Univerzita Karlova v Praze
Filozofická fakulta
Ústav informačních studií a knihovnictví

Studijní program: informační studia a knihovnictví
Studijní obor: informační studia a knihovnictví

Rostislav Koňařík

Počítačové zpracování textu a řeči a jeho využití v informačním prostředí

Text and speech computer processing and its use in information environment

Bakalářská práce

Praha 2009-08-10

Vedoucí bakalářské práce: Doc. RNDr. Jiří Souček, DrSc.

Oponent bakalářské práce:

Datum obhajoby:

Hodnocení:

Prohlášení:

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje.

V Praze, 10. srpna 2009

……..

podpis studenta

Identifikační záznam

KOŇAŘÍK, Rostislav. *Počítačové zpracování textu a řeči a jeho využití v informačním prostředí [Text and speech computer processing and its use in information environment].* Praha 2009-08-10. 47 s. Bakalářská práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Vedoucí bakalářské práce Doc. RNDr. Jiří Souček, DrSc.

Abstrakt

Bakalářská práce se zabývá popisem technických principů převádění textu do elektronické podoby pomocí skenování a následného rozpoznávání znaků pomocí technologie OCR. Dále seznamuje s principy analýzy, syntézy a rozpoznávání řeči za pomoci počítače a zaměřuje se také na hardware a software sloužící k těmto účelům. Opomenuta nezůstává ani problematika úspěšnosti a spolehlivosti těchto metod vzhledem ke zdrojovému materiálu a současným technologickým limitům. Dále práce obsahuje příklady použití těchto technologií v praxi, zmapování jejich významu a možnosti využití v informační společnosti. [Autorský abstrakt].

Abstract

The bachelor thesis deals with a description of technical principles transferring a text into an electronic form by scanning and consecutive recognizing of symbols by OCR technology. Further it introduces principles of analysis, synthesis and recognition of speech by a computer and is also concentrated on hardware and software subserving these purposes. What is also not forgotten is the problem of fruitfulness and dependability of these methods with regard to a resource material and nowadays limits of technology. In the following the bachelor thesis includes some examples of work experiences of use these technologies, exploration of their meaning and a possibility of use these technologies in information society.

Klíčová slova

analýza řeči, digitalizace, informační prostředí, OCR, počítačová lingvistika, počítačové zpracování, převod textu na řeč, rozpoznávání řeči, skenování, syntéza řeči, umělá inteligence

Keywords

speech analysis, digitizing, information environment, OCR, computational linguistics, computer processing, text-to-speech, speech recognition, scanning, speech synthesis, artificial intelligence

OBSAH

PŘEDMLUVA

Tato bakalářská práce se zabývá zpracováním textu a řeči za použití výpočetní techniky. Jejím cílem je nastínit technické a technologické principy fungování těchto procesů, jejich současné nedostatky a limity. V neposlední řadě bude také uvedeno praktické využití těchto technik a jejich uplatnění v informační společnosti.

Výběr tématu jsem zvolil vzhledem k mému zájmu o zvukovou a výpočetní techniku. Dále jsem se chtěl dozvědět více o principech, jak fungují tyto technologie z oblasti umělé inteligence.

Tato práce si neklade za cíl postihnout probíranou problematiku v celé její šíři, vhledem k odbornosti jejího obsahu, který zasahuje do mnoha oborů jako je například umělá inteligence, počítačová lingvistika, akustika a dalších. Aktuálnost dostupných informací k této problematice je navíc ohrožována jejich rychlým zastaráváním díky neustálému výzkumu a rozvoji informačních a komunikačních technologií.

Podstatou této bakalářské práce je tedy přiblížit dnešní možnosti a základní princip fungování těchto procesů a nastínit možnosti jejich praktického využití v informační společnosti.

Text je rozdělen do tří hlavních kapitol a dílčích podkapitol. První část je orientována na digitalizaci textových dokumentů a jejich následnou analýzu pomocí rozpoznávací technologie OCR. Druhá část se věnuje řečovým technologiím, mezi které patří syntéza a rozpoznávání řeči. Podrobněji je zpracována část o řečové syntéze metodou text-to-speech. Opomenuty nebudou ani aspekty hodnocení kvality syntetizované řeči, problémy a limity při její syntéze. Poslední kapitola je zaměřena na aplikaci a využití výše zmíněných technologických principů v praxi.

Bakalářská práce má celkem 47 stran. Je v ní použito 30 zdrojů, které jsou řádně odcitovány podle normy ISO 690 a ISO 690-2.

- ISO 690:1987. *Documentation – Bibliographic references – Content, form and structure.* 2nd ed. Geneva : ISO, 1987. 11 s. Výtah textu normy dostupný také z WWW: <http://www.collectionscanada.gc.ca/iso/tc46sc9/standard/690-1e.htm>.

- ISO 690-2:1997. *Information and documentation – Bibliographic references – Part 2: Electronic documents or parts thereof.* 1st ed. Geneva : ISO, 1997. 18s. Výtah textu normy dostupný také z WWW: <http://www.collectionscanada.gc.ca/iso/tc46sc9/standard/690-2e.htm>.

Na závěr bych chtěl poděkovat Doc. RNDr. Jiřímu Součkovi, DrSc. za jeho pomoc při psaní této práce a v neposlední řadě také své rodině a svým blízkým za podporu při studiu.

1. Počítačové zpracování textu

V následující kapitole se zabývám fungováním a použití počítače jako přístroje užívaného k rozpoznávání textu a další práce s ním. Dále se budu zabývat převáděním textových dokumentů do digitální podoby pomocí skenování. Vysvětlím co je to OCR, k čemu slouží a jak funguje tato technologie. Uvedu potřebný hardware a stručný přehled softwaru pro skenování a rozpoznávání pomocí OCR. Sémantickou povahu textu však ponechám stranou, vzhledem k tomu, že se jedná o další samostatnou vědní disciplínu.

S rozvojem moderních technologií posledních desetiletí 20. století došlo i k modernizaci přístupu k psanému a tištěnému textu. Psaný text byl postupně nahrazován textem elektronickým, který má oproti textu “na papíře” řadu nesporných výhod. Mezi jeho zásadní přednosti lze zařadit jeho velmi jednoduchou upravitelnost, snadný převod do papírové podoby, schopnost provazovat text navzájem pomocí tzv. hypertextových odkazů, interaktivitu textu, možnost elektronický text přečíst prostřednictvím vhodného software bez přítomnosti fyzického mluvčího apod. Moderní elektronickou komunikaci, posílání e-mailů, SMS, tvorbu elektronických knih (e-books), tvorbu a administraci internetových portálů, vývoj software i hardware si dnes již bez elektronického textu nelze představit. Přesto má oproti svému papírovému předchůdci jednu nespornou nevýhodu. Existuje tu pouze několik desítek let. (Kopecký; Nocar; Kopecký, 2003)

1.1. Digitalizace

S pojmem digitalizace se dnes setkáváme téměř na každém kroku. Termín digitalizace lze definovat jako technologii reformátování, která spočívá v převedení dokumentu do elektronické (digitální) podoby. Součástí digitalizace bývá často i tvorba metadat, které jsou ukládány společně s digitalizovanými dokumenty. (Digitalizace dokumentů, 2005)

Po konverzi "analogového" média do digitální podoby, jsou dokumenty reprezentovány formou dat uložených v binární (dvojkové) soustavě, ve které pracují osobní počítače.

Do digitální podoby lze převést téměř všechny typy dokumentů - obrazové, zvukové i audiovizuální. Mezi hlavní důvody, proč digitalizovat dokumenty, patří především úspora místa a vytvoření záznamu, který se dá dále efektivněji upravovat, distribuovat a archivovat. K dalším výhodám digitalizace různých typů dokumentů patří snížení prostorových nároků při archivaci a skladování dokumentu, další cesta jak zpřístupnit dokument uživateli např. online, odstranění prostorových bariér, rychlé prohledávání digitálního archivu, možnost snadného zálohování a v neposlední řadě zpřístupňování elektronického obsahu prodejem v elektronických obchodech. Nevýhodu digitálních dokumentů spatřuji zejména ve větší náchylnosti k

nelegálnímu pořizování jejich kopií, neautorizovanému šíření, úpravě obsahu či jiného porušení autorských práv. (Habiballa, 2004)

Ke správě, sdílení a archivaci digitální dokumentů se používají softwarové systémy typu DMS (Document Management Systems).

Vzhledem k obsahu této práce se budu dále věnovat, především digitalizaci textových dokumentů.

1.1.1. Digitalizace textu pomocí skenování

Jednou z metod digitalizace, která umožňuje převést dokument do elektronické podoby je skenování. Skenováním rozumíme proces snímání obrazové nebo textové předlohy pomocí skeneru, jímž se převádí optoelektronický signál do digitální podoby. Digitalizovaný obraz se ukládá v některém z grafických formátů. Hustota sítě snímání, tzv. rozlišení, určuje kvalitu získaného obrazu. (Skenování, 2005)

K digitalizaci či převodu textového dokumentu do el. podoby potřebujeme tedy kromě počítače ještě skener a software, který bude schopen se skenerem komunikovat.

Před vlastním zpracováním je vhodné připravit dokumenty tak, aby digitalizace jako celek probíhala co nejefektivněji tedy bez zbytečných časových ztrát či jiných komplikací. Proto je vhodné mít dokumenty ke skenování náležitě připraveny. Tím je myšleno uvést dokument do co nejlepšího stavu po fyzické stránce, např.: narovnání zohýbaných stran, odstranění sponek, svorek či jiných artefaktů, v případě nutnosti také opravit možná poškození, kterými mohou být např. potrhané listy. (Jančařík, 1980)

1.2. Hardware potřebný pro skenování

V této části se zaměříme na dnes nejrozšířenější skenery - skenery plošné. Uvedené zákonitosti však platí obecně pro jakékoliv snímání tištěných předloh.

Dokument je nutné ve skeneru nejdříve nasnímat (neboli zaznamenat, zachytit). Základním požadavkem je dobré a rovnoměrné osvícení předlohy po celé její ploše. To zajišťovala u plošných skenerů donedávna tzv. "Chladná katodová lampa", neboli zářivka. Výhodou tohoto řešení je vysoká intenzita produkovaného světla, nevýhodou je pak nerovnoměrné osvětlení (nejvíce světla je vyzařováno uprostřed). Aby byl tento nedostatek v co možná největší míře odstraněn, je zářivka obvykle doplněna systémem zrcadel, které vrací odražené světlo na místo, kde je ho potřeba. Novější řešení u tzv. CIS technologie využívá řadu luminiscenčních LED diod. Všechny použité diody jsou přirozeně stejné a to zaručuje maximální možnou stejnoměrnost osvětlení po celé šíři snímaného dokumentu. Osvětlovací a snímací mechanizmus se postupně posouvá po předloze a snímá jeden řádek za druhým. (Knotek, 1999)

Kombinace zářivka - optická soustava - snímací prvek CCD (elektronicky řízený senzor) je klasická technologie, nazývaná CCD (Charge-Coupled Device). Skenery vybavené tímto způsobem snímání jsou trochu dražší, choulostivější na poškození, ale mají lepší barevnou citlivost.

V poslední době přibývá na trhu množství skenerů s technologií osvitu CIS. Nejedná se o nic jiného, než o dvě řady diod, jednu vysoce svítivých LED diod a řadu diod snímacích. Kladem jsou nižší výrobní náklady a tudíž nižší cena CIS skenerů, menší rozměry a větší odolnost. Nevýhodou je naopak nižší svítivost a citlivost (to se projevuje například při snímání jemných barevných odstínů nebo třeba u silnější rozevřené knihy ve hřbetu).

Nyní se podívejme na proces převodu informace z dokumentu do počítače. Předloha je tedy patřičně nasvícena. Nyní vstupují do hry tzv. snímače (CCD nebo CIS). Snímač pracuje tak, že intenzita světla, které dopadá na jeho jednotlivé buňky je přeměněna na elektrický náboj o různé síle. Každý bod elektronické podoby obrazu je složen ze tří informací - intenzity tří základních barev - R (červená), G (zelená) a B (modrá). Každý bod snímané předlohy je tedy měřen třemi buňkami snímače - každá buňka pomocí speciálních filtrů vyhodnocuje jednu z uvedených

barevných složek bodu. V plošných skenerech jsou použity tzv. řádkové CCD nebo CIS snímače, použitý snímač tedy určuje maximální možné optické rozlišení skeneru. (Knotek, 1999)

Z výše uvedeného vyplývá, že kvalita skeneru je přímo závislá na kvalitě použitého snímače a počtu jeho buněk. V současné době většina plošných skenerů používá snímače s rozlišením 1200 nebo 2400 dpi. Označení dpi (dot per inch) udává, kolik bodů je snímač schopen změřit na vzdálenosti jednoho palce (asi 2,5 cm). CCD snímač s rozlišením 1200 dpi má tedy 3600 buněk (každý bod je snímán třikrát) na každých přibližně 2,5 cm. Plošný skener určený pro formáty A4 má přibližně 15 000 buněk. Skenery s udávaným rozlišením 1200 dpi mívají někdy snímací prvek s rozlišením 600 dpi. Pohybující se snímací mechanismus je schopen na dráze dlouhé jeden palec změřit 1200 řádek předlohy, takže výsledné optické rozlišení elektronické podoby obrázku z takového skeneru je oněch 600 x 1200 dpi. Obdobně skenery označené rozlišením 600 dpi mají někdy snímač s rozlišením 300 dpi, který snímá předlohu v 600 krocích (řádkách) na palec. (Bennex, 2007)

Většina prodávaných skenerů umí dále softwarově upravit počet bodů na mnohem vyšší hodnotu, přičemž každý původně vyhodnocený bod rozdělí na několik dalších bodů a na kvalitě programového vybavení potom záleží, jak dobře si

skener poradí s barevnými odstíny přidělenými novým bodům. Tento proces zvládá však i naprostá většina dobrých programů pro úpravu obrázků. Kvalitu výstupu ze skeneru primárně a zásadně určuje jeho optické rozlišení a tím i ostrost výsledného elektronického obrazu. Kromě počtu buněk na snímači jsem se také zmínil o jeho kvalitě. Ta je dána tím, jak věrně je schopen tento převodník obrazové informace na elektronickou reprodukovat barvy.

Další z vlastností, ze které je částečně patrná výsledná kvalita barevného podání, je barevná hloubka. Ta udává, kolik možných hodnot může mít elektrický náboj produkovaný jednotlivými buňkami snímače. Je udávána v "bitech" a větší číslo udává větší počet barev, který je schopen skener rozeznat. V praxi se u barevných skenerů setkáme s hodnotami 24 až 48 bitů. Dalšími činiteli, které ovlivňují barevné podání obrazy při jeho dalším zpracování, je skutečná kvalita snímače, kterou žádný výrobce neuvádí, věrnost barevného podání monitoru a rovněž kvalita tiskárny, na které případně upravený dokument uživatel tiskne. (Knotek, 1999)

1.2.1. Typy skenerů

Při rozhodování o výběru skeneru je nutné brát v patrnost především to, co bude od tohoto zařízení uživatel očekávat a k jakému účelu jej bude převážně používat. Dalším hlediskem bude velmi pravděpodobně jeho cena a možnost technické podpory a servisu. Naprosté většině domácích i kancelářských uživatelů bude pro zpracování grafických i textových předloh dostačovat rozlišení 1200 x 2400 dpi, které dnes nabízí většina modelů. Standard se posouvá stále vzhůru. Při zvažování budoucího využití je též nutné brát v úvahu skutečnost, že dokumenty s vyšším rozlišením zaberou mnohem více místa v paměti počítače i na pevném disku a pro jejich svižné zpracování je zapotřebí odpovídající hardwarové vybavení - nelze stanovit pravidla (každý software pro zpracování obrázků má své minimální požadavky na hardware počítače), ale platí zde zásada, že čím výkonnější hardware, tím kratší dobu skenování zabere. (Bennex, 2007)

Rychlost práce však kromě počítače ovlivňuje právě skener. Většina přístrojů snímá předlohu o velikosti formátu A4 v rozmezí 20-30 sekund. Před vlastním snímáním je však obvykle nutné podle náhledu dokumentu nastavit oblast snímání a korekce. Náhled skener zvládne zhruba za polovinu uvedené doby. Pokud plánujeme skenovat desítky stran denně, velkým pomocníkem nám bude rychlý skener s rychlým rozhraním, který zabírá na stole sice více místa, ale předlohou formátu A4 zpracuje za cca 10 vteřin a to za pomoci podavače originálů.

Při výběru je rovněž nutné brát ohled na způsob připojení skeneru k počítači. Dnešním standardem jsou především rozhraní USB či starší SCSI. USB skener lze připojit k počítači, který je vybaven odpovídajícím konektorem. Konektory USB jsou v posledních letech již standartní součástí základních desek většiny PC. Do některých starších počítačů je možné doinstalovat kartu (řadič) s USB porty. SCSI skenery jsou určeny pro připojení k SCSI rozhraní, které většina domácích a kancelářských počítačů implementováno nemá. (Pecinovský, 1999)

1.3. Rozpoznávání textu pomocí OCR

OCR je zkratka pro optické rozpoznávání znaků (Optical Character Recognition). Jedná se o technologii převodu textu uloženého v bitmapovém formátu do formátu textového. Je to metoda, která pomocí skeneru umožňuje digitalizaci tištěných i psaných textů, s nimiž pak lze pracovat jako s normálním počítačovým textem, který je reprezentován znaky.

K největším výhodám OCR patří tedy již zmiňovaná možnost nasnímaný text dále upravovat v různých textových editorech a úspora času (kdyby systém OCR neexistoval, všechny texty by musely být přepisovány ručně). Kopecký; Nocar; Kopecký, 2003)

Technologie převodu dokumentu z digitální obrazové do textové podoby pomocí metody optického rozpoznávaní znaků, je založena na porovnání hustoty bodů předlohy na pomyslné síti s typickými znaky jednotlivých písmen uloženými v databázi programu. Míra schopnosti rozpoznávání závisí na úrovni jazykové analýzy a národních slovníků a na možnostech doplňovat porovnávací databázi o nové znaky. Převádět lze tištěné i rukopisné znaky. (OCR, c2005)

Účel a význam této technologie spočívá v rychlém a levném převádění tištěného nebo psaného textu do elektronické podoby (editovatelného počítačového souboru). Rozpoznávání textu je 20 až 25krát rychlejší než ruční přepisování. Pro srovnání si uveďme pár čísel: zatímco velmi rychlá sekretářka napíše 200 znaků za minutu, OCR rozpozná několik stovek znaků za sekundu. (Nutno přidat určitou dobu pro proces skenování a pro obsluhu software.) (Kopecký; Nocar; Kopecký, 2003)

1.3.1. Zpracování a rozpoznávání dokumentu

První fáze je naprosto stejná jako při procesu skenování běžného dokumentu. Ten se převede do počítače, tam proběhne druhá fáze, při které speciální software rozpoznává text znak po znaku a každý z nich porovnává se svou interní databází. Následně vyhodnotí, o jaký znak jde a zapíše jej do textového souboru. Pokud některý znak nerozpozná, má v záloze ještě analýzu celého slova podle vestavěného slovníku.

OCR je speciálním případem vektorizace. Při skenování musí dojít k vektorizaci. Vektorizací rozumíme převod dat z rastrového formátu do formátu vektorového. Jedná se o úlohu obtížnou, neboť informací uložených v rastrovém

formátu je méně, než informací uložených ve formátu vektorovém, a tak je potřeba nové informace automaticky generovat nebo je ručně do dat doplnit. Dnes nejčastěji používanou metodou je poloautomatická vektorizace, jíž provádí program, který je v případě sporných situací korigován a opravován uživatelem. Kvalita vektorizace a její rychlost závisí na stupni automatizace. Text uložený v bitmapě není chápán jako text, je to jen sada tmavých a světlých bodů v obrázku. OCR program tedy musí identifikovat v bitmapě různé tvary a porovnat je se znakovou předlohou a rozhodnout jaké písmeno, ten který shluk představuje. (Habiballa, 2004)

V současné době se rozpoznávání ručně psaných znaků potýká s řadou problémů, které snižují potenciální popularitu OCR. Tyto problémy se týkají zejména odstraňování pozadí, korekce sklonu a velikosti písma, digitálního způsobu přemýšlení, který se liší od lidského. Některé z těchto problémů se podařilo do jisté míry odstranit tzv. předprocesními a poprocesními úpravami (preprocessing a postprocessing), které se vykonávají před nebo po samotném OCR. Je rozumné

oddělovat fáze předprocesní a poprocesní od algoritmů OCR, a to z důvodu, že většina známých algoritmů OCR umí pracovat jen s černými znaky na bílém pozadí.

Typickým příkladem předprocesních algoritmů je odstraňování vzorů na pozadí, vypreparování textu a srovnání šikmo psaného textu, korekce velikosti a sklonu znaků. Postprocesní zpracování je velice důležité, protože slouží k napravování chyb, kterých se případně algoritmus OCR dopustí. Nejtypičtějším případem postprocesního zpracování je kontrola pravopisu (spell checking), která automaticky opraví drobné chyby a u větších chyb se dotáže uživatele na správný tvar.

Naskenujeme-li dokument běžným barevným skenerem, musíme provést prahování šedi (thresholding). Metoda zisku dat spočívá v přiřazení hodnoty 0 (bílá) každému pixelu na pozadí a 1 (černá) pixelům na popředí. OCR algoritmy používají ke své práci také kontextové informace. Počítače obecně nemají problémy s rozpoznáváním dobře napsaných znaků, které se moc neliší od daných vzorů, ale psané znaky jsou mnohdy víceznačné a nečitelné ani pro člověka.

Lidé jsou schopni číst slova s nečitelnými znaky, a tak by to mělo být i u počítačů. Algoritmy založené na principu rozpoznávání znak po znaku by na takovém slově neuspěly, proto technologie OCR pracuje také s informacemi získanými z kontextu. (Psutka, 2006)

1.4. Software využívající technologie OCR

Pro převod obrazového formátu do počítače srozumitelné textové podoby jsou určeny právě OCR aplikace. Převod však téměř nikdy není stoprocentně správný, a různé programy si s ním poradí s různou úspěšností. Výrobci spolu se skenerem často dodávají aplikaci, která je schopna rozpoznávat znaky. Většinou se ale jedná o "odlehčené" verze komerčně dostupných produktů. Některé z nich ale neumí pracovat s českou znakovou abecedou. Jiné nepodporují české znaky vůbec, případně je lze rozpoznávání českých znaků "doučit". Nezanedbatelnou výhodou ukládání na text převedených dokumentů je kromě možnosti úpravy či hledání v textu rovněž skutečnost, že obrazové dokumenty jsou datově mnohem větší, než

jejich textové podoby. Je tedy potřeba použít příslušný obslužný program, který bude schopen se skenerem komunikovat a ovládat ho. (Pecinovský, 1999)

Patrně mezi nejvíce rozšířený software u nás patří OmniPage Pro 17 a TextBridge Pro 11 od bývalé firmy Recognita, dnes Nuance (http://www.nuance.com/). Další populární aplikací je ABBYY FineReader 9 (http://finereader.abbyy.com/) a Readiris 12 (http://www.irislink.com/).

Všechny tyto nástroje podporují češtinu, liší se však cenou, rychlostí a funkcemi. Většina z těchto programů umožňuje exportovat vytvořený dokument do PDF, některé z nich mají dokonce hlasový výstup.

1.5. Spolehlivost a limity technologie OCR

Jak již bylo řečeno, technologie OCR není samozřejmě absolutně dokonalá. Ne vždy dojde k přesnému rozpoznání znaků a text rozpoznaný metodou OCR může obsahovat chyby. Jejich množství závisí na kvalitě původní předlohy. Problémy způsobuje především podtržený text, kurzíva, text s nepravidelnou roztečí a příliš těsný text. Výsledek negativně ovlivňuje především také malý kontrast mezi písmem a pozadím nebo drobné zbytky toneru na originálním dokumentu.

Zpracování textu z tištěné do elektronické podoby je použitelné pro všechny tištěné výstupy z laserových, inkoustových, termosublimačních a jehličkových

tiskáren a samozřejmě pro předlohy vytištěné knihtiskem. U nevhodných předloh např. slabě vytištěných jehličkovými tiskárnami nebo dohromady slitých písmen se z časového hlediska vyplatí spíše přepis textu.

Situace je navíc zkomplikována tím, že texty bývají vytištěny v různých fontech a dokumenty bývají často nekvalitní. Zvláště xeroxované dokumenty bývají "zašpiněné", tzn. obsahují rozmazaná písmena a šmouhy. Program se tedy např. snaží určit, zda tečka poblíž identifikovaného písmena „c“ je háček a nebo jen nějaké artefakt. Většina programů pracuje tak, že dokument procházejí několikrát za sebou a při posledních průchodech už spolupracují s programem na kontrolu pravopisu (spell checker). Mnohé programy se také umí "učit". Takže když chceme převést do textového formátu sadu dokumentů psaných na jednom psacím stroji, můžeme OCR program naučit, že dotyčnému stroji např. ustřelovalo určité písmeno. (AIM, 2000)

2. Počítačové zpracování řeči

V následující kapitole se budu zabývat počítačovým rozpoznáváním lidské řeči a její syntézou. Stručně uvedu její principy a modely se kterými pracuje, stejně tak seznámím se softwarovým vybavením, které slouží k těmto účelům. Zaměřím se také na hodnocení a posuzování kvalitativních aspektů syntetizované řeči. Na závěr kapitoly nebudou opomenuty ani problémy a limity při využití těchto modelů umělé inteligence a jejich použitelnost.

Obor počítačového zpracování řeči prošel v průběhu posledních třiceti let obrovským rozvojem. Z původně úzce zaměřené větve rozsáhlé oblasti číslicového zpracování signálů se postupem času vyvinul samostatný obor s výrazně multidisciplinárním charakterem. Od prvních pokusů s jednoduchými zařízeními rozpoznávajícími či syntetizujícími mluvenou řeč, prováděných koncem 60. let, se pokrok ubíral až k dnešním vysoce výkonným systémům, na jejichž vývoji se podílejí specializované týmy složené z počítačových odborníků, matematiků, lingvistů, fonetiků, zvukových inženýrů i např. psychologů. Problematika komunikace člověka s počítačem pomocí mluvené řeči je v současnosti jednou z nejaktuálnějších oblastí umělé inteligence. Intenzivní výzkum dnes probíhá prakticky na celém světě, což je též odrazem určité specifičnosti oboru, úzce svázaného s

konkrétním národním a jazykovým prostředím. U nás se touto problematikou zabývá např. Katedra kybernetiky ZČU (http://www.kky.zcu.cz/).

I přes významný a pozoruhodný pokrok, kterým se obor za tři desetiletí své existence může prokázat, existují stále nedořešené problémy, bránící širšímu uplatnění výsledků výzkumu v praxi. Většina problémů se týká zejména oblasti automatického rozpoznávání řeči, kde se snad nejvýrazněji střetávají omezené možnosti technicky orientovaného přístupu s dosud nepříliš objasněnou biologickou povahou řečové komunikace. (Mařík; Štěpánková; Lažanský, 1997)

2.1. Syntéza řeči

Syntéza řeči představuje důležitou oblast problematiky zpracování řečového signálu a je již po dlouhá léta předmětem výzkumu. Patří mezi významné úlohy komunikace člověk - počítač. Jde o proces, při němž se uměle vytváří řeč. V dnešní době se řeč vytváří výhradně softwarově s využitím počítače. Před nástupem počítačů a výpočetní techniky se k vytváření umělé řeči používaly mechanická zařízení nebo později elektronické obvody. Umělé vytváření řeči počítačem si klade za cíl “zpřirozenit” komunikaci člověka s počítačem a stát se tak rovnocenným partnerem tradiční viziuální komunikaci. (Mařík; Štěpánková; Lažanský, 2007, s. 299)

2.1.1. Historie

První pokusy o syntetickou řeč byly učiněny již v roce 1779, kdy Christian Kratzenstein vymyslel sadu píšťal, které napodobovaly samohlásky. V polovině 19. století došlo k vytvoření mechanického přístroje, který byl schopen reprodukovat některá slova. Tyto syntetizéry vycházely z fyzického napodobování mluvícího ústrojí. Roku 1922 byl vyvinut plně elektrický přístroj. Jedno z prvních významných zařízení v historii řečové syntézy byl řečový syntetizér zvaný VOCODER, který byl uveden v New Yorku v roce 1939. Původně sloužil ke komunikaci, pro přenos zvuku

radiem a transkontinentálním kabelem. První TTS (text-to-speech) engine (přístroj, který dokázal přečíst libovolný text) se objevil v roce 1968. Od 80. let pokračuje vývoj rozvojem softwarových aplikací, jejichž účelem je reprodukovat vstupní text co nejvěrněji a nejpřirozeněji v porovnání s lidskou řečí. (Bartůšek; Nygrýn, 2000)

2.1.2. Model vytváření řeči

←

Lidská řeč je charakterizována:

- akustickou strukturou (zvukovým spektrem měnícím se v čase)
- lingvistickou strukturou (gramatikou a skladbou)
- subjektivním vlivem osobnosti řečníka (intonace, rytmus, barva hlasu)

Fyzikální reprezentací řeči jsou akustické řečové kmity, které jsou tvořeny lidskými řečovými orgány. Ty jsou tvořeny hlasivkami, dutinou hrdelní, ústní a nosní, měkkým a tvrdým patrem, zuby a jazykem. Jako základní zdroj hlasové energie slouží plíce a s nimi spjaté dýchací svaly. Zdrojem (generátorem) znělých zvuků jsou kmitající hlasivky uložené v horní části hrtanu. Pod tlakem vycházejícího z plic dochází ke kmitání hlasivek. Frekvence kmitu hlasivek určuje základní tón lidského hlasu. Ten se liší u dětí, dospělých, žen i mužů. Většinou se pohybuje v rozmezí 150 až 400 Hz. V akustickém spektru každé samohlásky se objevují zesílené tóny vznikající rezonancí v dutinách hlasového traktu, nazývají se formanty.

Při vyslovení určitého zvuku musí různé části hlasového ústrojí zaujmout tomuto zvuku odpovídající počáteční polohu a v procesu realizace tohoto zvuku měnit polohu hlasových orgánů předem definovaným způsobem. Tímto vznikají odpovídajícím polohám odpovídající akustické signály, ze kterých se formují zvukové elementy - fonémy - umožňující rozlišovat zvukovou podobu slova. Foném je základní jazyková jednotka schopná rozlišit význam, např. les - pes.

Problémem je, že v různém kontextu dochází ke změnám výslovnosti jednotlivých fonémů. Tento jev označujeme jako koartikulace, kdy akustická realizace fonémů závisí jak na předcházejícím a na následujícím zvuku, tak i na tempu a intonaci řeči. Koartikulace tedy značně komplikuje složitost dalších postupů zpracování, ať už pro účely hlasové syntézy nebo rozpoznávání řeči.

Cílem modelování produkce řeči je nalézt matematické vztahy, které by mohly být využity pro reprezentaci akustických fyzikálních dějů spojených s touto produkcí. Ideální by bylo, kdyby takový model měl minimální složitost, byl lineární a časově neproměnlivý. Lidská řeč je však souvislý, časově proměnný proces vykazující i nelineární charakteristiky. Nelze se proto divit, že zatím nebyl předložen jediný univerzální model, který by respektoval tyto požadavky. (Mařík; Štěpánková; Lažanský, 1997, s. 216)

2.1.3. Principy syntézy řečového signálu

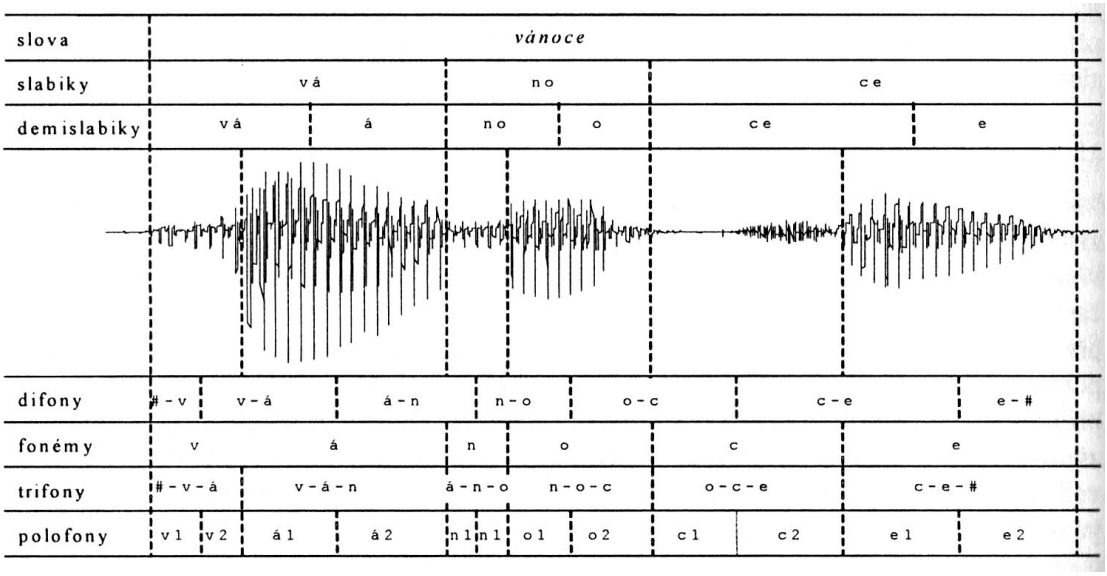
Obvyklý způsob, jak vytvořit řečový syntetizovaný signál, je vybrat základní stavební akustické jednotky, zpracovat a uložit je v paměti počítače a posléze, ve

vhodném okamžiku je generovat tak, že se pospojují dohromady vhodné segmenty z tohoto předběžně uloženého inventáře. (Mařík; Štěpánková; Lažanský, 1997, s. 224)

Při takovéto počítačové syntéze řeči je třeba uvážit dvě důležitá hlediska. První se týká fonetických a lingvistických jednotek, ze kterých budeme řeč vytvářet a druhé způsobu zpracování a vytváření akustických kmitů.

Jako základní stavební jednotka syntézy může sloužit věta, slovo, slabika či jen foném. Čím je základní řečová jednotka delší, tím více promluv musí být v systému zpracováno a uloženo. Vzhledem k neúměrným požadavkům na takto zaznamenané promluvy se jako optimální řešení nabízí využití fonému jako základní stavební jednotky syntetizované řeči. Bohužel s čím menší stavební jednotkou pro syntetizovanou promluvu pracujeme, tím více se projevuje vliv nesprávné koartikulace při spojování těchto jednotek. Nesprávné koartikulační efekty mají vliv především na plynulost syntetizované řeči.

obr. 1 Ukázka segmentace řečového signálu slova Vánoce na základní řečové jednotky. *Zdroj:* (Mařík; Štěpánková; Lažanský, 2007, s. 306)



Další významný jev, objevující se v přirozeném mluveném projevu, se nazývá prozódie. Prozódie je informace přítomná v řeči v podobě prozodických příznaků, mezi které patří intonace, hlasitost a časování (rytmus) řeči. Změna intonace se může projevovat ve změnách melodie hlasu např. v závěrečné části tázacích vět. Intonace a časování se během řeči spojují v jev zvaný přízvuk. Poslední

prozodickou charakteristikou je hlasitost (amplituda) promluvy. Ta však hraje z hlediska reprezentace významu daleko menší roli než změny časování a intonace. Nutno říci, že obraz změn hlasivkového tónu a rytmu překlenují mnohem delší rozpětí řeči, než jsou jednotlivá slova. Tyto příznaky tedy nelze uplatnit při syntéze promluvy jednoduchým řazením slov uložených ve slovníku. Při syntéze řeči lze postupovat v časové nebo frekvenční oblasti. (Bartůšek; Nygrýn, 2000)

Syntéza v časové oblasti (konkatenační syntéza)

Tato syntéza je také nazývána konkatenační syntézou, neboli syntézou spojováním. V současné době se využívá při metodě TTS (text-to-speech).

Při tomto typu zpracování provedeme digitalizaci reálného řečového signálu, tyto vzorky potom uložíme v paměti a opačným procesem je zpětně generujeme, nebo zaznamenaná data editujeme za účelem vytvořit soubor kratších jednotek (slov, slabik, fonémů) a jejich spojováním řeč syntetizovat. Při výběru jazykových jednotek se vědci řídí cílovou aplikací a požadovanou kvalitou syntézy. Například při výběru slova jako základní jazykové jednotky není technicky dobře možné dát do větných celků intonaci a koartikulační efekty. Za těchto okolností zní syntetizovaná řeč monotónně, avšak nemá strojový charakter a je dobře srozumitelná.

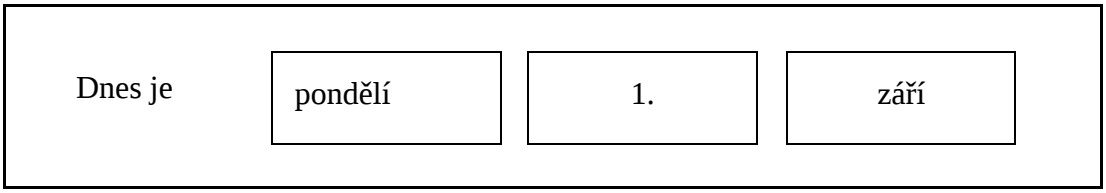
Problémům s koartikulací a prozodií v syntetizované řeči se lze vyhnout, pokud budeme zaznamenávat celé fráze a věty. Takto reprodukovaný signál potom působí zcela přirozeně. Toto řešení je kvůli velkým paměťovým nárokům využitelné pouze v situacích, kdy stačí mít k dispozici omezený slovník promluv.

V takovýchto případech lze ke snížení paměťových nároků dospět pomocí využití metody struktury rámců. Jako příklad si můžeme uvést hlasový automat, který nám větou "Dnes je pondělí 1. září" oznámí aktuální datum a den v týdnu. V této větě se mění položky jen na určitých pozicích. Těmto položkám říkáme tzv.

sloty a tvar věty nazýváme větným rámcem. Změnou těchto položek (slotů) dostaneme větu s jiným významem.

obr. 2 Znázornění principu syntézy vět na základě využití struktury rámců
Zdroj: (Mařík; Štěpánková; Lažanský, 1997, s. 255)

Pro dobrou srozumitelnost syntetizované promluvy je zde velmi vhodné dosazovat do rámcové struktury slova, která jsou foneticky konzistentní s jejich



pozicí v promluvě i s intonací promluvy. Sloty proto obvykle obsahují vedle znalosti o přesném datu i znalosti o koartikulačních efektech, které mohou doprovázet spojování slov.

Syntéza ve frekvenční oblasti

Syntéza v časové oblasti se týkala reprodukce řečových vzorků, které byly uloženy v paměti počítače. Tato syntéza zvaná též syntéza kódováním zdroje se týká reprodukce frekvenčního spektra řečového signálu. Syntéza ve frekvenční oblasti se nesnaží rekonstruovat řečové kmity jedna k jedné, jako tomu bylo při reprodukci kódovaných kmitů v časové oblasti, ale místo toho ve fázi analýzy zpracovává řečový kmit a uchovává ho zvlášť jako matematický model jeho frekvenčního spektra. Tento syntetizér potom napodobuje funkci lidských hlasových orgánů využitím frekvenčních generátorů, filtrů a zesilovače, které jsou řízeny uchovanými parametry. Protože lze nezávisle ovládat zisk, periodu základního tónu i parametry modelu hlasového traktu, nabízí syntetizér na bázi kódování zdroje možnosti ovlivňovat prozodické charakteristiky syntetizované řeči. (Mařík; Štěpánková; Lažanský, 1997, s. 225)

Hlavím typem syntézy ve frekvenční oblasti je formantová syntéza. Ta je založena na akustickém modelování hlasového traktu pomocí formantů. Formanty, neboli vyšší harmonické či částkové tóny, jsou frekvence, které jsou spolu se základní harmonickou frekvencí obsaženy v každém tónu lidského hlasu a vznikají rezonancí v hlasovém ústrojí. Frekvenční syntetizér, který syntézu obstarává, pracuje s matematickým modelem artikulačního ústrojí člověka. Formantová syntéza se používala s úspěchem pro úlohy TTS v 60.-80. létech, ale vzhledem ke složitosti vytvoření dostatečně kvalitního modelu hlasového ústrojí, nejsou dnes frekvenční syntetizéry v této oblasti téměř používány. (Bartůšek; Nygrýn, 2000)

2.2. **Syntéza řeči z textu (text-to-speech)**

←

Jednou z řešených úloh řečových technologií je i syntéza řeči z textu, někdy také nazývaná převod textu na řeč. Patří mezi nejobtížnější úlohy počítačové řečové syntézy.

Cílem metody TTS (text-to-speech) je generování řeči z libovolného textu. Konečným cílem syntézy řeči z textu je převádět psaný text na řeč v takové formě a kvalitě, jakou by stejný text přečetl člověk s dobrým přednesem. Syntetická řeč by tedy neměla působit jednotvárně, měla by znít přirozeně a její poslech by neměl unavovat ani vyžadovat zvýšenou pozornost.

Při řešení tohoto úkolu jsou nejčastěji využívány principy produkčních systémů, které známe z umělé inteligence. Nyní se podíváme na jednotlivé komponenty systému TTS.

Systém pro syntézu řeči z textu se skládá ze dvou základních modulů:

- modul pro zpracování přirozeného jazyka NLP (Natural Language Processing)
- syntetizér řeči

Oba tyto moduly jsou na sobě relativně nezávislé, ale oba mají svoji důležitost. Modul pro zpracování přirozeného jazyka zajišťuje analýzu textu, který do systému přichází na jeho vstupu, a snaží se získat pokud možno co nejvíce informací o tom, co se má syntetizovat. Tento modul provádí fonetickou transkripci textu, díky které se potom na výstupu objeví posloupnost hlásek příslušného jazyka doplněná o

prozodické značky. Poté přichází na řadu řečový syntetizér, který na základě těchto informací generuje řečový signál. (Mařík; Štěpánková; Lažanský, 2007, s.. 301)

Modul pro zpracování přirozeného jazyka se skládá z následujících komponent:

- modulu fonetické transkripce
- generátoru prozodie
- morfologicko-syntaktického analyzátoru MSA (Morphological Syntactic Analyzer)

Morfologicko-syntaktický analyzátor slouží pro rozpoznání kontextu a syntaktické informace analyzovaných slov. Tento analyzátor se skládá z několika dalších komponent, ty jsou v současných systémech TTS řazeny paralelně. Každá z komponent totiž může do zpracovávaného textu v průběhu zpracování přidat novou informaci, ostatním komponentám do té doby nepřípustnou. Tato nová informace pak může způsobit opětovné "zavolání" příslušné komponenty, která na jejím základě např. zpřesní odhad některých vlastností zpracovávaného textu. Všechny tyto komponenty se tedy mohou vzájemně ovlivňovat. (Mařík; Štěpánková; Lažanský, 2007, s. 302)

Mezi komponenty MSA patří:

- blok předzpracování textu (detekuje typ vstupního textu, filtruje jeho znaky, např. nadbytečné znaky formátování)
- blok normalizace textu (normalizuje text do plné slovní formy např. číslovky, zkratky a symboly přepisuje na jejich slovní vyjádření - 200 Kč na dvě stě korun českých)
- morfologická analýza (navrhuje všechny možné mluvnické kategorie každého slova věty izolovaně bez kontextu sousedních slov)
- kontextová analýza (o jednotlivých slovech uvažuje v kontextu okolních slov a redukuje tak seznam všech možných mluvnických kategorií slov, získaných v předchozím kroku)
- syntakticko-prozodický rozbor (slouží k nalezení větných úseků ve zpracovávané větě)

Nyní se ještě stručně vraťme k charakteristice dvou zbývajících modulů pro zpracování přirozeného jazyka:

- modul fonetické transkripce (jeho cílem je popis fonetické podoby textu)
- generátor prozodie (na základě zpracovávaného textu odhaduje a generuje průběhy prozodických charakteristik, tj. melodie, časování a intenzity vytvářené řeči)

Při syntéze TTS musí nejprve dojít k fázi zpracování textu (tzv. fonetické transkripci) a dále navržení prozodických charakteristik. Nakonec přichází na řadu morfologicko-syntaktický analyzátor.

2.3. Rozpoznávání řeči

2.3.1. Principy rozpoznávání jednotlivých slov

Základní principy rozpoznávání řeči lze poměrně dobře ilustrovat na úloze rozpoznávání izolovaně pronášených slov patřících do malého či středně velkého slovníku. Omezení, která tato úloha přináší, jsou pro řadu aplikací a pro srozuměného uživatele přijatelná a z hlediska návrhu systému představují podstatné zjednodušení problému. U slov oddělených dostatečně dlouhou pauzou (např. 1 s) lze použít poměrně jednoduchou strategii hledání jejich začátku a konce. Systém může vycházet z předpokladu, že ve vymezeném intervalu se nachází právě jedno slovo. Je-li navíc slovník tvořen pouze desítkami či několika málo stovkami slov, může být slovo považováno za základní (dále nedělitelnou) jednotku a takto zpracováváno. Lze tím obejít úlohu, která je ve své podstatě mnohem složitější, a to identifikaci jednotlivých hlásek (fonémů). Přitom stejný postup je možné uplatnit i na sousloví a krátké fráze, pokud jsou vyslovovány souvisle a ve slovníku jsou reprezentovány jedinou položkou.

2.3.2. Základní schéma rozpoznávání

Na vstupu systému je analogový signál snímaný mikrofonem, na výstupu se objevuje informace o přiřazení slova (pokud bylo v signálu detekováno) k jedné z položek slovníku. Jednotlivé fáze zpracování tvoří digitalizace signálu, jeho následná parametrizace, určení začátku a konce slova a samotná klasifikace. Převod

analogového signálu přicházejícího z mikrofonu do odpovídající číslicové (digitální) podoby je dnes rutinní záležitostí, kterou zajistí každý A/D (analogově digitální) převodník či zvuková karta. Pro účely rozpoznávání je dostačující použít převodník se šestnáctibitovým rozlišením, nastavený na vzorkovací frekvenci alespoň 8 kHz. Bitovou hloubku 16 bitů a vzorkovací frekvenci 8 kHz lze zároveň považovat za jistý standard, se kterým pracuje většina komerčních produktů v oblasti rozpoznávání řeči. Vychází se z faktu (ověřeného dlouholetými zkušenostmi s telefonním signálem), že k porozumění obsahu mluvené řeči stačí přenosové pásmo s horní frekvencí okolo 3 000 Hz. Ve zvukovém spektru tedy pásmo středů, kde je lidské ucho nejcitlivější. Potřebné zesílení mikrofonního signálu vykonává mikrofonní předzesilovač, který může být integrován ve zvukové kartě.

Další zpracování číslicového signálu se děje po úsecích, které se v angličtině i v češtině označují jako framy. Frame představuje krátký segment signálu o délce 10 - 25 milisekund. Délka framu bývá v daném systému konstantní a volí se tak, aby nebyla delší než trvání nejkratší hlásky. Pak lze signál v rámci jednoho framu považovat za stacionární a lze jej popsat menším počtem parametrů. To je také hlavním cílem segmentace: snížit tok dat směřujících do vyšších úrovní systému. Parametry vybrané pro popis signálu, neboli příznaky, musí splňovat několik požadavků. Měly by umožnit dostatečné odlišení framů reprezentujících různé akustické a řečové jevy, zároveň by měly být voleny tak, aby pokud možno potlačovaly ty rysy signálu, které jsou svázány s individualitou řečníka (např. výška a síla hlasu), a v neposlední řadě by měly být výpočetně dostupné. (Habiballa, 2004)

V současné době jsou nejčastěji používanými příznaky spektrální a zejména kepstrální koeficienty, které se obvykle doplňují o hodnotu energie signálu a o odvozené dynamické parametry. Typický příznakový vektor popisující jeden frame signálu se skládá z 12 až 40 parametrů. Některý z příznaků, nejčastěji to bývá energie, slouží v následujícím bloku k určení začátku a konce slova.

Detektor řeči většinou funguje na principu stavového automatu, jehož činnost je určena aktuální hodnotou daného příznaku a sadou několika řídicích parametrů.

Některé z těchto parametrů musí být určeny předem, např. maximální a minimální přípustná délka slova či maximální délka pauzy uvnitř promluvy, jiné, např. prahové úrovně související s hladinou šumu prostředí, mohou být nastavovány automaticky a průběžně aktualizovány. Výsledkem činnosti detektoru je stanovení počátečního a koncového framu řeči a určení délky slova. Celé slovo je zredukováno na matici hodnot typu F x P, kde F je počet framů náležejících slovu a P je počet příznaků používaných k popisu signálu. Tato matice pak reprezentuje slovo při jeho rozpoznávání. (Kopeček; Politzer, 1999)

Podle několika základních kritérií lze reálné rozpoznávací systémy dělit na:

- systémy určené pro spojitou řeč, resp. systémy pro izolovaná slova
- systémy nezávislé, resp. závislé na mluvčím (s nutností adaptace na konkrétní osobu)
- systémy s omezeným, resp. neomezeným (parametricky definovaným) slovníkem

2.3.3. Metody rozpoznávání řeči

I přes určitou odlišnost, pramenící zejména z rozdílné složitosti a rozdílného počtu úrovní zpracování informace, mají všechny uvedené systémy řadu společných prvků a postupů. Patří k nim především proces parametrizace akustického signálu a dále určité přístupy a procedury klasifikace parametrizovaného signálu. Parametrizace sleduje dva základní cíle, a to: průběžně reprezentovat signál vybranými parametry neboli příznaky, které jsou vhodné pro jeho další zpracování,

a zároveň tím významně snížit datový tok směřující do vyšších úrovní rozpoznávacího systému.

V průběhu minulých desetiletí byla problému parametrizace věnována velká pozornost. Od původně jednoduchých, snadno dostupných, avšak nepříliš účinných příznaků se vývoj ubíral přes spektrální příznaky dosažitelné softwarově prostřednictvím rychlé Fourierovy transformace (FFT) či hardwarově pomocí soustavy filtrů, až k dnešnímu, téměř univerzálně používanému popisu pomocí

parametrů tzv. kepstra. (Název kepstrum, vzniklý přesmyčkou od slova spektrum, naznačuje, že jde o určitý inverzní obraz zvukového spektra.) (Speech recognition, 2001)

Cílem klasifikace je rozhodnout o zařazení určitého úseku signálu řeči do některé z předem definovaných kategorií, např. hlásek fonetické abecedy nebo slov z daného slovníku. Protože ale jde o signál do značné míry náhodný, lze takové rozhodnutí učinit jen s určitým stupněm věrohodnosti, vycházejícím z míry podobnosti mezi parametry daného úseku řeči a jistými referenčními vzory, které má klasifikátor k dispozici. Úloha se však dále komplikuje tím, že řeč je proces vyvíjející se v čase, tedy proces popsaný ne jedním vektorem příznaků, ale posloupností těchto vektorů. Procedura porovnávání signálu se vzory se tak stává úlohou dynamickou.

I v této oblasti přinesl vývoj celou řadu metod. Z nich stojí za zmínku metoda dynamického borcení času (Dynamic Time Warping DTW), slavící úspěchy zejména v 70. a 80. letech.

Metoda DTW se obecně hodí pro rozpoznávací systémy, které jsou určeny pro jediného uživatele. Ten musí ještě před prvním použitím namluvit (jednou či vícekrát) všechny položky slovníku. Má-li systém současně sloužit více mluvčím, měl by každý z nich mít v paměti své vzory. Při vícenásobných referencích pak vzrůstá naděje, že systém bude poměrně dobře spolupracovat i s člověkem, na něhož není adaptován. Nároky na paměť slovníku a tudíž i na čas klasifikace tím samozřejmě úměrně vzrůstají.

V současné době se daleko častěji používají systémy založené na parametrických modelech, z nichž nejvýznamnější roli hrají tzv. skryté markovské

modely (zkr. HMM z anglického Hidden Markov Model). V případě HMM jde o abstraktní statistický popis, který má jen málo společného s předchozími referencemi. Model slova je tvořen několika vzájemně propojenými stavy, přičemž každému z nich je distribuční funkcí přiřazena jistá část prostoru příznaků. Také přechody mezi jednotlivými stavy jsou charakterizovány pravděpodobnostními koeficienty. Díky tomuto statistickému popisu se chování modelu jeví jako částečně náhodné, jakoby skryté za rouškou utkanou z dvojnásobně nejisté (pravděpodobnostní) mlhy.

Právě tato míra náhody se však dobře hodí pro signál typu lidské řeči. Modely jednotlivých slov mají stejnou strukturu, často i stejný počet stavů, liší se pouze statistickými parametry. Jejich hodnoty se stanovují v průběhu tzv. trénování, kdy se speciálními algoritmy extrahují z předložených vzorků daného slova. Obecně platí, že čím více trénovacích vzorků (nejlépe od mnoha různých mluvčích) použijeme, tím kvalitnější a univerzálnější jsou modely.

Klasifikace probíhá podobně jako u metody DTW. Každou položku slovníku zastupuje jeden model, který je porovnáván s reprezentací neznámého slova. Tentokrát se však neměří vzdálenost, ale pravděpodobnost, s jakou se model blíží této reprezentaci. Výpočet je poměrně náročný, ale i zde existují urychlující postupy vycházející z principu dynamického programování. Ze všech modelů se pak nalezne ten, jehož pravděpodobnostní hodnota je maximální.

Hlavní výhodou techniky HMM je fakt, že umožňuje reprezentovat slova ve slovníku takovým způsobem, který je téměř nezávislý na konkrétním mluvčím. Je-li rozpoznávací systém jednou natrénován, může jej s velkou úspěšností používat prakticky kdokoliv. (Habiballa, 2004)

2.4. Software využívající řečové technologie

Softwarové programy, které jsou pro účely analýzy a syntézy řeči k dispozici, lze rozdělit podle funkcí do následujících kategorií:

- programy pro převod textového souboru na řečový výstup
- programy pro převod obsahu obrazovky (v textovém režimu) na řečový výstup
- programy, které čtou menu a zprávy různých aplikací

- programy rozpoznávající určitou množinu příkazů v lidské řeči (například "nahoru", "vlevo", "ukončit", "uložit" a podobně) a ovládající jiné programy v závislosti na těchto příkazech.
- programy, které dokáží převádět mluvenou řeč ve formě oddělených slov na psaný text, který může být později např. v textovém editoru upraven
- programy, které převádí přirozený plynulý mluvený projev na psaný text

2.4.1. Aplikace s hlasovým výstupem

- Voice Leader - aplikace od společnosti Linguatec dostupná v různých variantách za velmi příznivou cenu, umožňuje hlasový výstup v českém jazyce. (http://www.linguatec.net/products/tts/voice_reader)
- KUK - česká aplikace umožňující číst obsah textové obrazovky, pracovat s různými předem definovanými programy, u kterých umí číst menu a také předčítat text ze souboru. Tento program je určen pro prostředí MS DOS.
- Text Assist - anglická aplikace pro Microsoft Windows. Umožňuje číst text z textového souboru, lze nastavit různé vlastnosti hlasu jako je výška, rychlost.
- CS Voice - česká aplikace pro Microsoft Windows. Umožňuje číst text z textového souboru v českém jazyce. (http://www.frog.cz/prod04.htm)

←

2.4.2. Aplikace pro ovládání počítače jednoduchými povely

Tyto aplikace jsou nezávislé na jazyce, protože povely jsou uložené jako zvuky. Jejich náročnost na výkonnost počítače není příliš vysoká, dostačuje i méně výkonný počítač. Všechny tyto programy jsou závislé na mluvčím, hlavní podmínkou je předem namluvit všechny příkazy.

- Voice Assist - aplikace pro MS Windows dodávaný se zvukovými kartami Sound Blaster
- In Cube - aplikace MS Windows
- IBM VoiceType Control - aplikace pro systém MS Windows 95

- JetVoíce - hlasové záznamy jsou zde vidět v obrazové podobě, kde lze rozpoznat jak vaše intonace závratně mění způsob, jakým počítač zvuk vnímá.
- MyVoice - umožňuje ovládat počítač a na něm nainstalované programy pomocí hlasových povelů, diktovat text, zadávat nové hlasové povely, řídit pohyb myši, ovládat MS Word. (http://www.fugasoft.cz/index.php?cont=myvoice) (Vorlíček, c1999)

2.4.3. Aplikace pro převod mluveného projevu na text

Tyto aplikace jsou nezávislé na mluvčím, i když vyžadují určitou adaptaci. Tato adaptace se provádí za provozu programu pomocí zpětné vazby. Při každém rozpoznaném slovu je nabízeno více variant a uživatel má možnost pomocí myši vybrat tu správnou. Pokud tak neučiní, počítač sám vybere nejpravděpodobnější variantu. Tím se zároveň adaptuje na konkrétního mluvčího. Nevýhodou je, že pro tuto adaptaci by nevidomá osoba musela mít nějakého vidomého pomocníka. Lze ale říci, že po určité době je program tak dobře zadaptovaný, že nepotřebuje další úpravy.

- Dragon Naturally Speaking od společnosti Nuance dokáže převádět řeč diktovanou jako oddělená slova na text. Také zahrnuje funkci ovládání programů slovními povely. (http://www.nuance.com/naturallyspeaking/)
- IBM ViaVoice (http://www.nuance.com/viavoice/)

Aplikace IBM Via Voice stejně jako produkt firmy Nuance, zahrnuje rozpoznávání řeči a její převod na text (speech-to-text) ale disponují i technologií TTS.

2.4.4. ATToolkit (Automatic Transcription Toolkit)

Tento systém vymyslela a vyvíjí Laboratoř počítačového zpracování řeči na TU v Liberci. (https://www.ite.tul.cz/speechlab/)

Systém umožňuje plnit tyto funkce:

- plně automatický systém pro on-line sledování a transkripci televizních a rozhlasových pořadů
- paralelní sledování několika stanic (klastr počítačů s paralelním distribuovaným zpracováním dat)
- automatická detekce řeči a klasifikace neřečových zvuků
- rozpoznávání identity mluvících osob (několik set osob v databázi –moderátoři, redaktoři, politici)
- rozpoznávání řeči následované formátovaným přepisem řeči
- ukládání dat v datových skladech (video, audio, přepis, časové značky)
- možnost vyhledávání libovolného slova (ze slovníku) a jeho okamžité přehrání.

(Laboratoř počítačového zpracování řeči, c2009)

2.5. Problematika řečových technologií, spolehlivost a limity

Je nutno uvést, že současný stav poznatků zatím neumožňuje člověku komunikovat s počítačem plynulou řečí, bez omezení a na jakékoliv téma. Je to způsobeno stálými obtížemi s rozpoznáváním plynulé řeči a i omezenými možnostmi součinnosti procesu klasifikace řečového signálu a procesu porozumění smyslu posloupnosti klasifikovaných slov. To má za následek, že systémy hlasového dialogu s počítači jsou zatím využívány pouze v jednodušších úzce problémově zaměřených aplikacích (konzultace databázových systému, automatické informační a rezervační systémy apod.), a nebo jsou nabízeny systémy, které jsou schopny řešit jednotlivé

subúlohy objevující se při hlasovém dialogu s počítačem, tedy např. počítačovou syntézu. (Mařík; Štěpánková; Lažanský, 1997, s. 215)

Při všech pokusech směřujících k návrhu efektivního systému rozumějícího přirozené řeči se až dosud naráželo především na tyto stěžejní problémy:
Spojitý charakter mluvené řeči, kdy lze v toku řeči jen velmi obtížně určit hranice jednotlivých slov, efekt koartikulace (vzájemné ovlivňování sousedních hlásek), značná variabilita řeči, jak u různých mluvčích, tak i u jednotlivého řečníka, jehož

projev může být ovlivněn mnoha faktory, jako např. stres, únava, nemoc, atd. Pro kvalitní analýzu audio signálu mohou být nežádoucí zvuky z prostředí (ruchy), v němž je řeč zaznamenávána (hluk, šum, vítr, déšť), dále je třeba brát ohledy na kvalitu přenosové linky (telefon), která může způsobovat degradaci kvality signálu a v neposlední řadě i velké množství informací obsažené v řečovém signálu, z něhož nás však ve většině případů zajímá pouze malá část (nejčastěji jen obsah sdělení).

Všechny dosud vyvinuté systémy rozpoznávání řeči jsou proto vždy kompromisem mezi dostupnými znalostmi o tvorbě a vnímání řeči na jedné straně a současnými technickými možnostmi řešit výše uvedené problémy na straně druhé. V řadě konkrétních aplikací však i tato kompromisní řešení mohou plnit a plní požadovaný účel. Nyní se podívejme, jak lze hodnotit přirozenost a kvalitu syntetické řeči. (Psutka, 2006)

2.5.1. Hodnocení kvality syntetické řeči

Chceme-li zhodnotit kvalitu promluvy syntetickou řečí a máme k dispozici původní promluvu, není problém změřit rozdíly obou promluv a na jejich základě zhodnotit kvalitu rekonstruované řeči.

V případě syntézy libovolné promluvy metodou text-to-speech, kdy nemáme původní promluvu k dispozici, přicházejí na řadu poslechové testy. Při těchto testech skupiny osob subjektivně hodnotí kvalitu syntetizované řeči. Kvalitou máme na mysli celkovou srozumitelnost, přirozenost a plynulost řeči posuzované uživateli. Je nutno uvést, že vzhledem ke komplexnosti řeči neexistují žádné objektivní testy či metody. Jistá "objektivita" je zajištěna dostatečně velkým počtem osob (posluchačů), kteří se prováděných testů účastní. Poslechové testy slouží k upozornění na chyby, kterých se syntetizér může dopouštět a pomáhají tak k jejich odstraňování. Rozlišujeme dva základní typy poslechových testů:

- testy srozumitelnosti
- testy přirozenosti

Testy srozumitelnosti se soustředí na porozumění syntetické řeči posluchači. Jeden z nejpoužívanější poslechových textů se nazývá MRT (Modified Rhyme Test) neboli test modifikací rýmu. Tento typ testu byl původně navržen pro angličtinu, ale postupně byl aplikován do mnoha jiných jazyků. Posluchačům je přehráno 50 skupin slov po 6ti slovech. Úkolem každého posluchače je identifikovat o jaké slovo ze skupiny se jedná. Tato slova se od sebe liší počátečním nebo koncovým fonémem. Jak je možno vidět na příkladu, používají se vždy jednoslabičná slova. (Psutka, 2006, s. 627)

obr. 3 Ukázka skupin slov pro hodnocení srozumitelnosti české řeči pomocí testu MRT *Zdroj:* (Psutka, 2006, s. 628)

Velkou výhodou těchto testů je jejich velká spolehlivost, možnost porovnání s testy provedenými s jinými syntetizéry a dostačující poměrně malý počet "nekvalifikovaných" posluchačů. Nevýhodou je především možnost korigovatelnosti

pyl	pih	pij	piš	piv	pin
pes	les	ves	bez	děs	rez
důl	hůl	vůl	sůl	půl	kůl
rak	tak	sak	vak	lak	pak
lev	les	lem	lep	led	len
byt	lid	kyt	Žid	hit	vid
loď	lom	lov	lok	los	lob
kos	bos	los	nos	šos	sos
bál	šál	tál	vál	sál	kál
suk	puk	kuk	luk	muk	fuk

rozhodnutí posluchače podle seznamu slov v dané skupině.

Druhým typem testu srozumitelnosti je test SUS (Semantically Unpredictable Sentences) neboli test za použití sémanticky nepredikovatelných vět. Za sémanticky nepredikovatelné věty se považují takové věty, které ač jsou gramaticky správné, přesto nedávají smysl. Příkladem budiž věta: ,,Zkažený automobil jedl nové kohouty''. Úkolem posluchačů v tomto testu je přesně zapsat slovo od slova z věty, které jim jsou přehrávány. Cílem těchto testů je minimalizovat možnost odvození

nesrozumitelných slov z kontextu věty a okolních slov. Posluchač tedy nemůže využít zkušenosti a neznáme slovo si domyslet. (Psutka, 2006, s. 627)

Druhým typem poslechových testů, jak už jsem zmínil, jsou testy přirozenosti syntetické řeči. Někdy jsou také nazývány celkovými testy kvality. Jejich cílem je porovnávat a hodnotit řeč podle celkové kvality vjemu, tedy z komplexního hlediska. Soustředí se tedy nejen na srozumitelnost řeči, ale také na celkovou kvalitu řečového projevu, posluchatelnosti a přirozenosti promluvy. Opět se jedná o subjektivní testovací metody.

Do této skupiny paří test MOS (Mean Opinion Score). Výsledkem testu je hodnocení posluchačů, kteří hodnotí jednotlivé promluvy tak, jak jsou jim přehrávány stupnicí bodů 1 - 5. K výsledku testu dospějeme vypočtením průměru všech výsledků u všech posluchačů.

Druhou často používanou metodou testu přirozenosti promluv je test CCR (Comparsion Category Rating) čili test porovnávání párů, při němž dostává posluchač hodnocenou větu vždy ve dvou vzorcích po sobě, pojmenovaných např. vzorek 1, vzorek 2. Každý z těchto vzorků je vytvořen např. odlišným systémem TTS. Posluchač volí mezi oběma vzorky ten, který mu přijde lepší. Pro dodržení objektivity je voleno náhodné pořadí vzorků ke každé větě. Celkově však musí být dodrženo, že v padesáti procentech bude jako první v pořadí přehráván vzorek vytvořený prvním TTS systémem a ve zbylých případech vzorek vytvořený druhým typem systému TTS. V tomto textu se tedy vždy porovná kvalita dvou vzorků, které jsou generovány odlišnými text-to-speech systémy. (Matoušek, 2006)

Všechny zde zmíněné metody byly jen nástin do oblasti rozpoznání a syntézy řeči. Celá tato problematika je velice náročným tématem oblasti umělé inteligence, kterým se neustále zabývají vědci ze špičkových ústavů (patří sem i např. špičkové pracoviště na ZČU či liberecká laboratoř Speechlab), protože stále je co zdokonalovat. Použití výsledných aplikací má své nepostradatelné místo v oblastech dnešního života.

3. Praktické využití technologií v informačním prostředí

V poslední kapitole mojí práce se budu zabývat aplikací již zmíněných technologií a praktickým využitím těchto metod v informačním prostředí společnosti kde žijeme.

Současný stav technologie OCR a technologie rozpoznávání řeči spolu s možnostmi dnešních počítačů dovolují, aby řada konkrétních aplikací našla své potenciální uživatele. Komerčně dostupné systémy se objevují zejména v těchto

oblastech: jednoduché (několikapovelové) ovládání různých zařízení a přístrojů (např. doplňků v automobilu), hlasové řízení některých počítačových aplikací, hlasový diktát rozšiřující možnosti stávajících textových procesorů, veřejné informační systémy s hlasovým vstupem a výstupem, automatické informační, dopravní a rezervační služby s přístupem po telefonu, nástroje a kompenzační pomůcky pro tělesně postižené. Neméně důležitou funkcí je zásluha o rozvoj informační společnosti a začlenění handicapovaných jedinců do běžného života. Nyní se podrobněji podíváme na jednotlivé oblasti jejich využitelnosti.

3.1. Využití digitalizace dokumentů a technologie OCR

Digitalizace je využívána hlavně v rozsáhlých archivech. Jde především o stavební archivy, které obsahují různorodé materiály. Dost často se digitalizuje i živnostenský archiv, protože jde o relativně nový archiv, který obsahuje dokumenty vytvořené až po roce 1989. K hlavním přednostem digitalizovaných dokumentů patří např. možnost jejich trvalé archivace, uchránění původních nosičů před fyzickým znehodnocením a zlepšení služeb s ohledem na uživatele (lepší dostupnost dokumentů, odstranění prostorových bariér). V současné době jsou v provozu projekty digitalizující různé archivy např. v Praze 6, Praze 2 a Praze 8.

Projekty digitalizování knihovních fondů patří k současným podporovaným trendům. Svědčí o tom mj. i tematický okruh FRVŠ, vypsaný právě v kategorii E tj. knihovny. Pokud se knihovna rozhodne pro tuto činnost, je třeba počítat s tím, že na splnění úkolu její vlastní síly s velkou pravděpodobností nebudou stačit. Převod

klasických fondů do digitalizované podoby a zejména jejich následné efektivní zpřístupnění je do značné míry otázkou technickou a je zapotřebí za tímto účelem získat spolupracovníky, mající na starosti správu a provoz počítačových sítí a informačního systému dané instituce. Míra schopnosti a ochoty ke vzájemnému dialogu a schopnost tvůrčí spolupráce se přímo promítne do celkového výsledku.

Tímto dochází k rozšíření a zkvalitnění nabídky elektronických publikací. Dochází k minimalizování časových ztrát uživatelů: uživatel se "obslouží" sám, může studovat ihned po vyhledání příslušného titulu v elektronickém katalogu,

nemusí žádat knihovníka o primární dokument, čekat na jeho expedici a po práci jej vracet. Zmenšují se časové ztráty a zefektivňuje se práce knihovníků: požadovaný primární dokument není třeba vyhledávat a expedovat ze skladiště a opět jej vracet na místo.

Díky technologii OCR lze převádět tištěné dokumenty do el. podoby, tisknout je např. jen po částech, editovat, archivovat, velmi rychle v nich vyhledávat. Další nespornou výhodou jsou malé fyzické nároky na skladovací prostory (HDD, ROM či flash média), prolomení lokačních bariér při shánění dokumentů, nabízí se také snadná možnost kopírování (což může být zároveň nevýhodou kvůli autorským právům). Dále snazší přístup k starším, vzácnějším a jinak těžko dostupným dokumentům (odpadá nutnost prezenčního půjčování), vznik elektronických archivů, knihoven, elektronického publikování, elektronických knih a obchodů. V neposlední řadě je také tato technologie využívána zrakově postiženými či nevidomými osobami, kteří používají navíc software, který umožňuje konverzi z elektronické podoby textu zvukové podoby. (Knotek, 1999)

3.2. Využití řečových technologií

Pomůcky pro handicapované osoby

Systémy TTS mají obrovský přínos především pro handicapované osoby, pomáhají jim se lépe začlenit do života dnešní společnosti. Pomocí speciálně upravených klávesnic mohou zapsat svoji řeč a nechat si ji systémem TTS převést do hlasové podoby. Profitovat ze systémů TTS mohou i nevidomí. Např. přístroje na

automatické čtení novin a knih jim umožní přístup k textovým informacím (Psutka, 2006, s. 631)

Aplikace s hlasovým výstupem

Syntetizéry řeči z textu je možné použít ve všech systémech, kde je požadován hlasový výstup (např. automatické čtení e-mailů, SMS, knih, atd.). Tímto se v současné době zabývá např. Katedra kybernetiky (KKY) na ČZU v Plzni. V rámci projektu "Vizuální syntéza češtiny metodou parametrického modelu jako doplněk řečového syntetizéru" napomáhá integraci handicapovaných lidí do běžného

života. Zdokonalují a testují software zobrazující "mluvící hlavu", která by mohla pomoci lidem nedoslýchavým (odezírání) i zrakově postiženým.

Telekomunikační služby

Telefonních přístrojů se dnes stále více využívá i k neinteraktivní komunikaci, kdy je volajícímu podána určitá informace z dialogového systému. Také je zde možné využití automatického rozpoznávání řeči a zadání dotazu přímo hlasem. Systém poté dotaz rozpozná a nalezne ve své databázi vhodnou odpověď, kterou předá na vstup systému TTS a ten jej reprodukuje.

Výuka jazyků

Řečové technologie najdou své využití také při výuce jazyků. Jejich použití v aplikacích pro výuku cizích jazyků může sloužit jako vhodný doplněk ke "klasickým" učebnicím. V budoucnu by mělo být možné učit se s výukovým systémem formou dialogu.

Automatické čtení textu

Systémy TTS jsou dále hojně využívány v celé řadě aplikací, kdy je potřeba převádět vstupní text na řeč. Tedy ve všech případech, kdy je vhodné mít k dispozici zvukový výstup. Příkladem mohou být různé automatické hlásiče (např. odjezdů vlaků na nádraží), automatické čtení knih, webových stránek a SMS zpráv.

Hlasové monitorování

Hlasové monitorování může sloužit jako doplněk při získávání informaci vizuální formou. Lze ho vhodně využít při monitorování v měřicích a řídících systémech například v podobě varovných hlášení v automobilech či letovém provozu.

Multimédia, komunikace člověk-počítač

Z hlediska budoucího vývoje komunikace člověka s počítačem hrají vedle automatického rozpoznávání mluvené řeči svou nezanedbatelnou roli i systémy TTS.

Je zřejmé, že pro kvalitní komunikaci se bez kvalitních systémů TTS nelze obejít. Již nyní se hlasové výstupy stávají standardní výbavou počítačů. (Psutka, 2006, s. 632)

Výzkum

Systémy TTS se vedle praktických aplikací také používají pro výzkumné účely. Určité typy systémů jsou využívány fonetiky, kteří díky nim mohou studovat dosud neobjasněné problémy, které souvisí se vznikem a šířením řeči v hlasovém traktu člověka. Vzhledem k tomu, že systémy text-to-speech dávají v konkrétních případech vždy stejné výsledky (na rozdíl od lidské řeči), slouží také lingvistům k experimentálním účelům. Své uplatnění najdou také v lékařství při léčení vad poslechu, hluchnutí nebo poruch vnímání, psychologii a psychiatrii.

Zábava

V neposlední řadě mohou být systémy pracující na bázi řečových technologií zakomponovány do produktů zábavního průmyslu. Konkrétním příkladem jsou různé dětské hračky, mluvící automaty a počítačové hry. (Psutka, 2006, s. 632)

4. Závěr

Nyní se nacházíte na konci bakalářské práce, ve které se autor snažil postihnout principy fungování, limity a význam technologií sloužících k digitalizaci a rozpoznávání textu a syntéze a analýze řeči.

Řečové technologie stejně jako technologie rozpoznávání znaků probíhají bouřlivým vývojem. Jsou dnes předmětem mnoha výzkumů a nepředpokládá se, že by v dohledné době došlo ke změně této situace.

Převádění dokumentů do elektronické podoby je dnes běžnou záležitostí v mnoha archivech, knihovnách, korporacích a různých institucí státního sektoru.

Technika OCR skrývá velký potenciál. Snaha získat upravitelný text s minimem chyb vede výrobce software k neustálému zdokonalování metody rozpoznávání znaků.

Jinak tomu není ani v oblasti zpracování řeči, přestože kvalita syntézy řeči zatím nebyla uspokojivě vyřešena a lidé dávají přednost lidskému hlasu před syntetickým, což je přirozené.

Je zřejmě jen otázkou času, kdy bude možno syntetizovat řeč, která bude k nerozeznání od přirozené lidské řeči. Potom teprve bude moci dojít k opravdovému rozšíření těchto systémů. Generování naprosto přirozené lidské řeči s sebou přináší i jistá bezpečnostní rizika. Pokud by k tomu došlo, lidstvo by muselo čelit potenciálnímu problému zneužití identity mluvčího.

Seznam použité literatury

1) AIM. *Optical Character Recognition (OCR).* Pittsburgh, 2000. Výzkumná zpráva. AIM, Inc. Dostupný také z WWW: <http://www.aimglobal.org/technologies/othertechnologies/ocr.pdf>.

2) BATŮŠEK, L.; NYGRÝN, P. D. Vyhoďte monitor, nastražte uši : jak naučit počítač mluvit. *Computer : počítačový čtrnáctideník.* 2000, roč. 12, č. 5, s.4-5. ISSN1210-8790.

3) Bennex. *Bennex* [online]. Bennex, c2007 [cit. 2009-07-29]. Podpora. Chci skenovat. Dostupný z WWW: <http://www.bennex.cz/podpora/skenovat/>.

4) Digitalizace dokumentů. *KTD : Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna České republiky, c2005- [cit. 2009-05-29]. Dostupný z WWW: <http://sigma.nkp.cz/F/HLAYPC7PRSP4VLDAPH7N2EBVTCV9PDS9LRSPDY72YYFEE6NHA7-33662?func=full-set-set&set_number=015240&set_entry=000001&format=999>.

5) DUTOIT, Thierry. *An introduction to text-to-speech synthesis.* 1st ed. Dordrecht : Kluwer Academic, 1997. xi, 285 s. ISBN 0-7923-4498-7.

6) HABIBALLA, Hashim. *Umělá inteligence*. 1. vyd. Ostrava : Ostravská Univerzita, 2004. 83 s. Učební texty Ostravské Univerzity. Dostupné také z WWW:<http://www.volny.cz/habiballa/publ/umint.pdf>.

7) CHAFE, W. L. Prosodic and Functional Units of Language. In *Talking Data: Transcription and Coding in Discourse Research.* Hillsdale, NJ: Lawrence Erlbaum, 1993, s. 33-43. ISBN 0805803491.

8) JANČAŘÍK, Miloslav. *Sběr a příprava vstupních dat metodou OCR*. 1. vyd. Praha : UVTEI, 1980. 41 s.

9) *Katedra kybernetiky (KKY)* [online]. Plzeň : ZČU, c2009 [cit. 2009-7-15]. Dostupný z WWW: <http://www.kky.zcu.cz/cz>.

10) KNOTEK, Pavel. *Velká kniha o skenování.* 1.vyd. Brno : Unis, 1999. 180 s. ISBN 80-86097-37-4.

11) KOPECKÝ, K.; NOCAR, D.; KOPECKÝ, R. OCR technologie v pedagogických disciplínách. *e-Pedagogium* (on-line), 2003, roč. 3, č. 3. [cit. 2009-06-30]. Dostupné také na WWW:

<http://pandora.idnes.cz/part/2006/9/22343/3/OCR_pedagog.doc>. ISSN 1213-7499.

12) KOPEČEK, Ivan. *Úvod do počítačového zpracování řeči : materiály k předmětu.* Brno : Masarykova Univerzita, 2009.

13) KOPEČEK, I.; POLITZER, M. Mluvíme s počítačem : svět rozpoznávání řeči. *Computer : počítačový čtrnáctideník.* 1999, roč. 11, č. 5, s. 4-8. ISSN1210-8790.

14) *Laboratoř počítačového zpracování řeči* [online]. Liberec : TUL, c2009 [cit.2009-08-12]. Dostupná z WWW: <https://www.ite.tul.cz/speechlab/>.

15) MAŘÍK V.; ŠTĚPÁNKOVÁ, O.; LAŽANSKÝ, O., aj. *Umělá inteligence.* (2). 1. vyd. Praha : Academia, 1997. 373. s. ISBN 80-200-0504-8.

16) MAŘÍK V.; ŠTĚPÁNKOVÁ, O.; LAŽANSKÝ, O., aj. *Umělá inteligence.* (5). 1. vyd. Praha : Academia, 2007. 544. s. ISBN 978-80-200-1470-2.

17) MATOUŠEK, Jindřich. Syntéza řeči. *In Portál ZČU* [online]. Plzeň : ZČU, c2009 [cit. 2009-07-22]. Dostupný z WWW: <http://docs.google.com/gview?a=v&q=cache:llx1Yp0Fe5IJ:portal.zcu.cz/wps/PA_Courseware/DownloadDokumentu%3Fid%3D5733+konkatenan>.

18) OCR. KTD : *Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna České republiky, c2005- [cit. 2009-05-29]. Dostupný z WWW: <http://sigma.nkp.cz/F/HLAYPC7PRSP4VLDAPH7N2EBVTCV9PDS9LRS

PDY72YYFEE6NHA7-04021?func=full-set-set&set_number=015858&set_entry=000002&format=999>.

19) Optical character recognition. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikimedia Foundation, 2001- , last modif. on 31 May 2009 [cit. 2009-06-01]. Anglická verze. Dostupné z WWW: <http://en.wikipedia.org/wiki/Optical_character_recognition>.

20) PECINOVSKÝ J., PECINOVSKÝ R. *Skenery a skenování.* 1. vyd. Praha, 1999. 121 s. ISBN 80-7169-8444-X.

21) PECINOVSKÝ, Josef. *Skenujeme na počítači.* 2. vyd. Praha : Grada, 2005. 84 s. ISBN 80-247-1244-X.

22) PSUTKA, Josef. *Mluvíme s počítačem česky.* 1. vyd. Praha : Academia, 2006. 746 s. ISBN 80-200-1309-1.

23) SIGMUND, Milan. *Speaker recognition : identifying people by their voices.* Brno : VUTIUM, 2000, 21 s. ISBN 80-214-1590-8.

24) SIGMUND, Milan. *Voice analysis and recognition.* Brno : VUTIUM, 2007. 24 s. ISBN 978-80-214-3396-0.

25) Skenování. *KTD : Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha : Národní knihovna České republiky, c2005- [cit. 2009-05-29]. Dostupný z WWW: <http://sigma.nkp.cz/F/HLAYPC7PRSP4VLDAPH7N2EBVTCV9PDS9LRSPDY72YYFEE6NHA7-22609?func=find-b&find_code=WTD&x=0&y=0&request=skenování&adjacent=N>.

26) Speech recognition. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikimedia Foundation, 2001- , last modif. on 3 June 2009 [cit. 2009-06-03]. Anglická verze. Dostupné z WWW: <http://en.wikipedia.org/wiki/Speech_recognition>.

27) Speech synthesis. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikimedia Foundation, 2001- , last modif. on 2 June 2009 [cit. 2009-06-03]. Anglická verze. Dostupné z WWW: <http://en.wikipedia.org/wiki/Speech_synthesis>.

28) *Speech Technology Magazine* [online]. Medford, NJ: Information Today, 2007- , last modif. on 1 June 2009 [cit. 2009-06-2]. Anglická verze. Dostupné z WWW: < http://www.speechtechmag.com/>.

29) UHLÍŘ, Jan. *Technologie hlasových komunikací*. 1. vyd. Praha : Nakladatelství ČVUT, 2007, 276 s. ISBN 978-80-01-03888-8.

30) VORLÍČEK, Jan. Zvuková uživatelská rozhraní. In *Braillnet* [online]. SONS, c1999 [cit. 2009-7-22]. Dostupný z WWW: <http://www.braillnet.cz/sons/docs/tl97/zvuk_rozhrani.html>.

Evidence výpůjček

Prohlášení:

Dávám svolení k půjčování této bakalářské práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

V Praze, 10. 8. 2009

Rostislav Koňařík

Jméno	Katedra / Pracoviště	Datum	Podpis