

eScribe: ONLINE PŘEPISOVACÍ CENTRUM

Zdeněk BUMBÁLEK, Jan ZELENKA

Research and Development Centre, Fakulta elektrotechnická

České vysoké učení technické v Praze

Technická 2, 166 27, Praha 6

Email: {bumbazde, zelenj2}@fel.cvut.cz

Anotace: Komunikační a hlasové technologie mají potenciál snížit komunikační, vzdělávací a sociální bariéry osob s postižením sluchu. Cílem projektu eScribe je omezení těchto bariér a pomocí zapojení moderních komunikačních technologií (mobilní sítě, Internet, internetová telefonie, automatické rozpoznávání řeči) přispět ke všeobecné dostupnosti přepisu mluvené řeči.

1. Úvod

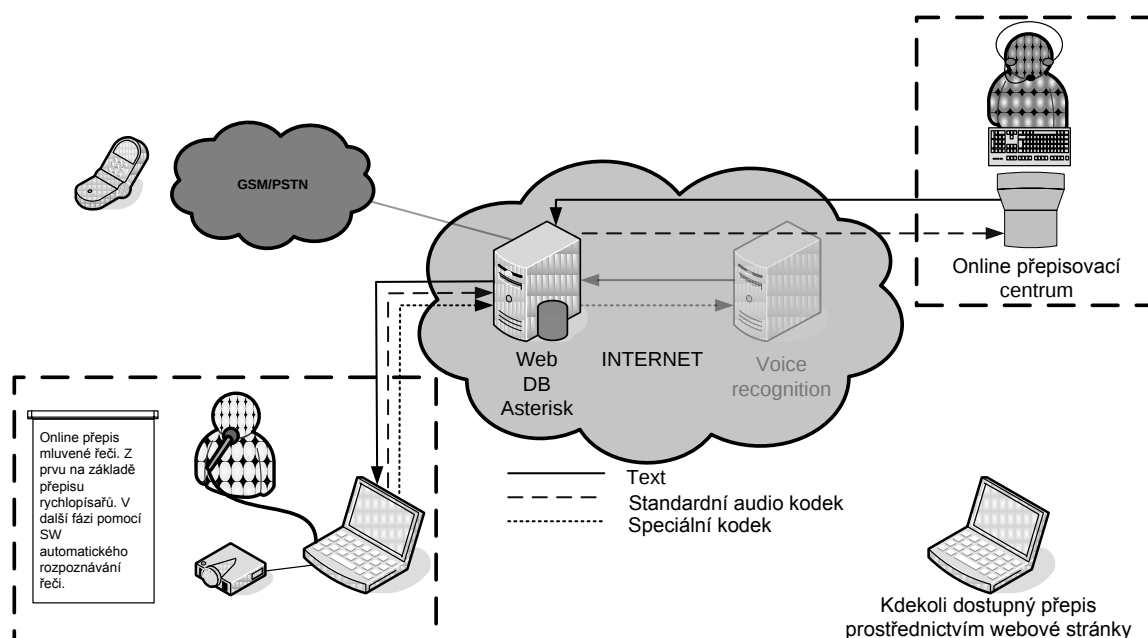
Schopnost porozumět mluvené řeči hraje základní úlohu v komunikaci mezi lidmi, ve vzdělávání, sdílení myšlenek a udržování mezilidských vztahů. Lidé s postižením sluchu tak ve většinové „slyšící“ společnosti musí čelit komunikačním bariérám. Tyto bariéry jsou o to více závažné při komunikaci s úřady nebo soudních jednáních, kde porozumět správně každému slovu je obzvláště důležité. Podle zákona [1] o komunikačních systémech neslyšících a hluchoslepých osob, mají sice tyto osoby právo svobodně si zvolit takový komunikační systém, který nejlépe odpovídá jejich potřebám, ve skutečnosti však dosud nebyly vytvořeny podmínky pro naplňování tohoto práva, zejména pokud jde o zajištění přepisu mluvené řeči. Jeden z cílů projektu eScribe, společné aktivity Českého vysokého učení technického v Praze (ČVUT) a České unie neslyšících (ČUN), podpořené Nadací Vodafone ČR a částečně Magistrátem hl. m. Prahy, je tento stav napravit. Řešením je systém okamžitého, snadno dostupného a všudypřítomného přepisu mluvené řeči. Využit k tomu lze služeb vyškolených rychle píšících přepisovatelů. Koncept fyzicky přítomných přepisovatelů byl v ČR použit v pilotním projektu ČUN: „Simultánní přepis mluvené řeči“ [2]. Avšak nedostatek přepisovatelů a náklady na jejich služby spojené s cestováním limitují počty uživatelů, kteří by chtěli takovou službu využít. Alternativou by se v budoucnu mohly stát systémy automatického rozpoznávání řeči (Automated Speech Recognition - ASR). Bohužel v současnosti jsou tyto systémy limitovány svou přesností, zejména pokud se jedná o hovorovou řeč a složitost českého jazyka.

Hlavní přínos systému eScribe spočívá v navržení prototypu pro všudypřítomné, ekonomicky efektivní přepisovací centrum pracující v reálném čase, které bude zprvu založené na službách rychle píšících přepisovatelů a později na kombinaci systému automatického rozpoznávání řeči ve spolupráci s lidským korektorem. Cílem projektu je vytvořit online přepisovací centrum pro neslyšící.

2. Princip a funkce

Technické řešení projektu eScribe je založeno na IP telefonii a online zobrazení přepisu řeči na webových stránkách. Přístup do systému je možný jednak z klasických telefonních přístrojů, mobilních telefonů, ale i z HW SIP telefonů a SW SIP klientů. Nejsnadnější způsob pak tvoří webový telefon, díky kterému je celý systém dostupný pouze z webového rozhraní bez nutnosti jakékoli instalace a konfigurace pro uživatele.

Princip systému eScribe je znázorněn na obr. 1. Z místa konání přednášky pro neslyšící je přenášen hovor pomocí VoIP telefonie do přepisovacího centra nebo kamkoli k přepisovateli. Přepis zajišťují speciálně vyškolení rychlopisáři, kteří používají velký seznam zkratk. Ten se expanduje na celá slova nebo věty. K tomu využívají MS Word s velkým seznamem automatického vkládání. Použití MS Word bylo jednou ze základních podmínek přepisovatelů na vývoj aplikace pro přepis. V praxi probíhá přepis tak, že si přepisovatel pomocí webového rozhraní aplikace vytvoří a otevře dokument MS Word. Na jeho pozadí běží programový kód, který s malým zpožděním, prakticky v reálném čase, odesílá text na server, odkud je dále zobrazován na webovou stránku. Na místě přednášky je k dispozici projektor a přepisovaný hlas je zobrazován na plátno. Dostupný je však odkudkoli pomocí webového prohlížeče. Do budoucna projekt zamýšlí postupné zapojování automatického SW na rozpoznávání řeči. Tím se zpřístupní služba ještě širšímu počtu neslyšících [3].



Obr.1: Princip systému eScribe

V současnosti implementovaný prototyp online přepisovacího centra je připraven k ostrému testování, které bude spuštěno během měsíce března na přednáškách FEL ČVUT, kterých se účastní neslyšící studenti.

3. Architektura systému

Na obr. 2 je naznačena architektura systému eScribe. Základem komunikačního systému projektu eScribe je SW ústředna Asterisk [4]. Systém Asterisk provozujeme na platformě Linux OpenSUSE 11.1 a plnohodnotně plní veškeré funkce pobočkové ústředny PBX. Dalším velmi důležitým blokem je server zajišťující online zobrazování přepsané řeči na webových stránkách, který úzce spolupracuje s webovým serverem Apache.

3.1 SW ústředna Asterisk

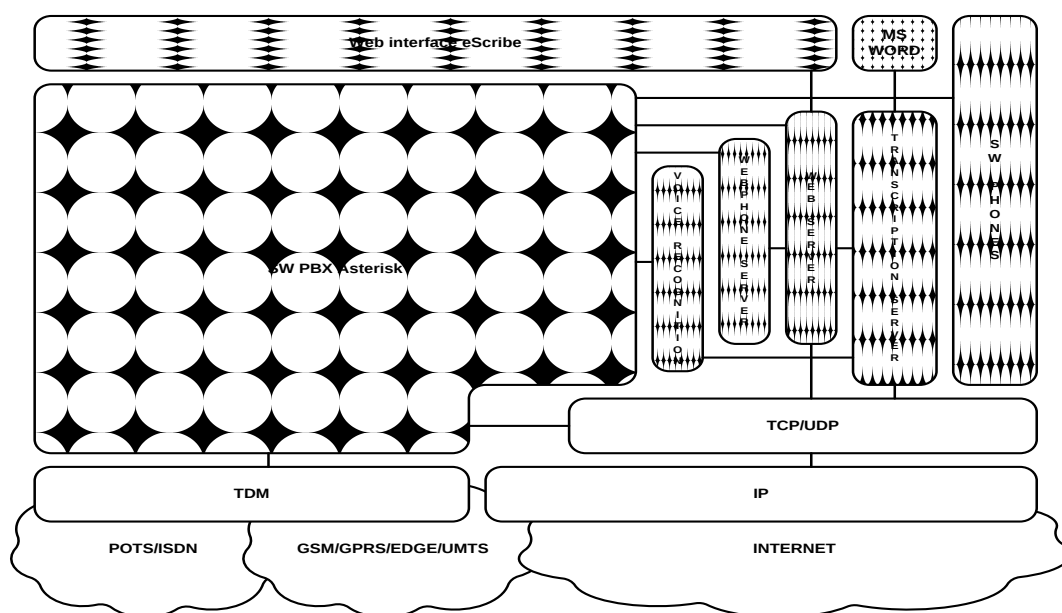
Asterisk je open source SW určený k provozování telefonních služeb jak na úrovni přepínání okruhů (TDM), tak v síti s přepínáním paketů. Systém Asterisk lze právem považovat za velmi rozšířené, flexibilní a silné řešení v oblasti telekomunikačního SW.

Asterisk je schopný zpracovávat různé druhy protokolů jak ve VoIP telefonii (SIP, MGCP, H.323 a IAX), tak v pevných i mobilních telefonních sítích. Velká výhoda systému spočívá právě v jeho otevřenosti a možnosti přizpůsobení nejrůznějším standardům. Otevřenost systému je velmi důležitým požadavkem projektu eScribe, zejména z důvodu pozdějšího snadného propojení se SW pro automatické rozpoznávání hlasu.

V rámci eScribe jsou využívány protokoly pro internetovou telefonii SIP (Session Initiation Protocol) a IAX (Inter-Asterisk eXchange). Pomocí těchto protokolů je možné uskutečňovat hovory zdarma odkudkoli s přístupem na Internet. Pro komunikaci se sítí PSTN využívá projekt eScribe propojení pomocí soukromého VoIP operátora, do mobilní sítě GSM je systém připojen pomocí GSM bran.

3.2 Online zobrazování přepisu

Dalším velmi důležitým blokem systému je server a aplikace pro přepis. Zatímco u komunikačního řešení byl použit otevřený a často používaný systém Asterisk, u aplikace pro přepis bylo zapotřebí vyvinout vlastní řešení, které by splňovalo požadavky přepisovatelů na minimální změnu v jejich dosavadní práci. Pro přepis využívají přepisovatelé standardně MS Word. Vstupem dat pro online zobrazování textu na webových stránkách je tak dokument MS Word, na jehož pozadí běží programový kód, který zajišťuje odesílání textu na server. Projekci přepisu na webových stránkách zajišťují skripty v PHP, které generují automaticky se obnovující dokument, který periodicky přenáší celý text přepisu.



Obr.2: Architektura systému eScribe

Cílem online přepisu je poskytnout neslyšícím komunikaci v reálném čase, tzn. přenos a zobrazování textové informace znak po znaku s možností oprav a využití kláves Delete a Backspace. Za takovýchto podmínek není možné využívat standardních instant messengerů. Vyvinutá aplikace pro přepis tyto podmínky částečně splňuje, umožňuje především zpětné editování textu, ovšem přenos znak po znaku je díky nativním vlastnostem MS Word nemožný, neboť umožňuje odesílat událost max. jedenkrát za vteřinu. Další zpoždění nastává na straně zobrazení dat ze serveru na webovou stránku. Celkové zpoždění činí okolo 1,5 vteřiny a text je přenášán v dávkách odpovídajících maximálnímu množství textu, které stačí přepisovatel za tuto dobu napsat. V současné době pracujeme v rámci eScribe také na aplikaci pro přepisovatele, která by byla dostupná přímo na webových stránkách a která by simulovala prostředí a chování textového editoru MS Word.

Možným řešením přenášení textu znak po znaku je využití RTP protokolu, který se běžně používá pro přenosy zvuku a videa v reálném čase (real time). Ideálním řešením pro uživatele je integrace přenosu real time textu [5] s přenosem zvuku a obrazu do jednoho zařízení (klienta). Takový způsob komunikace se nazývá Total communication a je popsán v doporučení ITU-T F.703. Zavedením bezchybného real time přenosu textu znak po znaku přes IP se zabývá standard T.140 a IETF RFC 4103. Takovýto způsob přenosu textu se v literatuře označuje jako Text over IP (ToIP) [6] a je doplňkem k VoIP a Video over IP. Text over IP najde uplatnění kromě komunikace s lidmi se sluchovým postižením také při komunikaci ve velmi hlučných prostorách, diskrétní komunikaci, přenosu přesných informací a jako doplněk k hlasové a video komunikaci.

Projekt eScribe se v současné době zabývá implementací ToIP do systému Asterisk s podporou rychlosti psaní cca 30 znaků/sec.

4. Další pokračování projektu

Komunikační a hlasové služby nacházejí čím dál častěji široké uplatnění i v oblasti asistivních technologií. Asistivní hlasové služby jsou založeny na ASR (Automated Speech Recognition) a TTS (Text to Speech) technologiích a popsány jsou v řadě studií [7] [8] [9]. Pro rozpoznávání anglického jazyka existuje v současnosti již i řada komerčních řešení (např. Microsoft, IBM). Vedoucí postavení ve využití ASR v asistivních technologiích zaujímá LLP (Liberated Learning Project) [10] ve spolupráci s IBM Human Ability and Accessibility Center [11]. Rozpoznávání národních jazyků jako je čeština zůstává mimo zájem velkých společností a rozvíjí se především na univerzitní půdě (v ČR: TU Liberec – prof. Nouza a kol., ZCU – doc. Müller a kol., ČVUT – Speech Processing Group). Hlavním cílem pro následující období je propojení projektu eScribe se systémem rozpoznávání řeči ASR a automatizovat tak částečně proces přepisu. Pro rozpoznávání češtiny bude použit systém ASR společnosti Newton Technologies (systém vyvinutý na TU Liberec, prof. Nouza a kol.), která se stala partnerem projektu eScribe. Současné ASR systémy jsou navrženy především pro diktovací účely. Při rozpoznávání běžného hovorového jazyka však zatím nedokážou pracovat s dostatečnou přesností. Úkolem projektu je proto vývoj korektorských nástrojů, pomocí kterých by korektor dokázal snadno a rychle editovat výstup systému ASR [12]. Kombinací systému automatického rozpoznávání řeči a dodatečné editace

bude moci přepis dostatečně rychle a kvalitně zajišťovat i méně zkušený a pomalejší přepisovatel. Zajímavou alternativou je také využití tzv. „stínového mluvčího“, na jehož hlasový profil se systém ASR natrénuje. Stínový mluvčí řeč přemlouvá podle zásad diktování, čímž se výsledná přesnost rozpoznávání značně zvyšuje.

V další etapě zamýšlíme zohlednit v architektuře projektu model „Cloud Computing“ [13] a poskytovat služby založené na využití kontextových informací. Zatímco vstupní kontext obsahuje informace např. o tel. čísle volajícího a jeho polohu, výstupní kontext obsahuje informace, které přinesou uživateli užitek. Takto navržená architektura umožní propojení současných, ale i budoucích komunikačních a hlasových technologií a bude tak položen základ pro vývoj nových hlasových aplikací a jejich přizpůsobení potřebám handicapovaných.

5. Závěr

Široce dostupné přepisovací centrum umožní poskytovat službu přepisu mnohem většímu okruhu osob s postižením sluchu. Prostřednictvím přepisu na dálku s využitím systému ASR a korektorského pracoviště bude možné zajišťovat službu přepisu levněji než při osobní přítomnosti přepisovatele na místě konání akce a bude možné také podstatně zvětšit počet „přepisovaných“ akcí.

Díky online přepisu se neslyšícím zpřístupní kulturní, vzdělávací, společenské či jiné akce, kterých by se kvůli komunikační bariéře nemohli zúčastnit. Přepis lze využít také při individuálních jednáních neslyšících osob např. u soudu nebo na úřadě, kde je neschopnost domluvit se pro neslyšící osoby zvlášť tíživým problémem.

Námi navržená otevřená architektura asistivních hlasových služeb umožní poskytování těchto služeb nezávisle na místě, času a i typu uživatelského terminálu, protože službu je možné využívat na jakémkoliv zařízení, které umožňuje prohlížet webové stránky. Svou modulárností umožní zohlednit jednotlivé potřeby nejen neslyšících uživatelů.

Literatura

- [1] Zákon č. 155/1998 Sb., o komunikačních systémech neslyšících a hluchoslepých osob, ve znění zákona č. 384/2008 Sb <http://www.escribe.cz>
- [2] <http://www.prepis.cz>
- [3] Bumbálek, Z.: Využití IP telefonie v asistivních technologiích pro neslyšící. Access server [online]. 2010, Internet: <http://access.feld.cvut.cz/view.php?nazevclanku=vuziti-ip-telefonie-v-asistivnich-technologiich-pro-neslyšici&cislocclanku=2010020003>
- [4] Meggelen, J. – Madsen, L. – Smith, J.: Asterisk™: The Future of Telephony. 2nd Edition. Sebastopol: O'Reilly Media, 2007
- [5] <http://www.realtimetext.org>
- [6] Hellström, G.: IETF RFC 4103: RTP Payload for Text Conversation, the Internet Society, 2005
- [7] Miyoshi, S., Kuroki, H., Kawano, S., Shirasawa, M., Ishihara, Y.: Support Technique for Real-Time Captionist to Use Speech Recognition Software. In: ICCHP 2008, LNCS 5105, pp. 647–650, 2008
- [8] Forman, I., Brunet, T., Luther, P., Wilson, A.: Using ASR for Transcription of Teleconferences in IM Systems. Universal Access in HCI, Part III, HCII 2009, LNCS 5616, pp. 521–529, 2009
- [9] Kheir, R. and Way, T.: Inclusion of Deaf Students in Computer Science Classes Using Real-Time Speech Transcription. In Proceedings of the 12th Annual SIGCSE Conference on Innovation & Technology in computer Science Education (Dundee, Scotland, June 25-27, 2007). ITiCSE'07. ACM, New York, 192-196.
- [10] The Liberated Learning Consortium, <http://www.liberatedlearning.com/>
- [11] IBM ViaScribe, http://www-03.ibm.com/able/accessibility_services/ViaScribe-accessible.pdf
- [12] Wald, M.: Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time. In: ICCHP 2006, LNCS 4061, pp. 683 – 690, 2006.
- [13] Miller, M.: Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, Quo, 2009