

AUTOMATICKÝ PŘEPIS MLUVENÉ ŘEČI NA MOBILNÍ TELEFON

Magdaléna NEKOLNÁ

NEWTON Technologies, a.s., Na Pankráci 127, Praha 4, magdalena.nekolna@newtontech.cz

***Anotace:** Pomocí programu Newton Dictate Mobile se řeč diktujícího přenesse přes internet na vzdálený server, který v reálném čase mluvené slovo převede do textu a pošle zpět na mobilní telefon. Uživatel tak má k dispozici mobilní zařízení pro rozpoznávání mluvené řeči. Podmínkou využívání této technologie je mobilní připojení na internet a osvojení si základních principů diktování.*

1. Automatický přepis mluvené češtiny

Rozvoj moderních hlasových technologií založených na počítačovém zpracování signálů a metodách umělé inteligence umožnil vznik softwarových nástrojů, které jsou schopny automaticky přepisovat mluvenou řeč. První programy umožňující spojitě diktování do počítače v češtině byly vyvinuty na Technické univerzitě v Liberci mezi lety 2003 až 2005. Trvalo však ještě několik let, než mohly být dovedeny do podoby komerčně dostupného softwaru.

Čeština patří mezi jazyky s velkou mírou ohebnosti. Podstatná jména, přídavná jména, zájmena a číslovky se skloňují, slovesa se časují. Při tomto procesu vznikají nové slovní tvary, jejichž existenci je třeba brát v úvahu při procesu rozpoznávání řeči. Pokud jsou obsaženy ve slovníku rozpoznávacího systému, je šance, že budou správně rozpoznány. V opačném případě, tedy když některý slovní tvar ve slovníku není, nemůže být nikdy správně rozpoznán a systém se při přepisu dopustí chyby. (Kvůli kontextu může tato chyba zasáhnout i okolní slova.)

Rozdíl mezi lexikálním inventářem angličtiny a češtiny si můžeme demonstrovat na příkladu v tabulce 1. Zatímco anglické slovo „driver“ se může vyskytnout pouze ve třech různých tvarech, v češtině stejnému slovu odpovídá několik desítek různých forem podstatného jména v mužském i ženském rodu, a dále přídavného jména, včetně přivlastňovacích tvarů.

Angličtina	Čeština
driver, drivers, driver's	řidič, řidiče, řidiči, řidičem, řidičů, řidičům, řidičích řidička, řidičky, řidiče, řidičku, řidičkou, řidičkám, řidičkách řidičský, řidičská, řidičské, řidičského, řidičskému, řidičským, řidičův, řidičova, řidičovu , řidiččin, řidiččina,

Tab. 1 – Porovnání angličtiny a češtiny z pohledu ohebnosti slov

Morfologie češtiny je natolik bohatá, že počet slovních tvarů vyskytujících se v psané a mluvené podobě dosahuje několika milionů, zatímco v angličtině se jedná spíše o desítky tisíc. Systém pro rozpoznávání češtiny musí proto pracovat s několikanásobně větším slovníkem, což samozřejmě komplikuje návrh algoritmů, zpomaluje jejich činnost a zvyšuje nároky na výpočetní výkon. Nejvýkonnější diktovací systém navržený pro češtinu nyní používá slovník obsahující cca 500 tisíc nejběžnějších slovních tvarů.

Dalším velkým problémem češtiny je relativně volné pořadí slov ve větě. Zatímco angličtina vyžaduje, aby věta měla pevnou strukturu (např. podmět – přísudek – předmět – příslovečné určení), v češtině lze pořadí slov často změnit, aniž by věta získala odlišný význam (Např. věta „Pavel potkal Evu“ může existovat v podobách „Evu potkal Pavel“ nebo „Potkal Pavel Evu“.) Tato velká volnost způsobuje, že metody rozpoznávání řeči založené na statistikách výskytu sousedních slov nepracují v češtině tak dobře jako v angličtině.

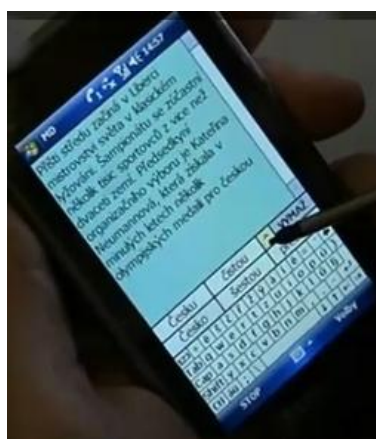
Nakonec zmíníme ještě jeden velký problém rozpoznávání řeči, který je ale společný pro většinu jazyků. Zatímco v písemné podobě děláme mezi jednotlivými slovy mezeru a díky tomu pak přesně vidíme, kde slovo začíná a končí, v mluvené podobě žádná hranice mezi slovy neexistuje. V běžné plynulé řeči totiž jedno slovo navazuje na druhé a pauzy děláme většinou jen na místech, kde se potřebujeme nadechnout, případně

rozmyslet. Jevy jako asimilace a koartikulace navíc výrazně zasahují do výsledné podoby promluvy. Rozdíl mezi psanou a mluvenou řečí demonstruje příklad v tabulce 2.

Psaná věta	Až se vrátíš, zavolej mi na linku pět šest sedm.
Mluvená věta (fonetický přepis)	ašsevrát'ízzavolej minalinkupječesedum

2. Program Newton Dictate Mobile

Od roku 2008 je zájemcům k dispozici také první český program umožňující diktování do počítače běžnou plynulou řečí. Vznikl na základě požadavků zejména soudců, právníků, lékařů a novinářů, tedy profesí, v nichž je potřeba vytvářet velké množství textových dokumentů. Do té doby si tyto osoby texty psaly samy, nebo je diktovaly asistentům či sekretářkám, případně je zaznamenávaly na diktafon a pak je nechávaly přepsat. Program, který je výsledkem spolupráce výzkumného týmu na Technické univerzitě v Liberci a firmy Newton Technologies a.s. a který proto nese název Newton Dictate, jim tuto práci významnou měrou usnadňuje a zefektivňuje.



Obr.1: Automatický přepis mluvené řeči pro mobily

Program Newton Dictate Mobile řeč diktujícího přenese přes internet na vzdálený server, který v reálném čase mluvené slovo převede do textu a během několika vteřin pošle zpět na mobilní telefon. K přepisu plynulé řeči využívá rozsáhlé slovníky s podporou statistického jazykového modelu. Úspěšnost přepisu závisí na složitosti textu (zejména na počtu méně frekventovaných slov a exotických vlastních jmen), dále na stylu diktování, výslovnosti a také na určitém cviku uživatele. Bez adaptace na konkrétní hlas lze běžně dosáhnout přesnost kolem 95 %. Pokud uživatel provede adaptaci na svůj hlas (pomocí dodávaného modulu), tato hodnota se zvyšuje obvykle o další 2 až 3 %. Program Newton Dictate Mobile umožňuje diktování prakticky libovolného tématu, přičemž u odborných textů jsou třeba speciální slovníky. Program pracuje se všeobecným slovníkem, který v současné době obsahuje kolem půl milionu nejfrekventovanějších českých slov. Pokud se člověk seznámí se všemi možnostmi programu a naučí se s ním efektivně pracovat, představuje pro něj diktování výraznou úsporu času a nákladů. Program má velký potenciál i pro hendikepované osoby.

Literatura

- [1] Jan Nouza, Petr Červa, Jindřich Žďánský: Programy pro diktování do počítače a možnosti jejich využití osobami se speciálními potřebami, článek bude uveřejněn v roce 2010 v časopisu Speciální pedagogika
- [2] Webové stránky společnosti NEWTON Technologies, a.s. www.diktovani.cz
- [3] Webové stránky Laboratoře počítačového zpracování řeči na Technické univerzitě v Liberci www.ite.tul.cz/speechlab/

PROJEKT DUHOVÝ MOST: NEPOMOHLA BY POLOVINA NEBE I VÁM?

Prof. Ing. Jan NOUZA, CSc., Ústav informačních technologií a elektroniky (ITE) Fakulta mechatroniky, informatiky a mezioborových studií (FMIMS) Technická univerzita v Liberci (TUL) Studentská 2, 461 17 Liberec 1, [jan.nouza@tul.cz]

Ing. Milan BOEHM, Megaflex, Osiková 3, 13000 Praha 3, [milan.boehm@megaflex.cz]

***Anotace:** Projekt Duhový most občanského sdružení Polovina nebe je zaměřen na výuku hlasového ovládní počítače osobami s kvadruplegií/kvadraparezou (či obdobným postižením), které by jinak nebyly schopny ovládat počítač úkony drobné motoriky. Tento tříletý projekt podporovaný EU a spolufinancovaný ESF má za cíl proškolit 50 osob z cílové skupiny, a díky tomu zvýšit jejich uplatnitelnost jak na pracovním trhu, tak v osobním životě.*

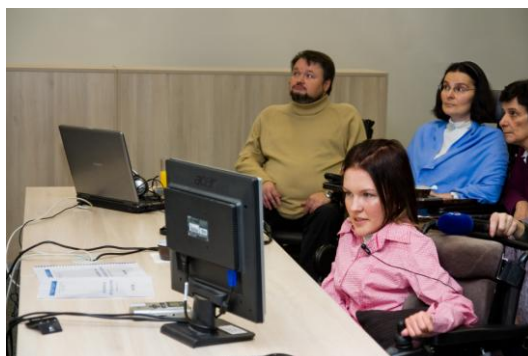
3. Projekt Duhový most

3.1 Ovládní počítače hlasem

Občanské sdružení Polovina nebe začíná v březnu tohoto roku realizovat projekt Duhový most, který je zaměřen na vzdělávání talentovaných jedinců, jimž tělesné postižení dosud bránilo ve větším uplatnění v životě i v práci. Projekt podporovaný Evropskou unií je součástí operačního programu Praha – Adaptabilita, který je spolufinancován ESF, a je registrován pod číslem CZ.2.17/2.1.00/32644.

Byl zahájen v roce 2010 a potrvá celkem 3 roky. Projekt je orientován zejména na skupinu osob s kvadruplegií/kvadraparezou případně paraplegií horních končetin nebo jejich těžkou paraparezou v takovém rozsahu, že nemohou provádět ani minimální úkony drobné motoriky. Tato zdravotní znevýhodnění se snažíme odstranit prostřednictvím nových hlasových technologií umožňujících ovládní počítače prostřednictvím hlasu klienta a dále technologií zaznamenávající diktovaná slova v písemné podobě. Projekt Duhový most řeší vytvoření tréninkové místnosti vybavené IT a hlasovými technologiemi, zajištění školení v používání hlasových technologií zdravotně znevýhodněnými a případně i nutné přizpůsobení hlasové technologie pro konkrétního klienta s dystoniemi - funkcionálními změnami hlasu.

Tato nová technologie – hlasová technologie – je dosud minimálně využívána v této oblasti a přitom může zásadním způsobem odstranit sociální izolaci zdravotně znevýhodněných osob přístupem na internet atd. Zdravotně znevýhodnění mohou po zaškolení bez pomoci pracovat na počítači pomocí svého hlasu a zvládnout práci, která by pro ně jinak byla prakticky nemožná. Zvládnutím těchto dovedností stoupne výrazně jejich uplatnitelnost na trhu práce a jejich zaměstnatelnost.



Obr. 1 Snímek zachycuje první zkušební trénink, který proběhl 9. listopadu 2009 v bezbariérové budově Gemini na pražské Pankráci za účasti médií a hlavně dvou prvních hendikepovaných žáků, které Dita školila. Byli jimi dva členové Asociace muskulárních dystrofií v ČR, pan Zdeněk Janda a paní Jitka Kačírková.