

# Bezdrátový systém pro přenos synchronního zápisu v reálném čase (speech-to-text reporting) určeného pro větší skupinu uživatelů se sluchovým postižením nejen v podmínkách vysoké školy

Christoph DAMM, Svatoslav ONDRA, Jiří TUŽIL

Masarykova univerzita, {damm, ondra, tuzil}@teiresias.muni.cz

***Anotace:** Jednou z technických služeb, které Masarykova univerzita poskytuje pro svých více než 100 studentů se sluchovým postižením, je vizualizace mluvené řeči prostřednictvím synchronního zápisu (speech-to-text reporting). S rostoucím počtem studentů, kteří jsou závislí na tomto typu služby a kteří navštěvují tytéž výukové hodiny současně, hledalo naše pracoviště vyhovující způsob, jak v případě skupin se třemi a více studenty zajistit, aby byl textový zápis pohodlně sledovatelný všemi účastníky. Dočasné způsoby řešení, které Masarykova univerzita aplikovala dříve, byly založeny na duplikování obrazového výstupu zapisovatelova počítače, ale žádný z nich nesplňoval požadavky a očekávání studentů, ani zapisovatelů.*

*Z těchto důvodů Masarykova univerzita vyvinula svůj vlastní systém pro bezdrátový přenos synchronního zápisu v reálném čase (**Polygraf**), který tento příspěvek popisuje. Systém prostřednictvím bezdrátové sítě průběžně odesílá textový zápis pořizovaný zapisovatelem na libovolný počet příručních displejů, které jsou k dispozici uživatelům textového zápisu.*

## 1. Vstupní situace a dřívější řešení

Středisko pro pomoc studentům se specifickými nároky Masarykovy univerzity poskytuje studentům se sluchovým postižením (obvykle nedoslýchavým studentům opírajícím se o psaný český jazyk) vizualizaci mluvené řeči prostřednictvím synchronního zápisu (speech-to-text reporting). Mluvený projev (v podmínkách vysoké školy obvykle na přednášce či semináři, obecně však na jakékoliv akci podobného charakteru i mimo vysokou školu) je průběžně zaznamenáván zapisovatelem na počítači a student takto vznikající záznam sleduje na displeji.

S rostoucím počtem studentů, kteří jsou závislí na tomto typu služby a kteří navštěvují tytéž výukové hodiny současně, hledalo naše pracoviště vyhovující způsob, jak v případě skupin se třemi a více uživateli zajistit, aby byl textový zápis pohodlně sledovatelný všemi účastníky. V minulosti jsme problém dočasně řešili dvěma způsoby:

1. **Připojením druhé obrazovky k zapisovatelově notebooku** – část uživatelů sledovala zápis přímo na obrazovce notebooku a část na sekundární obrazovce. Toto řešení vyžadovalo pokaždé sestavit kabelové propojení notebooku a sekundární obrazovky (zajistit její elektrické napájení), což nebylo ideální organizačně ani časově, nakonec ani z důvodů estetických.
2. Zřízením uzavřené bezdrátové sítě, jejímž prostřednictvím byl **přenášen obrazový výstup zapisovatelova notebooku** (pomocí aplikace VNC) **na malé notebooky**, které měli k dispozici uživatelé. Protože přenášený obraz přesně kopíroval obrazovku zapisovatelova notebooku (už to nemusí být vždy výhodné a žádoucí), systém na straně uživatelů neumožňoval žádné individuální možnosti zobrazení sdíleného zápisu.

Ačkoliv oba způsoby uspokojivě splňovaly požadavek na přenos zápisu v reálném čase, nebyly vyhovující v případě, kdy měl být zápis sledován více než třemi studenty zároveň. Navíc vstupní procedura – spojení všech komponent systému – nebyla triviální a její úspěšnost byla do značné míry závislá na technických kompetencích uživatelů.

## 2. Souhrn požadavků na nové řešení

- **Co nejjednodušší a co nejméně časově náročná příprava technického prostředí.**

Smyslem tohoto požadavku je snížení závislosti systému na fyzické přítomnosti technika během přípravy prostředí ve výukových prostorách a snížení času potřebného k přípravě na minimum.

- **Co nejmenší závislost na technické infrastruktuře místa, kde zápis probíhá.**  
Protože nasazení systému se předpokládá v libovolném prostoru s různou a často předem neznámou vybaveností síťové konfigurace a možností, systém by měl být na nich závislý co nejméně.
- **Zobrazení zápisu na straně uživatele musí být v reálném čase synchronizováno s vytvářeným textem na straně zapisovatele.**  
Jde o rozhodující požadavek; aby uživatelé závislí na zápisu mluveného projevu měli s intaktními účastníky co nejsrovnatelnější podmínky, prodleva mezi mluvcím a zobrazením textového zápisu musí být minimální.
- **Základní parametry zobrazeného textu na straně uživatele by měly být nastavitelné.**  
Protože každý z uživatelů textového zápisu může mít odlišné individuální požadavky na zobrazení textu, systém by měl umožňovat nastavení stylu zobrazení textu (typ, velikost a barva písma, řádkování apod.) každému uživateli nezávisle na ostatních.
- **Možnost alespoň základní obousměrné komunikace (tzn. umožnit textový vstup ze strany uživatele).**  
Jako doplňující vlastnost, která by uživatelům textového zápisu umožňovala vznášet technické komentáře zapisovateli, nebo vkládat vlastní příspěvky do diskuze, která ve výuce/při jednání právě probíhá (u nich by zapisovatel figuroval jako prostředník mezi dalšími partnery diskuze).
- **Kompatibilita systému s českým a anglickým jazykem**  
Systém by měl být připraven korektně zpracovat text nejméně v českém a anglickém jazyce (diakritika, interpunkce atp.)

### 3. Analýza obdobných systémů

V analýze existujících systémů záměrně vynecháváme všechny, u nichž je převod mluveného slova do textu primárně zabezpečen nikoliv lidským zapisovatelem, ale jakoukoliv variací strojového rozpoznávání řeči do textu (*speech-to-text recognition*), i když nejen pro jazyk anglický, ale i pro český je k dispozici několik aplikací s poměrně zajímavými výsledky (český jazyk: *NewtonDictate*, *SpeechTech ASR – MegaWord*; anglický např. *Dragon Naturally Speaking* ad.). K vynechání těchto technologií nás vedou nejméně tyto důvody:

- Pro účely výuky na vysokých školách, při které mluvcí využívají specializovanou slovní zásobu mnoha oborů, je přesnost technologií automatického rozpoznávání řeči nedostačující.
- Dosažení přesnosti rozpoznávání je velmi závislé na technických podmínkách při snímání zvuku, jehož kvalitu nelze často ovlivnit.
- Přesnost rozpoznávání je sice možné zvýšit „trénováním“ systému na konkrétní mluvcí, to však není reálné v podmínkách běžné výuky zajistit.

Společným jmenovatelem dalších systémů, které pro převod mluvené řeči do textu primárně **nepoužívají strojové rozpoznávání** (počítají s lidským zapisovatelem) a které jsou analyzovány dále, je jejich závislost na internetovém připojení z výukových prostor (obecně z místa, kde zápis probíhá). Tento požadavek nezbytný pro jejich správné fungování považujeme za jejich základní nevýhodu, neboť jejich použití se pak stává kriticky závislé na technické infrastruktuře místa zápisu. Další nevýhody analyzovaných systémů z našeho pohledu uvádíme jednotlivě.

#### 3.1 eScribe ([www.escribe.cz](http://www.escribe.cz))

Nespornou organizační výhodou tohoto systému je dislokování zapisovatelů v prepisovacím centru, což jistě snižuje náklady a organizační aktivity, které by vedly k zajištění jejich přítomnosti v místě, kde se odehrává výuka/jednání a kde jsou zároveň přítomni uživatelé, kteří textový zápis sledují. Na druhou stranu tento systém **vyžaduje snímání a přenos hlasu** přednášejícího a případně obrazu s prezentačním materiálem mluvcího směrem do prepisovacího centra. To jednak **zvýšuje nároky na datovou propustnost** a stabilitu internetového připojení a jednak **zvýšuje náročnost přípravy technického prostředí** v místě, kde se výuka/jednání odehrává. Nepřímý kontakt zapisovatele s mluvcím, publikem a celkovým děním v místě považujeme za nespornou psychologickou i praktickou nevýhodu, která komplikuje řešení problematických situací. Nicméně ostatní naše požadavky, které jsme souhrnně definovali v úvodu, systém *eScribe* splňuje.

### 3.2 Google Documents (docs.google.com)

Jde o obecně dostupné prostředí pro sdílení dokumentů v reálném čase, jehož základní nevýhoda vyplývá právě z příliš obecného charakteru systému a jeho široké využitelnosti. Konkrétní nevýhody a komplikace použití systému *Google Docs* pro účely přenosu textového zápisu mluvené řeči lze shrnout v těchto bodech:

- Poměrně **zdlouhavý proces přípravy prostředí** – zapisovatel i uživatelé textového zápisu se musí do systému *Google Docs* přihlásit použitím svých osobních přístupových údajů a otevřít dokument s předem dohodnutým označením.
- V případě, že zapisovatel není přítomen v místě, kde se odehrává výuka/jednání (obdobně jako v případě systému *eScribe*), je **zapotřebí dalšími aplikacemi zajistit snímání a přenos hlasu přednášejícího**, příp. jeho prezentační obrazový materiál. To neúměrně zvyšuje náročnost na přípravu technického prostředí a z toho vyplývající větší závislosti takové přípravy na přítomnosti technika.
- **Styl zobrazení** sdíleného dokumentu s textovým zápisem lze na straně uživatele (nezávisle na ostatních) **přizpůsobit jen částečně** a netriviálním způsobem.
- V každém případě je využití *Google Docs* **závislé na spolehlivém internetovém připojení** z místa, kde probíhá výuka/jednání.

### 3.3 Bee Communications (bee-communications.com)

Komerční systém nabízený britskou stejnojmennou společností jako komplexní zapisovatelská služba – jejich zapisovatelé mohou být k dispozici vzdáleně i na místě, textový zápis je distribuován uživatelům prostřednictvím vlastního webového systému. Ze všech analyzovaných systémů splňuje definované nároky sice v největší míře, nicméně jeho hlavní nevýhody jsou zásadní:

- Náklady na adaptaci webového systému *Bee Communications* pro konkrétní prostředí Masarykovy univerzity by přesáhly výdaje potřebné pro vývoj systému vlastního.
- Odpovědnost za správu a údržbu adaptovaného systému by s velkou pravděpodobností zůstala ne zcela vyhraněná mezi Masarykovou univerzitou a zahraniční komerční společností.
- V případě, že by byl systém aplikován způsobem bez přítomnosti zapisovatele v místě výuky/jednání, bylo by zapotřebí zajistit snímání a přenos hlasu přednášejícího, což implikuje už dříve popsané nevýhody.
- V každém případě by bylo použití systému závislé na spolehlivém internetovém připojení z místa, kde probíhá výuka/jednání.

## 4. Vlastní systém Polygraf

Definované požadavky a analýza srovnatelných systémů vedly naše pracoviště k **vývoji vlastního systému (Polygraf)** a jeho nasazení v praxi. *Polygraf* splňuje všechny vymezené požadavky a je založen na technologiích, které jej umožňují použít nezávisle na technických možnostech místa, kde zápis probíhá. Ke svému chodu systém vyžaduje pouze elektrické napájení.

Klíčovými vlastnostmi systému jsou:

- Je složen pouze ze tří hardwarových součástí (viz níže), které jsou přenosné a které mohou být k činnosti připraveny během krátké chvíle.
- Ke své funkci nevyžaduje internetové připojení, předem zaručenou konfiguraci sítě, ani vlastní server.
- Spolupracuje s prostředím editoru *MS Word*, s nímž mnozí zapisovatelé rutinně pracují.
- Využívá zařízení *iPad* – tím je zajištěn základní uživatelský komfort: bezprostřední dotykové uživatelské rozhraní (včetně doprovodných informací ve znakovém jazyce), dlouhá výdrž v pracovním režimu bez nutnosti nabíjení baterie, snadná přenositelnost zařízení, rychlý a jednoduchý přechod do pracovního režimu.

- Pro zobrazování zápisu lze kromě nativního aplikačního prostředí (v systému *iOS*) použít i standardní internetové prohlížeče.
- Umožňuje obousměrnou komunikaci, tedy: zapisovatel > uživatel, uživatel > zapisovatel.
- Softwarové řešení umožňuje promítat zápis ve formě dynamického titulku i na počítači mluvčího, a může tak okamžitě doplňovat jeho případnou obrazovou prezentaci.

#### 4.1 Hardwarové prostředky systému

1. Osobní počítač (notebook) s operačním systémem *Microsoft Windows* a aplikací *Microsoft Word* (verze 2003 a novější) vybavený síťovou kartou pro připojení k bezdrátové síti Wi-Fi. Tento počítač je určen pro zapisovatele.
2. Libovolný přístroj s operačním systémem *iOS* (typicky *iPad*, *iPhone* nebo *iPod*) s možností připojení k bezdrátové síti Wi-Fi. Na tomto přístroji, který zde de facto plní funkci příručního displeje, sleduje zápis uživatel.
3. Vysílač uzavřené bezdrátové sítě Wi-Fi – ta je využita pro přenos textového zápisu mezi počítačem zapisovatele a příručními displeji uživatelů. Tato bezdrátová síť nemusí být dále připojena k internetu.



Obr. 1: Příklad základních hardwarových prostředků systému

#### 4.2 Softwarové prostředky systému

1. Část aplikace určená pro počítač zapisovatele je z technického hlediska doplněk (*add-in*) pro textový editor *Microsoft Word*. Prostřednictvím bezdrátové sítě zajišťuje vysílání textového zápisu pořizovaného v textovém editoru.
2. Část aplikace určená pro příruční displeje zajišťuje příjem textového zápisu a jeho průběžné zobrazování podle individuálního nastavení na straně uživatele (typ, velikost a barva písma, řádkování apod.).
3. Část aplikace (doplňující) určená pro prezentační počítač mluvčího průběžně zobrazuje textový zápis ve formě titulku pod prezentačním obrazovým materiálem.



Obr. 2: Základní schéma fungování systému Polygraf

## 5. Popis softwarových součástí systému

### 5.1 Část aplikace určená pro počítač zapisovatele (doplňk pro Microsoft Word)

Jako jediná nezbytná softwarová součást systému přebírá průběžně text z otevřeného dokumentu (příp. dokumentů) a nabízí ho ke sdílení ostatním součástem systému. Doplňk je spuštěn automaticky spolu s textovým editorem Microsoft Word a umožňuje uživateli (zapisovateli) pouze dvě volby:

1. Nastavení, které z otevřených dokumentů mají být určeny ke sdílení, příp. rychlé nastavení sdílení všech otevřených dokumentů.
2. Nastavení, zda má být přenášen text dokumentu s formátováním nebo bez něj (jako prostý text). Přenos textu bez formátování zajistí na straně notebooku zapisovatele rychlejší práci editoru MS Word, bez občasných přerušení práce editoru, během kterých je v opačném případě aktualizován formát textu v rámci celého dokumentu. Zamezit tyto prodlevy je vhodné zejména v případě rozsáhlejších dokumentů a v případě použití méně výkonného notebooku zapisovatele.

Aby počítačová příprava systému byla z uživatelského hlediska co nejjednodušší a zároveň co nejméně závislá na síťové infrastruktuře místa, kde zápis probíhá, byl pro účely zveřejnění existence sdílených dokumentů pro další hardwarová zařízení (typicky příruční displeje iPad) zvolen protokol MDNS (přesněji jeho varianta *Apple Bonjour*), který je použitelný v systémech *MS Windows*, *MacOS*, *iOS* a dalších. Ostatní obdobné protokoly (SSDS, LLMNR, SLP) jsme vyloučili buď z důvodu vysokých požadavků na používané prostředky systému nebo se jednalo o proprietární služby. Protokol MDNS nevyžaduje internetové připojení, předem známou konfiguraci vnitřní sítě, ani žádnou jinou službu poskytovanou jakýmkoliv serverem.

Jakmile jsou prostřednictvím protokolu MDNS zveřejněny sdílené dokumenty, jejich obsah ve formátu HTML je protokolem HTTP na zařízení uživatelů přenášena. Doplňk pro MS Word de facto pracuje jako samostatný webový server, který nabízí obsah dokumentů.

### 5.2 Část aplikace určená pro příruční displeje

Při volbě zařízení, které bude plnit funkci příručního displeje, na němž bude textový zápis průběžně zobrazován a sledován uživatelem, jsme měli na paměti požadavek na spolehlivý a lehký přístroj s dlouhou výdrží akumulátorů. V době, kdy jsme systém vyvíjeli, byly k dispozici v zásadě jen dvě zařízení – *Apple iPad*

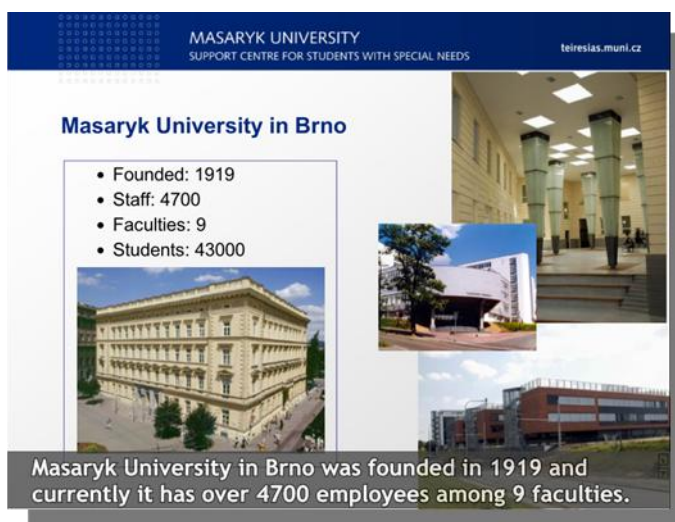
a *Amazon Kindle*. Protože Amazon Kindle neumožňuje vyvíjet samostatné aplikace pro svůj operační systém a jeho displej má pomalou odezvu, byl zvolen přístroj Apple iPad.

Ačkoliv je sdílený dokument počítačem zapisovatele nabízen protokolem HTTP ve formátu HTML, není v systému Polygraf textový zápis zobrazován ve standardním webovém prohlížeči přístroje iPad, ale ve vlastní aplikaci. K tomu nás vedly tyto důvody:

- Standardní webový prohlížeč přístroje iPad nespolupracuje s jinak výhodným protokolem Bonjour.
- Standardní prohlížeč sice umožňuje konfigurovat styl zobrazení dokumentu, ale z našeho hlediska méně efektivním způsobem.
- Pro plynulé zobrazení průběžně aktualizovaného textového zápisu by bylo zapotřebí doplnit standardní prohlížeč několika funkcemi, což obvykle nebývá spolehlivé.
- Uživatelské rozhraní vlastní aplikace lze lépe přizpůsobit dílčím požadavkům a doplnit o doprovodnou dokumentaci (náповěda atp.).

### 5.3 Část aplikace určená pro prezentační počítač mluvího

Tato doplňující součást systému je určena pro počítač, který zajišťuje zobrazení obrazového prezentačního materiálu mluvího (typicky snímky prezentace). Aplikace podobně jako příruční displeje Apple iPad přijímá textový zápis a transformuje ho na průběžně obměňovaný titulek zobrazovaný v plovoucím okně, které je připojeno k prezentačnímu materiálu mluvího. Výhodou je zobrazení textového zápisu a prezentačního obrazového materiálu mluvího současně na jedné obrazovce/plátně. Lze nastavit počet řádků (1–4), font a velikost písma a barvu pozadí a textu. Tato část aplikace je určena pro operační systém MS Windows XP a vyšší.



Obr 3.: Způsob doplňkového zobrazení textového zápisu jako titulku

## 6. Dosavadní praktické zkušenosti

Systém Polygraf byl poprvé pravidelně používán na Masarykově univerzitě během podzimního semestru 2010 v celkovém rozsahu přibližně 110 přednášek, každé z nich bylo současně přítomno 3–6 studentů sledujících textový zápis. Po celou dobu semestru byl během přednášek přítomen zástupce technického dozoru, ale po velmi krátké době se ukázalo, že jeho přítomnost není dále nezbytná.

Systém byl také bez větších komplikací využit pro zobrazení textového zápisu prezentací na mezinárodní konferenci Universal Learning Design 2011, kterou pořádala Masarykova univerzita ve dnech 8 – 11. února 2011 v Brně.

## 7. Další vývoj

Ačkoliv jsou v systému Polygraf řešeny dílčí drobné problémy a aplikovány drobné adaptace na základě zkušeností uživatelů, jsou plánovány následující úpravy a rozšíření většího charakteru:

- vedle textového zápisu možnost zobrazení obrazového prezentačního materiálu mluvího na příručním displeji (především pro uživatele se zrakovým postižením, kteří nemohou sledovat obraz na hlavním plátně)
- rozšíření aplikace pro další operační systémy (Android, Windows Mobile ad.)

### Souhrnný seznam funkcí systému Polygraf

- **Vizualizace mluvené řeči prostřednictvím synchronního zápisu** (speech-to-text reporting) pro větší skupinu uživatelů
- **Možnost přenosu textového zápisu bez původního formátování nebo s formátováním**

Funkci lze nastavit na straně zapisovatele.

- **Volitelný styl zobrazení textového zápisu na uživatelských zařízeních**

Pro větší uživatelský komfort může na příručním displeji uživatel nastavit individuální parametry zobrazení (typ, velikost a barva písma, řádkování apod.).

- **Komunikační rozhraní pro přenos textových zpráv mezi uživateli zápisu a zapisovatelem**  
Z příručních displejů lze na zapisovatelský počítač odesílat textové zprávy – ať už jsou to technické komentáře, příp. upřesňující dotazy k obsahu (jako adresát sdělení je zamýšlen pouze zapisovatel), nebo uživatelovy vlastní příspěvky do diskuze, která ve výuce/při jednání právě probíhá (u nich zapisovatel figuruje jako prostředník mezi dalšími partnery diskuze).

- **Zobrazení textového zápisu ve formě titulku**

Zápis je transformován na průběžně obměňovaný titulek zobrazovaný v plovoucím okně, které může být připojeno k prezentačnímu obrazovému materiálu mluvího. Výhodou je zobrazení textového zápisu a prezentačního obrazového materiálu mluvího současně na jedné obrazovce/plátně.

## Literatura

*eScribe – on-line centrum přepisu neslyšícím* [online]. R&D Centre ČVUT v Praze, (c) 2009 [cit. 2011-02-07]. Dostupné z <<http://www.escribe.cz/>>

*Bee Communications* [online]. Bee Communications, London, (c) 2011 [cit. 2011-02-07]. Dostupné z <<http://bee-communications.com/>>