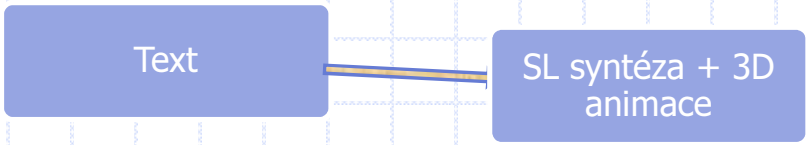
Řešení cíle 3: Generování znakového jazyka systémem Avatar

- **3D animace znakující postavy kopírující pohyby SL řečníka**
  - Promluvu vytváří znakující řečník ve studiu (u sebe doma)
  - Může být vyžadováno speciální zařízení – datové rukavice, kamerový systém
  - Každý obrazový snímek znakujícího „řečníka" je popsán sadou parametrů s extrémně nízkým datovým tokem
  - Doplnění znakujícího 3D Avatara do pořadu až na obrazovce neslyšícího diváka, rozlišení a kvalita obrazu může být velmi vysoká
  - 3D znakující Avatar „kopíruje" pohyby znakujícího řečníka (tempo řeči, volba znaků, styl apod. jsou odvozeny od znakujícího řečníka)
  - Není použita automatická syntéza ani automatický překlad do znakového jazyka

Generování znakového jazyka systémem Avatar

- **Automatická 3D animace znakující postavy z textu**
  - Plně automatický přístup – náhrada tlumočníka
  - Přenáší se pouze titulky a znakující 3D Avatar je vykreslován na straně diváka
  - Řídicí příkazy pro ovládání znakujícího 3D Avatara jsou vytvářeny modulem překladu českého znakového jazyka pouze ze vstupního textu
  - Pohyby paží, tvar rukou a tváře jsou syntetické (uměle vytvořené v počítači)

    Problémy: stále nevyřešená úloha pro automatický překlad a následnou 3D animaci znakující postavy



Závěr

Současné a budoucí úkoly:

Cíl 1: Podtitulkování „živých" pořadů

- příprava jazykových modelů a slovníků pro podtitulkování pořadů dalších žánrů (dramatická tvorba, sportovní pořady apod.)
- práce na zvyšování přesnosti systému automatického rozpoznávání řeči, který je základem procesu podtitulkování

Cíl 2: Vytváření doprovodné zvukové stopy TV vysílání

- příprava dalších vysoce kvalitních hlasů
- možnosti míchání hlasů ve vytvořené zvukové stopě
- příprava testovacího vysílání

Cíl 3: Generování znakového jazyka systémem Avatar

- výzkum v oblasti 3D animace znakujícího Avatara kopírujícího pohyby SL řečníka
- výzkum plně automatické 3D animace znakujícího Avatara přímo z textu (řeší se v rámci jiného projektu)