Univerzita Hradec Králové
Fakulta informatiky a managementu
Katedra informačních technologií

Analýza a využití TTS (Text-to-Speech) softwaru pro studijní účely

Bakalářská práce

Autor: **Vladimír Maliniak**
Studijní obor: **Informační management**

Vedoucí práce: **Ing. Tereza Otčenášková, BA**

Hradec Králové **28. srpna 2017**

Prohlášení:

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a s použitím uvedené literatury.

.....

V Hradci Králové dne 28. 08. 2017 Vladimír Maliniak

Poděkování:

Děkuji Ing. Tereze Otčenáškové, BA za metodické vedení práce a rady, které mi během psaní poskytla. Rád bych poděkoval společnosti Novasoft, a.s. za poskytnutý semistrukturovaný rozhovor, materiály a licencovanou verzi programu, kterou jsem následně využil v praktické části bakalářské práce. Dále bych rád poděkoval ostatním společnostem (Acapela Group, Linguatec a NaturalReader) za poskytnuté informace a užitečné odkazy, které jsem využil ve své bakalářské práci.

Anotace

Bakalářská práce se zabývá Text-to-Speech (TTS) softwarem a jeho možným využitím v dnešní společnosti. Cílem práce je poukázat na existenci TTS, informovat o principech jeho fungování a jeho možném využití v nejrůznějších oblastech.

Teoretická část práce pojednává o historii syntézy řeči, principech, kterými se provádí analýza a sběr dat potřebných pro tvorbu řeči, a která posléze vedou až k mluvené řeči. Dále jsou uvedeny funkcionality TTS softwaru, jeho podoby a také to, jak může přispět ke zvýšení kvality života lidí.

V praktické části práce je vybráno pět konkrétních TTS programů od různých softwarových firem. Jednotlivé programy jsou analyzovány a následně porovnány na základě zvolených kritérií, u nichž je stanovena jejich důležitost se zaměřením na využití ve vzdělávacích institucích. Výstupem tohoto porovnání je výběr konkrétního softwaru vhodného pro případnou implementaci.

Klíčová slova:
Analýza, Audio program, Speech-to-Text, Syntéza řeči, Text-to-Speech.

Annotation

Title: Analysis and Usage of TTS (Text-to-Speech) Software for Study Purposes

The bachelor thesis deals with the Text-to-Speech (TTS) software and its possible use in current society. The goal is to introduce the TTS, inform about its principles and potential use in various areas.

The theoretical part presents the history of speech synthesis, the principles of data analysis and collection which lead to spoken speech. Furthermore, the TTS

software functionalities are mentioned together with its forms and areas in which it might improve the quality of human life.

In the practical part, five TTS programmes developed by various companies are selected. These are analysed and consequently compared based on the chosen criteria. For each criteria, the importance is set with the focus on TTS use in the educational institutions. The output of this comparison is the selection of software appropriate for potential implementation.

Key words:
Analysis, Audio programme, Speech synthesis, Speech-to-Text, Text-to-Speech.

Obsah

1 Úvod ... 1

2 Cíl práce ... 2

3 Metodika zpracování ... 3

4 Teoretická část ... 5

4.1 Syntéza řeči ... 5

4.1.1 Historie syntézy řeči ... 6

4.1.1.1 Mechanické syntetizéry ... 6

4.1.1.2 Elektronické syntetizéry ... 8

4.1.1.3 Digitální syntetizéry ... 9

4.1.1.4 Historie české syntézy řeči ... 11

4.1.2 Základní přístupy syntézy ... 12

4.1.2.1 Artikulační syntéza ... 14

4.1.2.2 Formantová syntéza ... 14

4.1.2.3 Konkatenační syntéza ... 17

4.1.3 Druhy řečových syntetizérů ... 20

4.1.3.1 Text-to-Speech software ... 21

4.1.3.2 Speech-to-Text software ... 21

4.1.4 Využití řečových syntetizérů v praxi ... 21

4.1.5 Příklady českých a zahraničních systémů ... 23

5 Praktická část ... 25

5.1 Základní informace o použitých programech ... 25

5.1.1 Acapela Group ... 25

5.1.2 Microsoft ... 27

5.1.3 SpeechTech ... 29

5.1.4 Linguatec....31
5.1.5 NaturalReader....33
5.2 Kritéria hodnocení TTS softwaru....34
5.3 Hodnocení kritérií u konkrétních firem....40
5.3.1 Acapela Group....40
5.3.2 Microsoft....43
5.3.3 SpeechTech....45
5.3.4 Linguatec....47
5.3.5 NaturalReader....49
5.4 Porovnání jednotlivých kritérií TTS systémů....51
6 Shrnutí výsledků práce....61
7 Závěry a doporučení....64
8 Seznam použité literatury....66

Seznam obrázků

Obrázek č. 1 - Jednoduchý model Text-to-Speech syntézy. Zdroj: upraveno podle Lemmetty, (1999). 5
Obrázek č. 2 - Kratzensteinovy rezonátory. Zdroj: převzato z Lemmetty (1999). 6
Obrázek č. 3 - Mluvící stroj Wolfganga von Kempelena. Zdroj: upraveno podle Psutka, (2006). 7
Obrázek č. 4 - Blokové schéma Voderu. Zdroj: upraveno podle Psutka, (2006). 9
Obrázek č. 5 - Významné milníky ve vývoji řečové syntézy. Zdroj: upraveno podle Lemmetty, (1999). 11
Obrázek č. 6 - Zjednodušený model produkce řeči podle teorie zdroje a filtru. Zdroj: převzato z Psutka, (2006) 13
Obrázek č. 7 - Obecný model produkce řeči podle teorie zdroje a filtru. Zdroj: převzato z Psutka, (2006) 13
Obrázek č. 8 - Blokové schéma formantového syntetizéru založeného na pravidlech. Zdroj: upraveno podle Psutka, (2006). 16
Obrázek č. 9 - Základní struktura paralelního formantového syntetizéru. Zdroj: upraveno podle Lemmetty, (1999). 17
Obrázek č. 10 - Logo společnosti Expert Choice. Zdroj: AcapelaBox, (2017). 25
Obrázek č. 11 – Výstřižek z online demo verze Acapela TTS. Zdroj: Snímek obrazovky web-demo Acapela TTS. 26
Obrázek č. 12 - Logo společnosti Microsoft. Zdroj: Microsoft, (2017a). 27
Obrázek č. 13 - Logo Microsoft Office. Zdroj: Microsoft, (2017d). 27
Obrázek č. 14 - Výstřižek z programu Microsoft Office 365. Zdroj: Snímek obrazovky Office 365 – OneNote Online. 28
Obrázek č. 15 - Výstřižek z programu Microsoft Office 365. Zdroj: Snímek obrazovky Office 365 – OneNote – Imerzivní čtení 28
Obrázek č. 16 - Výstřižek z programu Microsoft Office Word 2016. Zdroj: Snímek obrazovky Microsoft Office – Word 2016 29
Obrázek č. 17 - Logo společnosti SpeechTech. Zdroj: SpeechTech, (2017a). 29
Obrázek č. 18 - Výstřižek z programu SpeechTech TTS verze 3.4.2. Zdroj: Snímek z obrazovky SpeechTech TTS verze 3.4.2 31

Obrázek č. 19 - Logo společnosti Linguatec. Zdroj: Linguatec, (2017b).....31
Obrázek č. 20 – Linguatec jazykově-technologický trojúhelník. Zdroj: Linguatec, (2017c).....32
Obrázek č. 21 - Výstřižek z programu Linguatec Voice Reader Studio Online Demo. Zdroj: Snímek obrazovky Voice Reader Studio Online Demo.....33
Obrázek č. 22 - Logo společnosti NaturalReader. Zdroj: NaturalReader, (2017b)...33
Obrázek č. 23 - Výstřižek z programu NaturalReader Free. Zdroj: Snímek obrazovky NaturalReader Free.....34
Obrázek č. 24 - Zadaná kritéria a hodnoty – Výstřižek z programu Expert Choice 2000. Zdroj: Vlastní zpracování. 61

Seznam grafů

Graf č. 1 – Graf srovnání TTS systémů z hlediska ceny. Zdroj: Vlastní zpracování. .52
Graf č. 2 - Graf srovnání TTS systémů z hlediska doplňků. Zdroj: Vlastní zpracování. 53
Graf č. 3 - Graf srovnání TTS systémů z hlediska jazykové rozmanitosti. Zdroj: Vlastní zpracování. 54
Graf č. 4 - Graf srovnání TTS systémů z hlediska lokalizace. Zdroj: Vlastní zpracování. 55
Graf č. 5 - Graf srovnání TTS systémů z hlediska možnosti úpravy hlasu. Zdroj: Vlastní zpracování. 56
Graf č. 6 - Graf srovnání TTS systémů z hlediska náročnosti na HW. Zdroj: Vlastní zpracování. 56
Graf č. 7 - Graf srovnání TTS systémů z hlediska uživatelské přívětivosti. Zdroj: Vlastní zpracování. 59
Graf č. 8 - Dynamický graf citlivosti – Výstřižek z programu Expert Choice 2000. Zdroj: Vlastní zpracování. 62
Graf č. 9 - Performační graf citlivosti – Výstřižek z programu Expert Choice 2000. Zdroj: Vlastní zpracování. 63

Seznam tabulek

Tabulka č. 1 - Kritéria s váhami. Zdroj: Vlastní zpracování. 35
Tabulka č. 2 – Kritéria TTS společnosti Acapela Group. Zdroj: Vlastní zpracování. 42
Tabulka č. 3 - Kritéria TTS společnosti Microsoft. Zdroj: Vlastní zpracování. 44
Tabulka č. 4 - Kritéria TTS společnosti SpeechTech. Zdroj: Vlastní zpracování..... 46
Tabulka č. 5 - Kritéria TTS společnosti Linguatec. Zdroj: Vlastní zpracování. 48
Tabulka č. 6 - Kritéria TTS společnosti NaturalReader. Zdroj: Vlastní zpracování. .. 50
Tabulka č. 7 - Tabulka srovnání TTS systémů z hlediska API. Zdroj: Vlastní zpracování. 51
Tabulka č. 8 - Tabulka srovnání TTS systémů z hlediska dostupnosti desktopové verze. Zdroj: Vlastní zpracování. 52
Tabulka č. 9 - Tabulka srovnání TTS systémů z hlediska frekvence aktualizace. Zdroj: Vlastní zpracování. 53
Tabulka č. 10 - Tabulka srovnání TTS systémů z hlediska helpdesk. Zdroj: Vlastní zpracování. 54
Tabulka č. 11 - Tabulka srovnání TTS systémů z hlediska plynulosti řeči. Zdroj: Vlastní zpracování. 57
Tabulka č. 12 - Tabulka srovnání TTS systémů z hlediska podpory OS. Zdroj: Vlastní zpracování. 57
Tabulka č. 13 - Tabulka srovnání TTS systémů z hlediska podpory textových souborů. Zdroj: Vlastní zpracování. 58
Tabulka č. 14 - Tabulka srovnání TTS systémů z hlediska převodu na zvukový soubor. Zdroj: Vlastní zpracování 58
Tabulka č. 15 - Tabulka srovnání TTS systémů z hlediska školení. Zdroj: Vlastní zpracování. 59
Tabulka č. 16 - Tabulka srovnání TTS systémů z hlediska verze pro mobilní zařízení. Zdroj: Vlastní zpracování. 60

1 Úvod

V dnešní době je využívání informačních technologií všední záležitostí pro většinu obyvatel. Je ovšem zvláštní, jak málo si mnozí z těchto lidí uvědomují potenciál Text-to-Speech (dále jen TTS, volně přeloženo jako převod textu do řeči) nástroji, kterému se tato bakalářská práce věnuje. Cílem je představení samotného tématu, informování o jeho podstatě a způsobu sběru dat včetně vytváření souvislé řeči a poukázání na zjednodušení, která to může přinést v životě mnoha lidí.

Bakalářská práce je rozdělena na část teoretickou a praktickou. V teoretické části bakalářské práce je částečně vysvětlena historie syntézy řeči, tedy to, jak vznikala v průběhu času. Dále jsou představeny existující přístupy pro zpracování řeči. Následně je vysvětlen základní rozdíl mezi TTS a STT (Speech-to-Text, volně přeloženo jako převod řeči na text) a ukázány jsou konkrétní příklady těchto druhů syntézy řeči. Na závěr teoretické části bakalářské práce jsou představeny oblasti, kde jsou tyto programy využívány. Zahrnuty jsou i příklady konkrétních programů.

V praktické části je vybráno pět konkrétních programů, které jsou otestovány a analyzovány. Dále je sestaven soubor jejich kladů a záporů se zaměřením na užití ve vzdělávacích institucích. S ohledem na zvolená kritéria je následně vybrán nejlépe vyhovující program, který je doporučen pro zkušební implementaci a praktické využití pro vzdělávací účely.

2 Cíl práce

Cílem této práce je informovat o existenci softwaru TTS, představit základní informace o tomto softwaru a jeho možném využití pro usnadnění mnoha lidských činností. Největší důraz je ovšem v této práci kladen na možná využití ve vzdělávacích institucích pro urychlení a zlepšení studia členů akademické obce včetně hendikepovaných osob. Na pěti vybraných TTS softwarech jsou sledovány aspekty, které nejlépe vyhovují pro plánované využití. Na závěr jsou data porovnána a pro následnou implementaci je vybrán konkrétní, nejlépe vyhovující, software z testovaných.

3 Metodika zpracování

Cílem této práce je informovat o existenci TTS a nalézt takový, který nejlépe vyhovuje pro implementaci ve vzdělávací instituci (například na Fakultě informatiky a managementu Univerzity Hradec Králové).

Pro splnění tohoto cíle byly použity různé zdroje. Největší zastoupení pro tvorbu teoretické práce mají tištěné zdroje (knihy) a digitální knihy (práce, články) dostupné v digitálních knihovnách a databázích i na internetu. Při tvorbě praktické části práce bylo využito převážně internetových (online) zdrojů. Většina firem, které jsou v práci analyzovány, má totiž dostupné informace na svých a partnerských webových stránkách.

Jednotlivé Text-to-Speech programy těchto firem byly testovány na základě pečlivě zvolených kritérií, které byly zkonzultovány s expertem Ing. Ladislavem Daníčkem ze společnosti CCNovasoft. Vybraným kritériím byla přiřazena důležitost pomocí vah (na škále od 1 do 10). Ta byla určena na základě autorových znalostí v návaznosti na konzultaci s expertem a na dostupné zdroje.

Na závěr byly jednotlivé TTS porovnány na základě stanovených kritérií a vah v programu Expert Choice. Tento program pro podporu rozhodování pomáhá rychleji a efektivněji rozhodovat. Zjednodušuje rozhodování nezávisle na odvětví, ve kterém se uživatel pohybuje. Program pracuje v reálném čase a kombinuje nejnovější (moderní) technologii s (časem otestovanou) matematikou tak, aby umožnil jednotné a dobře strukturované rozhodnutí.
Metodologie programu Expert Choice byla implementována na stovkách univerzit a přeložena do mnoha jazyků po celém světě.
Program Expert Choice je využíván například v těchto odvětvích:

- Vesmírné a obranné instituce
- Neziskové organizace
- Finančnictví
- Zdravotnictví
- Pojišťovnictví
- Investice a správa peněz

- Reality

(Expert Choice, 2017a).

Program Expert Choice byl vybrán pro svou jednoduchost a dostupnou školní licenci. Stejně tak poskytuje srozumitelné výstupy, které odpovídají potřebám této práce, Výsledným výstupem je nejlépe vyhovující TTS pro případnou implementaci.

Pozn.: Ceny uvedené v cizí měně (EUR a USD) jsou v celé práci kvůli srovnatelnosti převedeny podle kurzu EUR (26.085 Kč) a USD (22.098 Kč) na české koruny (Kurzy.cz, 2017).

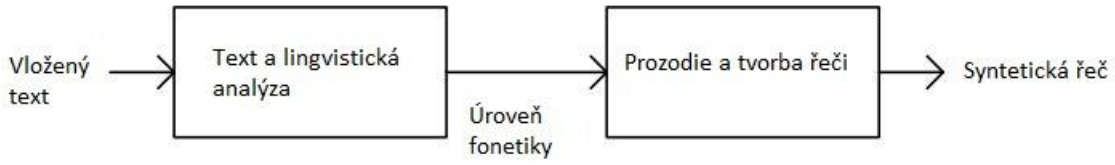
4 Teoretická část

Teoretická část práce je rozdělena do pěti hlavních podkapitol. V těchto podkapitolách je postupně vysvětlena syntéza řeči od historie přes přístupy, až po využití TTS v dnešní společnosti.

4.1 Syntéza řeči

Syntéza řeči představuje velmi důležitou oblast problematiky pro zpracování řečového signálu. V tomto případě se jedná o počítačovou syntézu, kde je řeč uměle vytvářena počítačem. Hlavním cílem počítačové syntézy je „zpřirozenit" komunikaci člověka s počítačem. Pro člověka je totiž řeč zcela přirozeným způsobem komunikace a syntéza řeči si klade za cíl vytvořit řeč v kvalitě a formě tak, aby byla zcela nerozeznatelná od běžné řeči člověka (Psutka, 2006).

Informace vytvořená mluvením může být analyzována mnoha způsoby. Syntéza řeči totiž pracuje na mnoha úrovních, mezi které patří akustická, fonetická, fonologická, morfologická, syntaktická, sémantická nebo pragmatická úroveň. Jak je zobrazeno na Obrázku č. 1, tyto úrovně jsou velmi příbuzné transformacím potřebným k sestavení mostu mezi dvěma částmi, mezi řečí a textem (Dutoit, 1997).



Obrázek č. 1 - Jednoduchý model Text-to-Speech syntézy. Zdroj: upraveno podle Lemmetty, (1999).

Mezi nejjednodušší cestu produkce syntézy řeči je nahrát části reálné řeči, jako jsou například jednotlivá slova nebo i celé věty. Jedná se o konkatenační metodu (viz popis v samostatné kapitole níže), která tímto způsobem zaručuje velmi vysokou kvalitu a přirozenost. Bohužel je značně omezena velikostí nahraného slovníku, který máme k dispozici. Stejně tak je téměř nemožné zachytit každé slovo, každý název, který na světě existuje a zahrnout jej do databáze. To je také důvod, proč je potřeba používat kratší části řečového signálu,

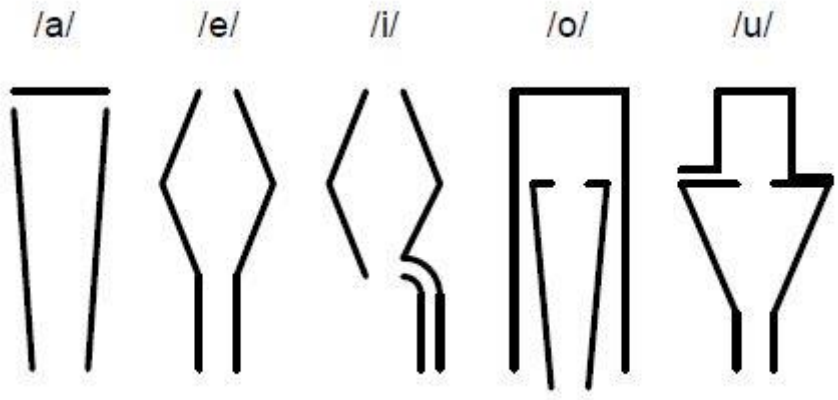
jako jsou například slabiky (nejmenší organizační jednotky řeči), fonémy (nejmenší jednotka zvuku) (CollinsDictionary.com, 2017c), diphony (jednotka řeči, která je vytvořena ze dvou jednoduchých řečových zvuků známé jako fonémy) (CollinsDictionary.com, 2017a) a jiné (Lemmetty, 1999).

4.1.1 Historie syntézy řeči

Lidská schopnost mluvit je v našem životě nesmírně důležitá a je to hlavní komunikační prostředek, který používáme. Není divu, že již od pradávna naše předky komunikace zajímala a studovali ji a zkoumali. S příchodem technologií se postupně začaly objevovat první pokusy o vytváření řeči umělé (Psutka, 2006).

4.1.1.1 Mechanické syntetizéry

O první pokus o umělé vytvoření řeči se zasloužil v roce 1779 německý profesor Christian Kratzenstein (Psutka, 2006). Vysvětlil fyziologické rozdíly mezi pěti dlouhými samohláskami (a, e, i, o, u) a sestavil zařízení, které je dokázalo poměrně dobře produkovat. Vytvořil akustické resonátory podobné lidskému řečovému ústrojí, které bylo možné aktivovat za pomoci vibrací, podobně jako u běžně známých hudebních nástrojů. Na Obrázku č. 2 níže je zobrazena základní struktura rezonátorů, které produkovaly zvuk při foukání vzduchu do dolní části trubky, čímž vznikal zvuk podobný flétně (Lemmetty, 1999).



Obrázek č. 2 - Kratzensteinovy rezonátory. Zdroj: převzato z Lemmetty (1999).

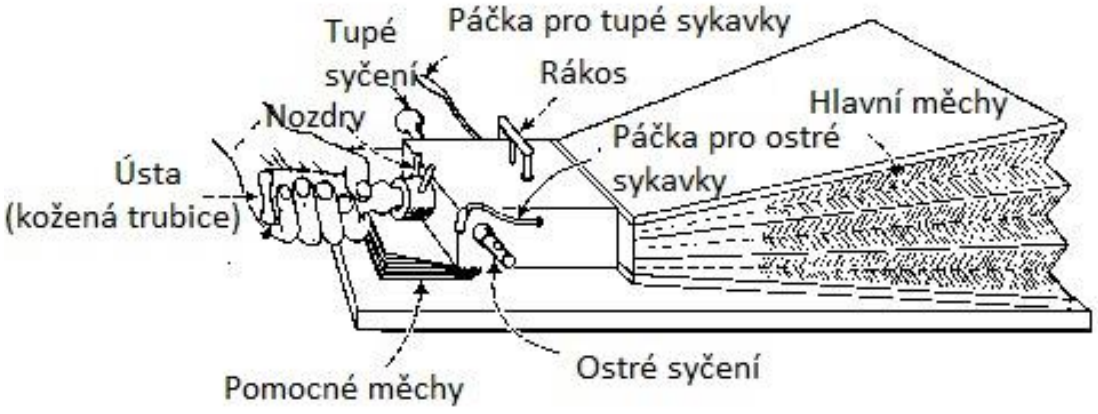
O několik let později (v roce 1791) byl maďarským šlechticem Wolfgangem von Kemplenenem představen zcela první "syntetizér řeči" (Psutka, 2006).

Ve skutečnosti ovšem von Kempelen začal pracovat na svém přístroji mnohem dříve než Kratzenstein, již v roce 1769. Přibližně po dvaceti letech svého výzkumu vydal knihu, ve které vysvětloval svá studia lidské produkce řeči a experimenty s přístrojem řečové syntézy (Lemmetty, 1999). O něco lepší, zdokonalenou verzi mluvícího zařízení, sestrojil Charles Wheatstone (viz Obrázek č. 3) (Psutka, 2006).

Tento přístroj se skládal „z hlavních a pomocných měchů, rákosu, pryže, kožené trubice a malých trubiček. Hlavní měchy simulovaly funkci plic, rákos činnost hlasivek, ústa se modelovala pomocí pryže. Kožená trubice přestavovala rezonanční prostor. Malé trubičky plnily funkci nozder. Jednotlivé zvuky se vytvářely modifikací budícího zvuku v rezonanční trubici." (Psutka, 2006, 533) Zařízení se ovládalo za pomocí páček a údajně dokázalo generovat primitivní slova a věty (Psutka, 2006).

V pozdější době, konkrétně okolo roku 1800 Alexander Graham Bell se svým otcem (inspirováni Wheatstonovým "mluvícím" zařízením) vytvořili velmi podobné "mluvící" zařízení (Lemmetty, 1999).

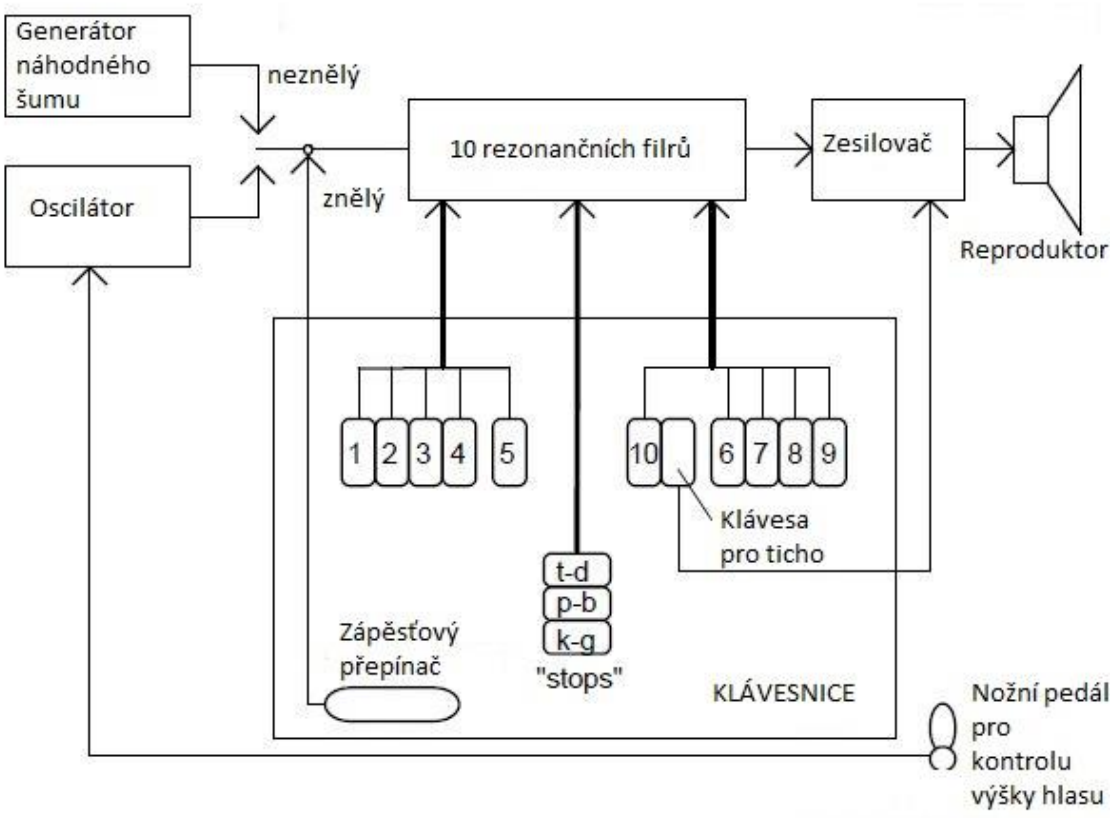
Podobné experimenty s mechanickými i polo-elektronickými analogickými a řečovými systémy pokračovaly nadále, avšak do roku 1960 nebylo dosáhnuto žádného významného objevu (Lemmetty, 1999).



Obrázek č. 3 - Mluvící stroj Wolfganga von Kempelena. Zdroj: upraveno podle Psutka, (2006).

4.1.1.2 Elektronické syntetizéry

V pozdější době, konkrétně ve 20. letech 20. století se zvyšoval zájem vědců o elektronické vytváření hlasu. Jedno z prvních zařízení představil v roce 1922 J. Q. Stewart. Tento přístroj využíval "bzučák jako buzení dvou rezonančních obvodů, které modelovaly první dva formanty hlasového traktu. Zařízení bylo schopné vytvářet jednotlivé samohláskové zvuky." (Psutka, 2006, 533) Zmíněné formanty lze v akustice a fonetice charakterizovat jako jakékoliv skupiny frekvencí, které se podílejí na zvuku, konkrétně na samohláskách ve svém maximu. (CollinsDictionary.com, 2017b) Ovšem později, v roce 1932, přidali japonští vědci Obata a Teshima třetí formant a tím dokázali zvýšit srozumitelnost syntetizéru (generovaných samohlásek). Opravdu první elektronický syntetizér schopný vytvářet souvislou řeč, Voder (Voice Operating Demonstrator, volně přeloženo do češtiny jako hlasový operační demonstrátor), uvedl v roce 1939 Homer Dudley. Byl složen "z deseti paralelně zapojených pásmových propustí, které byly buzeny buď periodickým signálem (v případě znělých zvuků), jehož perioda byla nastavována nožním pedálem, nebo v případě neznělých zvuků šumem." (Psutka, 2006, 533) (viz Obrázek č. 4).



Obrázek č. 4 - Blokové schéma Voderu. Zdroj: upraveno podle Psutka, (2006).

Po představení Voderu se začal vědecký svět zajímat o řečovou syntézu mnohem více než kdy předtím. Hlavním důvodem bylo, že díky Voderu se ukázalo, že je možné vyprodukovat opravdu srozumitelný zvuk. Momentálně je základní struktura a nápad Voderu velice podobný současnému systému, který je založen na "zdroj-filter" (podrobněji popsáno v kapitole níže) modelu řeči (Lemmetty, 1999).

4.1.1.3 Digitální syntetizéry

S nástupem číselných počítačů nastal v polovině šedesátých let zlom ve vývoji syntézy řeči, protože došlo k zásadní změně celého přístupu. Začaly se postupně objevovat úspěšné verze syntetizérů PAT (Parametric Artificial Talker, volně přeloženo do češtiny jako parametrický umělý řečník) a OVE (Orator Verbis Electris, volně přeloženo do češtiny jako řečník vybraných slov) (Psutka, 2006). PAT a OVE syntetizéry spustily konverzaci o tom, jak by správně měla být modelována akustická trubka, zda paralelně nebo kaskádově (Lemmetty, 1999). V roce 1968 Rex Dixon a David Maxey vytvořili syntetizér řetězící řečové jednotky na úrovni difonů

parametrizovaných pomocí formantových frekvencí, což lze označit za první výskyt konkatenační syntézy (Psutka, 2006).

Paralelně s tímto vývojem se vyvíjí syntézy řeči z textu (tedy TTS). Toto lze považovat za velmi významný krok ve vývoji, neboť TTS se soustředil na vytváření přirozené řeči z libovolného textu (Psutka, 2006).

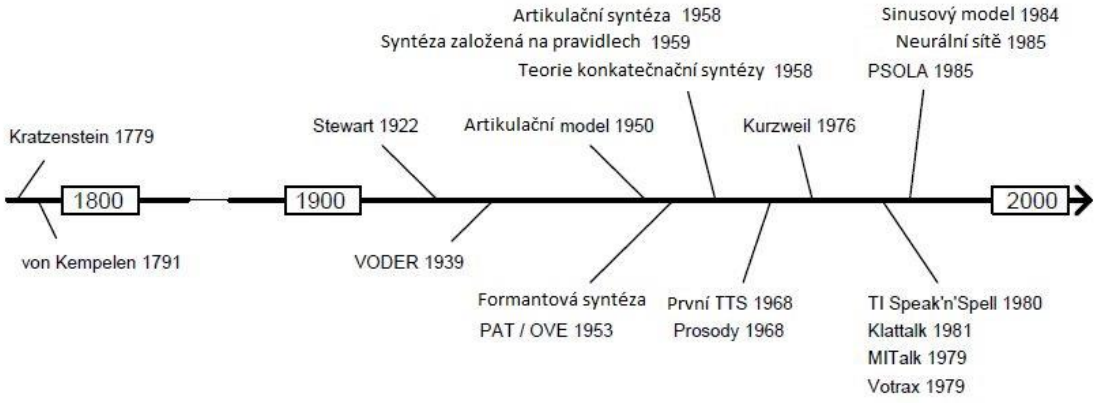
V roce 1968 byl vyvinut první systém TTS Norikou Umeda v Japonsku. Jednalo se zejména o generování řeči na základě artikulačního modelu, který samozřejmě nedosahoval úrovně, jakou známe u podobných TTS dnes (Psutka, 2006).

Jako významný pokrok bylo v roce 1976 vytvoření přístroje pro čtení tištěných stránek. O ten se zasloužil Raymond Kurzweil. Jednalo se o jedno-účelný počítač, který měl pomáhat nevidomým osobám, které v tehdejší době byly odkázány pouze na Braillovo písmo. Hlavním problémem byla značná nedostupnost z důvodu vysokých finančních nákladů, které dosahovaly 30 000 dolarů. I přes nemalou částku byl přístroj dostupný ve veřejných prostorech, jakými jsou například knihovny (Psutka, 2006).

Dále byly vyvíjeny systémy založené zejména na formantovém systému, jako TTS MITalk (systém převodu textu na řeč). Tento systém byl později použitý v Telesensory Systems Inc. (Lemmetty, 1999) a vylepšená verze Klattalk, která je ještě dnes základem mnoha prodávaných systémů (Psutka, 2006).

V 80. letech se rozmohly komerční systémy TTS (Psutka, 2006). Prvním integrovaným čipem vytvořeným pro řečovou syntézu byl Votrax, který obsahoval kaskádový formant syntetizér a jiné jednoduché čipy (Lemmetty, 1999). Mezi velmi známými lze uvést například DECtalk od společnosti Digital Equipment Corporation, který byl na trh uveden v roce 1983 (Psutka, 2006). V roce 1978 byl Richardem Gagnonem představen velmi levný (měl opravdu výhodnou cenu) systém Type-n-Talk, který byl založený na čipu Votrax. Mezi další poté patřil levný syntetizér "Speak-n-Spell“, který byl vyvinut o dva roky později (v roce 1980) firmou Texas Instruments. Tento nový systém se převážně užíval pro elektronické předčítání dětem, díky čemuž si získal nemalou pozornost. V roce 1982 byl firmou Street Electronics představen nízko-nákladový diphone (česky difon) syntetizér, který byl založen na novější verzi stejného čipu jako ten, jenž byl obsažen v systému

Speak-n-Spell. Ve stejnou dobu byl představen také Prose-2000 TTS systém (Lemmetty, 1999).



Obrázek č. 5 - Významné milníky ve vývoji řečové syntézy. Zdroj: upraveno podle Lemmetty, (1999).

S postupem vývoje počítačů docházelo i k vývoji konkatenační syntézy (viz Obrázek č. 5). Hlavním zlomem byla technika PSOLA (anglicky Pitch Synchronous Overlap and Add, volně přeloženo do češtiny jako fázový vokodér pracující
ve frekvenční oblasti), která je "metoda pro poměrně snadnou modifikaci prozodických charakteristik řetězených řečových jednotek." (Psutka, 2006, 536) Pro upřesnění je prozodie nauka zabývající se zvukovými vlastnostmi verše. (Goldmanová, 2015).

Dnešní systémy používají složité metody a algoritmy. V 90. letech 20. století dochází k nástupu korpusově orientované řeči, která je v dnešní době jednou z nejvyužívanějších metod (Psutka, 2006).

4.1.1.4 Historie české syntézy řeči

Ani v Čechách se, co se týče vývoje TTS, nezaostávalo. Mezi zcela první pokusy o vytvoření samohláskového systému se zasloužil František Kaňka. První český syntetizér řeči sestavil Přemysl Janota z Ústavu fonetiky Univerzity Karlovy v Praze v roce 1964. O rok později vytvořili Blanka Borovičková a Vlastislav Maláč dokonalejší český syntetizér **OVED 1** (první český syntetizér souvislé řeči) (Psutka,

2006). Dále jistě stojí za zmínku systémy **HO2** (první český TTS systém, 1977), **PCVOX** (první český LPC – Linear Predictive Coding, volně přeloženo lineárně prediktivní kódování – TTS systém, 1990), **Epos** (první český open source - s otevřeným zdrojovým kódem – TTS systém, 1996) a **ARTIC** (první český Unit Selection TTS, 2000) (Horák, 2009).

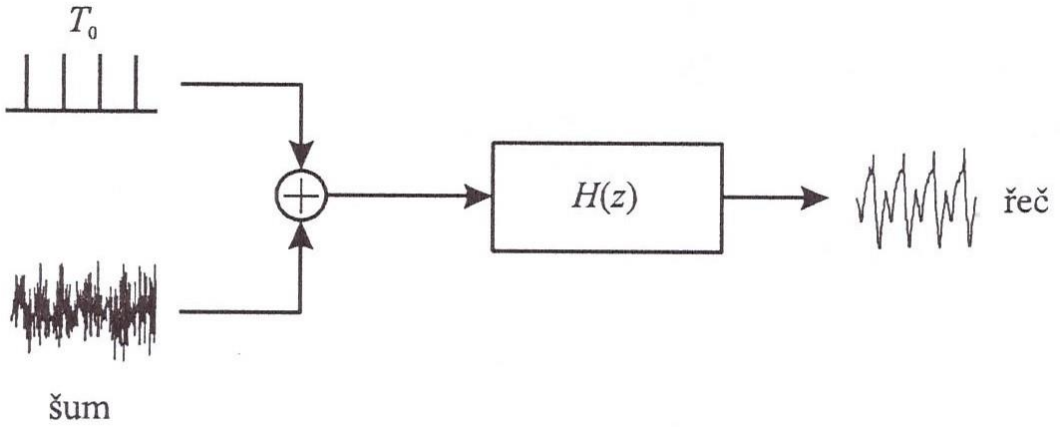
4.1.2 Základní přístupy syntézy

Přístupy se dělí na základě způsobu modelu použitého pro vytvoření výsledné řeči na artikulační, formantovou a konkatenační syntézu. Artikulační syntéza má za cíl modelování celého systému vytváření řeči, zatímco formantová a konkatenační se spíše soustředí na vlastní řečový signál a jeho modelování, místo snahy o přesnou imitaci hlasového ústrojí člověka (Psutka, 2006).

Teorie zdroje a filtru

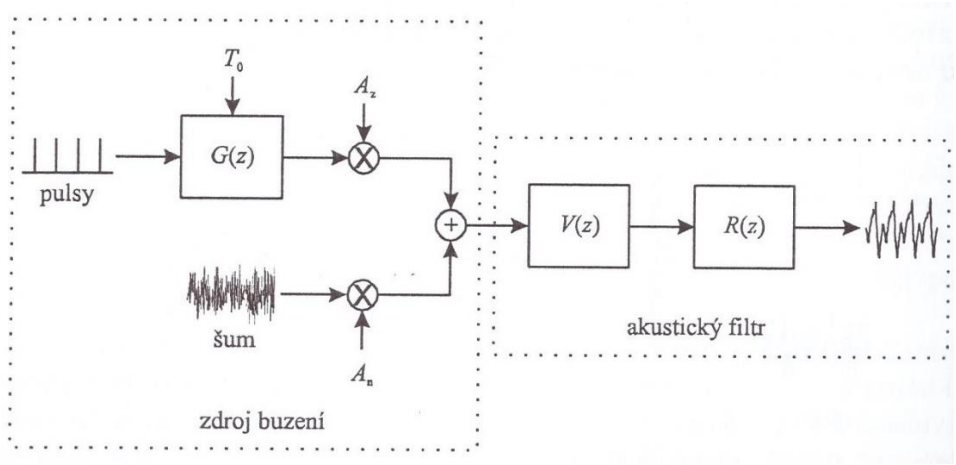
Tato teorie je také nazývaná akustická teorie utváření řeči. Vychází z akustického modelu produkce řeči, kterou vytvořil Gunnar Fant a akustické teorie artikulace samohlásek, kterou vytvořil Gerold Ungeheuer. Tento přístup popisuje v současné době nejpoužívanější model produkce řeči (Psutka, 2006).

Řeč je vytvářena dvěma nezávislými složkami: **Zdrojem buzení** a **lineárním akustickým** filtrem, který má reprezentovat frekvenční odezvu hlasového traktu. Model produkce řeči je znázorněn na Obrázku č. 6 a Obrázku č. 7. Sled hlasivkových pulsů (**hlasivkový signál**) je generován zdrojem buzení (Psutka, 2006).



Obrázek č. 6 - Zjednodušený model produkce řeči podle teorie zdroje a filtru. Zdroj: převzato z Psutka, (2006).

Podle teorie zdroje a filtru lze řečový signál po z-transformaci zapsat následovně: S(z) = H(z)U(z), kde U(z) je aproximace budícího signálu a H(z) = představuje přenosovou funkci akustického filtru, který reprezentuje radiační charakteristiku rtů a odezvu hlasového traktu (Psutka, 2006).



G(z) = přenosová funkce modelu hlasivek

Az = zesílení zdroje znělého buzení

An = zesílení zdroje neznělého buzení

V(z) = přenosová funkce hlasového traktu

R(z) = radiační charakteristiky rtů

Obrázek č. 7 - Obecný model produkce řeči podle teorie zdroje a filtru. Zdroj: převzato z Psutka, (2006).

4.1.2.1 Artikulační syntéza

Artikulační syntéza je jedním ze tří základních přístupů syntézy řeči. Na rozdíl od druhých dvou přístupů se artikulační syntéza více snaží věnovat procesu vytváření řeči. Artikulační syntéza ve své podstatě modeluje přímo systém vytváření řeči, jako u člověka včetně jeho fyzikálního chování. Tento model produkce řeči zahrnuje modely artikulátorů (čelisti, rty a podobně) a model hlasivek. Vytváření signálu se docílí simulací výdechového proudu vzduchu a činnost artikulačních modelů, s tím, že se klade důraz na fyzikální vlastnosti artikulátorů (Psutka, 2006).

Při mluvení svaly řečového ústrojí způsobují pohyb artikulátorů a změnu tvaru řečového ústrojí, čímž dochází ke změnám vytvářeného zvuku. Data u artikulačního modelu jsou nejčastěji odvozena z X-ray analýzy přirozené řeči. Hlavním problémem je, že tato data jsou dostupná pouze ve formě 2D, kdežto skutečný lidský řečový systém je přirozeně ve 3D. Pravidlově založené artikulační syntézy je tedy opravdu složité upravit a optimalizovat z důvodu nedostatku potřebných dat artikulátorů během řeči. Velkým nedostatkem artikulační syntézy je, že X-ray správným způsobem nedokáže charakterizovat množství a stupně volnosti artikulátorů (Lemmetty, 1999).

Jedná se vskutku o velmi zajímavou metodu pro syntézu řeči, jejíž snahou je přímo modelovat produkci řeči tak, jako je u člověka. Největším problémem je ovšem složitost takového modelu (výpočetní náročnost, náročná syntéza). V dnešní době bohužel ještě výsledky artikulační syntézy nedosahují takových kvalit jako formantová nebo konkatenační syntéza, což se samozřejmě může postupem času změnit. Zůstává tedy otázkou, zda bude mít tato snaha imitace lidského řečového ústrojí v budoucnu pozitivní výsledky (Psutka, 2006).

4.1.2.2 Formantová syntéza

Formantová syntéza je poměrně úspěšnou metodou v oblasti syntézy řeči, která je založena na takzvané teorii zdroje a filtru – imituje proces vytváření řeči jako u člověka (Psutka, 2006). Jedná se zřejmě o nejvíce používanou metodu posledních desítek let. Nejvíce oblíbená je formantová syntéza založená

na pravidlech. Existují dvě základní struktury, paralelní a kaskádová, avšak velmi často je nejlepší použití kombinace obou dvou. Formantová syntéza poskytuje téměř nekonečné množství zvuků, díky čemuž je opravdu flexibilním systémem, než je například konkatenační syntéza (Lemmetty, 1999).

Formantová syntéza se skládá z následujících částí:

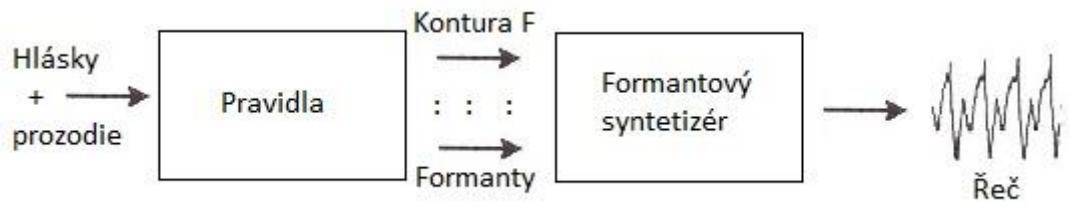
- Zdroj buzení – dále se dělící na zdroj znělých zvuků (simulující vibraci hlasivek) a zdroj neznělých zvuků (vytvářející šum).
- Hlasový trakt – modelován za použití filtru, který modeluje akustické vlastnosti řeči (Psutka, 2006).

Pro vyprodukování srozumitelného zvuku je zapotřebí alespoň tří formantů a pro vyprodukování vysoké kvality zvuku až pěti formantů. Každý formant je modelován s dvoupólovým resonátorem, který mu umožňuje jak specifikování frekvence formantů, tak druh přenosu (Lemmetty, 1999).

Formantová syntéza založená na pravidlech je postavena na množině pravidel, která určují konkrétní parametry pro syntézu. Vstupní parametry mohou být, jak uvádí ve své práci Lemmetty (1999), následující:

· „Voicing fundamental frequency (F0) *(volně přeloženo jako základní hlasová frekvence)*
· Voiced excitation open quotient (OQ) *(volně přeloženo jako hlasové (znělé) dráždění otevřeného kvocientu)*
· Degree of voicing in excitation (VO) *(volně přeloženo jako stupeň hlasového (znělého) dráždění)*
· Formant frequencies and amplitudes (F1...F3 and A1...A3) *(volně přeloženo jako formantové frekvence a amplitudy)*
· Frequency of an additional low-frequency resonator (FN) *(volně přeloženo jako frekvence dalšího (přídavného) nízkofrekvenčního rezonátoru)*
· Intensity of low – and high-frequency region (ALF, AHF) *(volně přeloženo jako intenzita nízké a vysoké frekvence)* " (Lemmetty, 1999).

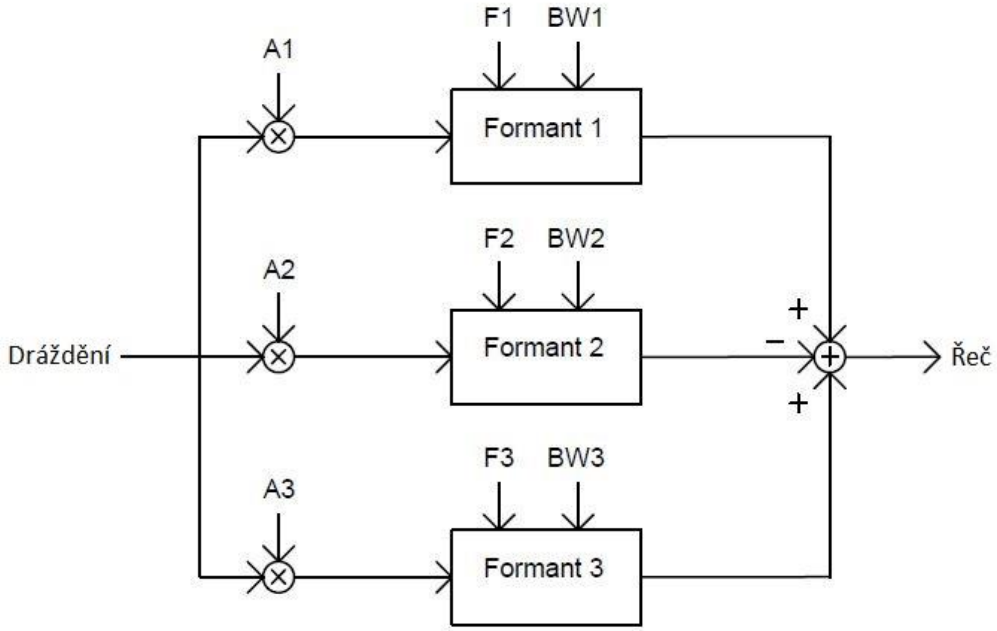
Jak uvádí Psutka (2006, 538): "Parametry syntetizéru úzce souvisí zejména s formantovými charakteristikami řeči (frekvencemi a šířkami pásem). Formantová syntéza bývá považována za typického zástupce syntézy podle pravidel." Na Obrázku č. 8 lze vidět ukázku blokového schématu formantového syntetizéru podle pravidel (Psutka, 2006).



Obrázek č. 8 - Blokové schéma formantového syntetizéru založeného na pravidlech. Zdroj: upraveno podle Psutka, (2006).

Kaskádová struktura je považována za lepší pro vytváření jednodušších zvuků a také proto, že vyžaduje mnohem nižší kontrolu informací než struktura paralelní (Lemmetty, 1999).

Paralelní formantový syntetizér se skládá z rezonátorů, které jsou zapojené paralelně. Vyprodukovaný signál je aplikován na všech formantech současně a na jejich výstupu opět dochází k jejich součtu (sumaci). U přilehlých výstupů formantových resonátoru se musí sečíst (dát do výstupního jednoho signálu) v opačném pořadí, aby se přecházelo nechtěným problémům. Struktura paralelního formantového syntetizéru kontroluje proud informací každého formantu zvlášť, čímž je zapotřebí více kontrolních informací (podrobněji zobrazeno na Obrázku č. 9) (Lemmetty, 1999).



Obrázek č. 9 - Základní struktura paralelního formantového syntetizéru. Zdroj: upraveno podle Lemmetty, (1999).

4.1.2.3 Konkatenační syntéza

Na rozdíl od jiných metod je konkatenační metoda více náročná na velikost paměti, jelikož je potřeba uchovat nahranou řeč (slova, věty, spojky a podobně). Nevýhodou je, že je většinou zapotřebí užití právě jednoho řečníka. Výhoda ovšem spočívá v tom, že výsledný zvuk je mnohem kvalitnější a plynulejší (Lemmetty, 1999).

Hlavním problémem v konkatenační syntéze je najít správnou jednotku délky. Výběr je většinou mezi jednotkou delší a kratší. Přirozenějšího hlasu lze dosáhnout použitím delší jednotky. V tomto případě však dochází i k vyššímu využití paměti. U kratší jednotky se nevyužívá takové množství paměti, ovšem výsledný zvuk není tak přirozený a sběrné a označovací procedury jsou mnohem složitější (Lemmetty, 1999).

V současné době má konkatenační syntéza jednoznačně hlavní (dominantní) postavení mezi všemi přístupy k syntéze řeči. Konkatenační syntéza je nejpoužívanější technikou k vytváření řeči. **Hlavními výhodami konkatenační syntézy,** jak uvádí Psutka (2006), jsou:

- **Používání přirozených segmentů řeči** – Konkatenační syntéza využívá segmenty skutečného řečového signálu, takže není potřeba je složitě vytvářet za pomocí navrhnutých pravidel. Výhodou je přirozenější vytvořená řeč.
- **Redukovaná znalost procesu vytváření řeči** – Na rozdíl od formantové syntézy odpadá potřeba detailní znalosti vytváření řeči člověkem. Pracujeme pouze s již vytvořenými segmenty přirozeného jazyka.
- **Rychlý návrh syntetizéru** – Na rozdíl od formantové syntézy je tento způsob (konkatenační syntéza) časově méně náročný. Odpadá zde nutnost hledání složitých pravidel, většina parametrů je totiž nastavena syntetizérem automaticky ze skutečných dat. Časově nejnáročnějším procesem v konkatenační syntéze tedy zůstává samotný sběr dat (tzn. namluvení řečového korpusu).
- **Kopírování hlasu řečníka** – V konkatenační syntéze dochází ke kopírování hlasu řečníka, čímž není potřeba umělého vytváření požadovaného hlasu (toho, jímž mluví syntetizér). Důležité tedy je mít k dispozici dostatečný počet namluvených vět řečníkem s požadovaným hlasem.
- **Vysoká kvalita syntetické řeči** – Konkatenační syntéza má obecně mnohem vyšší kvalitu syntetické řeči než ostatní druhy syntézy.
- **Transformace hlasu** – Závislost na konkrétním řečníku může být chápána jako nevýhoda. Konkatenační syntéza ze svého principu téměř neumožňuje syntetizování řeči jiným hlasem než tím, kterým byly namluveny jednotky v databázi.
- **Nebezpečí špatné kvality** – V případě korpusově orientované syntézy se může stát, že jisté slovo, nebo slabika bude znít velmi špatně. Je nutné toto nebezpečí redukovat pečlivým výběrem vět pro nahrávání. Problém může ovšem nastat u zřídka používaného slovního spojení.
- **Místa řetězení** – Největším nedostatkem konkatenační syntézy je její vlastní princip – nutnost řetězit jednotky, a to i přes to, že byly vyvinuty rychlé a efektivní algoritmy řetězení jednotek. Naštěstí lze tyto problémy redukovat stále více používanou korpusově orientovanou syntézou a on-line výběrem těch nejvhodnějších realizací řečových jednotek.

Korpusově orientovaná syntéza (Corpus-Based Synthesis)

Tento typ syntézy je v posledních letech hojně využívanou metodou, zejména díky neustále se zvětšujícímu výkonu moderních počítačů. Tato metoda syntézy řeči využívá rozsáhlé řečové korpusy (soubor počítačově uložených textů), které jsou segmentované až na úroveň hlásek a důkladně foneticky popsané. Každá operace během syntézy řeči probíhá naprosto automaticky na základě zvoleného řečového korpusu – kvalita je závislá na kvalitě korpusu. Velmi často se v současné době využívají obrovské řečové korpusy (až stovky hodin řeči), obsahující každou abstraktní řečovou jednotku v různých fonetických, spektrální i prozodických realizacích. Pro korpusově orientovanou syntézu je typické, že vedle uniformních řečových jednotek (difony nebo hlásky) pracuje i s neuniformními jednotkami (jednotky různé délky). V čase syntézy se tedy prohledá řečový korpus a hledá se optimální posloupnost řečových jednotek různého typu a jejich nejlepší realizace. V současné době existuje mnoho efektivních algoritmů, které využívají korpus k syntéze řeči (Psutka, 2006).

Nedávno vyvíjené korpusově orientované řečové syntetizéry měli tendence spoléhat na velkou databázi, v rozmezí od několika hodin do desítek hodin řečového korpusu k poskytnutí přirozené řeči. Ovšem navýšení velikosti korpusu ovlivní použitou metodu a oslabí proces syntézy (Dagba a spol., 2016).

Metody prozodických a spektrálních modifikací (tzn. různé metody konkatenační syntézy)

Lidská řeč se vyznačuje značnou proměnlivostí fonetických i prozodických charakteristik. Dokonce ani sebevětší korpus nemusí zahrnout každý řečový kontext, který je požadován při vytváření řeči. Z tohoto důvodu je potřeba jistým způsobem modifikovat vlastnosti řetězených segmentů (Psutka, 2006).

Modifikace jsou v zásadě dvojího druhu:

- „Prozodické modifikace upravující prozodické vlastnosti řetězených řečových segmentů za účelem co nejvíce je přiblížit požadovanému prozodickému kontextu syntetické řeči.

- Spektrální modifikace upravují frekvenční vlastnosti syntetické řeči, zejména v místech řetězení sousedních řečových jednotek, s cílem vyhladit přechody mezi řetězenými jednotkami.“ Psutka (2006, 574)

Mezi různé metody patří:

- Prostá konkatenace – primitivní a historicky nejstarší metoda konkatenační syntézy (čistá řeč, velké paměťové nároky).
- Lineární prediktivní syntéza–založena na teorii zdroje a filtru.
- PSOLA – patřící mezi nejmladší a nejúspěšnější techniky syntézy řeči.
- Harmonický šumový model vytváření řeči–složitější a výpočetně náročnější než TD-PSOLA (anglicky Time Domain PSOLA), avšak umožňuje efektivnější modelování řečového signálu (Psutka, 2006). Nejčastěji užívaná pro svou výpočetní účinnost. Nejzákladnější metoda spočívá v odvozování segmentů za pomoci Pitch (slovo Pitch ze zkraty PSOLA, více v kapitole výše) značek (Krothapalli, 2012).

Softwary, které jsou porovnávány níže v praktické části bakalářské práce, vycházejí z výše uvedených principů, které používají pro utváření mluvené řeči. Nejčastěji v praxi dochází k nahrání mnohahodinové nahrávky hlasu osoby předčítající vybraný text (obsahující co možná největší množství slov). Poté se tato nahrávka zanalyzuje a rozparceluje až na samotné hlásky. Z těchto hlásek se poté skládá strojová řeč. Tímto způsobem lze pomocí nahraných hlasových záznamů imitovat jakýkoliv hlas, a to i osob, které jsou již zesnulé.

4.1.3 Druhy řečových syntetizérů

V předchozích kapitolách byly popsány principy a metody, podle kterých se rozkládá text na určité části (fonémy, korpusy apod.) a za pomoci algoritmů se formuje do formy zvučné, tedy do mluvené řeči. Nyní

budou představeny dva (v textu výše zmíněné) druhy softwaru, které se v praxi hojně využívají v různých odvětvích.

4.1.3.1 Text-to-Speech software

TTS je druh řečové syntézy používané k vytvoření mluvené řeči z textu z počítačového dokumentu, jako je například webová stránka nebo soubor pdf. TTS umožnuje předčítání počítačového zobrazení (informací na monitoru) osobám se zrakovým postižením nebo může být použit k běžnému čtení textových zpráv. Dnešní TTS aplikace zahrnují předčítání obsahu emailových konverzací a jiné další užitečné aplikace (více informace níže) (Rouse, 2007).

4.1.3.2 Speech-to-Text software

Již méně využívaný typ softwaru, ale opět velmi užitečný, je Speech-to-Text neboli program pro převod mluvené řeči do textu. Původně byl tento typ softwaru vyvinut pro sluchově postižené. Jeho aplikace byla ovšem velmi omezená, protože starší STT musel být upravován (naučen) na rozpoznávání hlasu konkrétní osoby. V dnešní době se STT neustále vylepšuje, a kromě použití pro sluchově postižené osoby, je hojně využíván v mnoha oblastech. Mezi jeho využití patří například právnická praxe (zápis případů, výpovědi svědků a podobně), usnadnění práce v nemocnici (rychlejší zápis vyšetření rentgenových snímků a jiné) nebo využití ve vzdělávacích institucích (zápis ústního zkoušení, testů a podobně) (Rouse, 2009).

4.1.4 Využití řečových syntetizérů v praxi

Existuje mnoho využití pro řečové syntetizéry v praxi. Níže jsou uvedeny některé z nich, se kterými se může setkat kdokoliv při každodenních činnostech. Pro větší přehlednost praktického využití softwarů jsou informace rozčleněny dle kategorií uživatelů.

Osoby se zrakovým postižením či postižením pohybového aparátu:

Nejznámějším využitím TTS a STT je právě zjednodušení života osobám se zdravotním postižením, které jinak nejsou schopny (či s velkými obtížemi) vykonávat běžné činnosti jako psaní nebo čtení textu. TTS tedy mohou použít lidé se zrakovým postižením, kteří díky tomuto softwaru budou moct poslouchat jakoukoliv knihu, email a jiný text. STT bude naopak spíše sloužit osobám se sníženou schopností pohybu, které by za normálních okolností nemohly napsat zprávu nebo email na počítači.

Mezi nejvíce známé osobnosti využívající řečový syntetizér (konkrétně upravený na míru) je Stephen Hawking (Teoretický fyzik a jeden z nejslavnějších vědců dnešní doby, kterému byla ve věku 21 let diagnostikována amyotrofická laterální skleróza. Kvůli této nemoci je připoután na vozíčku a ke komunikaci používá právě výše zmíněný řečový syntetizér.) (Hawking.org.uk, 2017).

Běžní uživatelé:

Usnadnění učení:

Většina lidí TTS může použít pro usnadnění učení, a to v různých situacích, převážně předčítání dlouhých odborných textů, povinné četby nebo vlastních zápisků. Velkou výhodou je použití například v dopravních prostředcích, kde TTS může být nedocenitelným pomocníkem.

Zjednodušení užívání „chytrých" technologií:

Aniž by si to většina uživatelů uvědomovala, tak STT i TTS je implementován téměř ve všech známých chytrých zařízeních. Jako příklad STT lze uvést vytáčení čísel pomocí slov (vyslovení názvu kontaktu) nebo vyhledávání adres či bodů zájmu na mapách. Zařízení zachytí a převede uživatelova slova a pomocí adresáře identifikuje a následně vytočí konkrétní číslo v kontaktech nebo zobrazí na mapě výsledky vyhledávání. Podobné použití lze vidět i u inteligentního systému Siri, který je implementován v chytrých zařízeních společnosti Apple. Tato „aplikace" umožňuje vyhledávání na internetu za použití pouhého hlasu. Velmi obdobné funkce má u systému Android aplikace od společnosti Google.

Jako příklad TTS v chytrých zařízeních je předčítání SMS zpráv, textových dokumentů nebo výsledků překladačů. Stejným způsobem funguje i zadávání v těchto případech. Toto zadávání má výhody zejména pro seniory a osoby se zhoršenou jemnou motorikou.

Firemní uživatelé (komerční využití):

Stejně jako běžní uživatelé tak i firmy našly značné uplatnění pro TTS a STT. Zaměstnanci firem převážně využívají tyto softwary jako běžní uživatelé, tedy pro ulehčení svých činností. Důležitější je ovšem využití při činnostech jako je komunikace se zákazníky (zákaznické call centrum – zde může být hlas převáděn na text). Firma také může využít tzv. strojový hlas (TTS – zapsaný text na záznamníku je každému volajícímu přečten automatickým systémem). Dalším využitím firmami je automatický zápis ze schůzí a jiných meetingových akcí.

4.1.5 Příklady českých a zahraničních systémů

Bohužel naprostá většina uvedených systémů v knižních publikacích je již dnes nefunkční, proto jsou níže uvedeny pouze TTS softwary jiné, aktuální a plně funkční. Zmíněny jsou konkrétní české a zahraniční společnosti a software, který vyvíjejí. Pro získání podrobnějších informací je ke každé společnosti uveden odkaz na její domovskou stránku.

České systémy (software):

- SpeechTech: SpeechTech TTS (http://www.speechtech.cz/cz/)
- RosaSOFT: WinTalker 4, WinTalker 1.6 (http://www.rosasoft.cz/)
- EPOS: EPOS 2.4.85 (http://epos.ufe.cz/)
- CCNovaSoft: TTS a zejména STT (http://www.ccnovasoft.cz)

Zahraniční systémy (software):

- Acapela Group: Acapela TTS Voices (http://www.acapela-group.com/)
- LinguaTec: Voice Reader Home 15, Voice Reader Studio 15, Voice Reader Server 15, Voice Readr Web 15 (https://www.linguatec.de/en)
- ResponsiveVoice: ResponsiveVoice Free, ResponsiveVoice Pro (https://responsivevoice.org/)
- NaturalReader: NaturalReader 14 (https://www.naturalreaders.com/)

5 Praktická část

V praktické části této práce budou postupně představeny vybrané firmy vyvíjející TTS a k nim vždy jeden zástupce z řad jejich konkrétního softwaru. Jednotlivé programy firem byly podrobeny podrobné analýze a porovnány s přihlédnutím ke konkrétním vahám za pomoci programu Expert Choice (program pro podporu rozhodování umožňující porovnávání vlastností (charakteristik) s váhami určujícími jejich důležitost, více popsáno v kapitole níže). Výsledná analýza je následně rozebrána, okomentována a graficky zpracována. Na závěr je vybrán jeden TTS, který dle stanovených kritérií nejvíce vyhovuje pro implementaci ve vzdělávací instituci pro potřeby studentů.

5.1 Základní informace o použitých programech

V této kapitole jsou představeny jednotlivé TTS programy. U každého z programů jsou uvedeny základní informace o společnosti z informací dostupných na jejich oficiálních stránkách.

5.1.1 Acapela Group

Obrázek č. 10 - Logo společnosti Expert Choice. Zdroj: AcapelaBox, (2017).

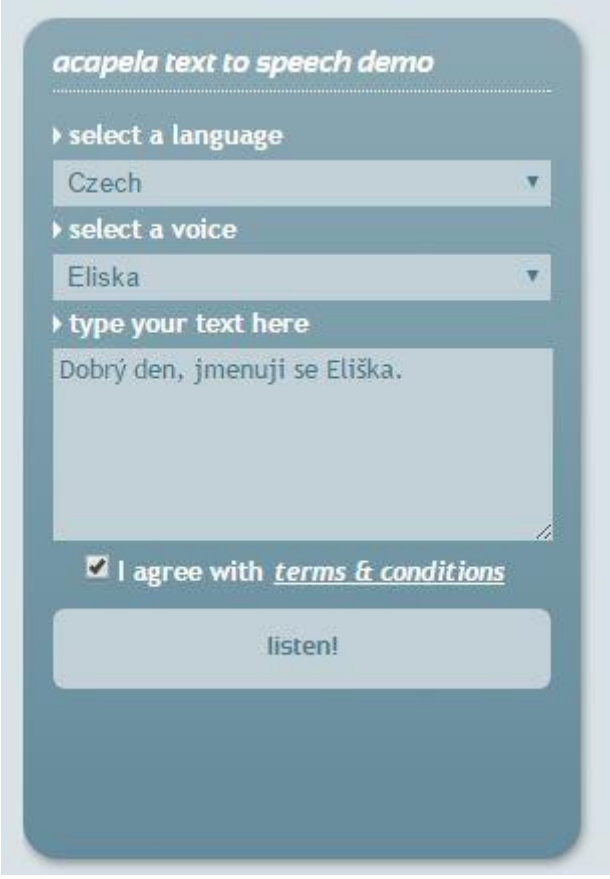
O společnosti:

Společnost Acapela Group (logo společnosti viz Obrázek č. 10) byla založena v roce 2003 spojením tří společností specializujících se na speech technologie: Babel Technologies (BE), Elan speech (FR) a Infovox (SWE). V průběhu dalších let se tyto tři společnosti zasloužily o vývoj mnoha zajímavých projektů jako TTS pro iPhone, různé mluvící programy pro děti, programy předčítající text, výzkumy v oblasti robotiky a zejména TTS, do kterých každým rokem přidávají podporu nových jazyků. Jako velmi zajímavý projekt této společnosti je i možnost použití vlastního hlasu pro jejich TTS (Acapela Group, 2017b).

Acapela Group již disponuje více jak100 hlasy ve 34 jazycích (Acapela Group, 2017a).

O programu:

Společnost Acapela Group má k dispozici webové demo svého TTS (viz Obrázek č. 11), které dokonce obsahuje český hlas pod názvem Eliška. Lze tedy vložit jakýkoliv text, který je následně možné poslechnout si jako audio. Dále je od této společnosti k dispozici verze pro Android, u které je možné zakoupení jednotlivých jazyků za cenu pohybující se okolo 130 Kč. Desktopová verze (verze pro stolní počítač) je dostupná přes obchodní partnery, kteří poskytují prodej TTS od Acapela Group ve svých zemích.



Obrázek č. 11 – Výstřižek z online demo verze Acapela TTS. Zdroj: Snímek obrazovky web-demo Acapela TTS.

5.1.2 Microsoft

Obrázek č. 12 - Logo společnosti Microsoft. Zdroj: Microsoft, (2017a).

O společnosti:

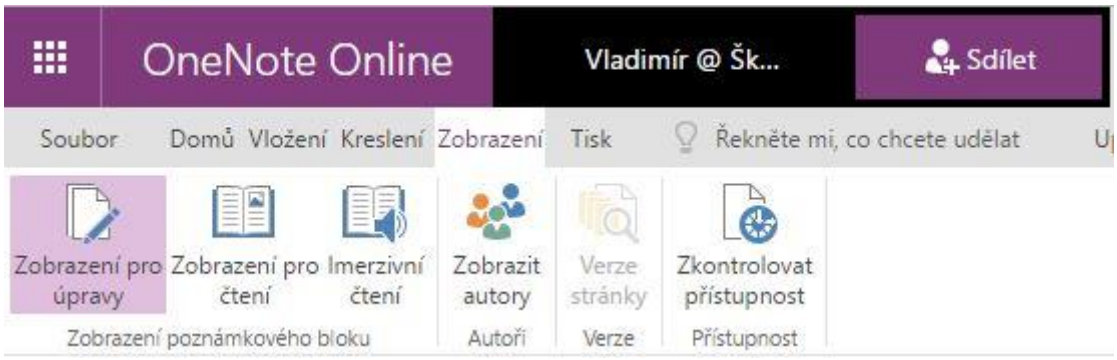
Společnost Microsoft (logo společnosti viz Obrázek č. 12) je přední výrobce technologií pro celý svět, pro který se stala mobilita a cloud naprostou nutností. Microsoft Corporation vznikla v dubnu roku 1975. Pobočka v České republice zahájila svou činnost v roce 1992. Na podzim roku 2015 se po zakoupení společnosti Skype společně s touto novou divizí přestěhovali do nových kanceláří vytvořených pro svět práce odkudkoliv. V roce 2016 se poté spojila česká a slovenská pobočka a vznikl Microsoft Česká republika a Slovensko. Vizí společnosti Microsoft je inspirovat zákazníky, jednotlivce i organizace v České republice a na Slovensku ke změně světa za pomoci digitálních technologií (Microsoft, 2017a).

Společnost Microsoft vyvinula velké množství technologií a aplikací, které denně používají miliony lidí po celém světě. Z hlediska problematiky TTS je představen nový Office 365, který obsahuje možnost čtení zapsaného textu s vysokou plynulostí, a to i v českém jazyce.

O programu:

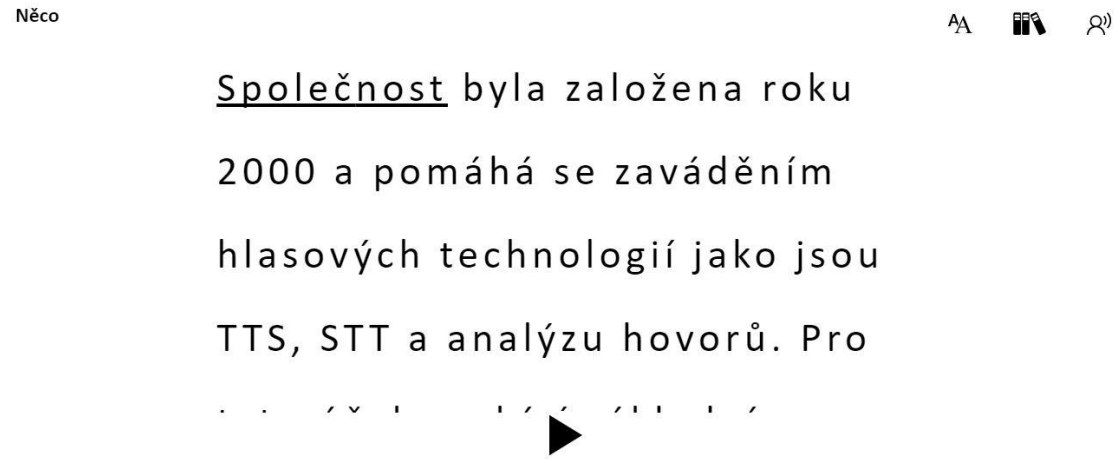
Obrázek č. 13 - Logo Microsoft Office. Zdroj: Microsoft, (2017d).

V novém Office 365 (logo viz Obrázek č. 13) je u programu „Online Word“ a „Online OneNote“ dostupná možnost přečíst jakýkoliv vybraný text za pomocí funkce „Imerzivní čtení“ na kartě „Zobrazení“ (viz Obrázek č. 14) (Microsoft, 2017a).



Obrázek č. 14 - Výstřižek z programu Microsoft Office 365. Zdroj: Snímek obrazovky Office 365 – OneNote Online.

Po stisknutí tlačítka „Imerzivní čtení“ se zobrazí nové okno s vybraným textem a jednoduchými ovládacími prvky pro úpravu hlasu a čteného textu (viz Obrázek č. 15).



Obrázek č. 15 - Výstřižek z programu Microsoft Office 365. Zdroj: Snímek obrazovky Office 365 – OneNote – Imerzivní čtení.

Celé rozhraní je velmi jednoduché a uživatelsky přívětivé. Velkou výhodou je ovšem to, že i přes omezené množství úpravy hlasu je toto TTS dostupné v již velmi známém a oblíbeném programu. Uživatel se tedy velmi snadno adaptuje na tuto novou funkci bez nutnosti složitého zaškolení v ovládání. Další výhodou je to, že mnoho společností a vzdělávacích institucí má již zakoupené licence pro Office 365, čímž lze ušetřit nemalé náklady.

O něco jednodušší forma TTS je dostupná i v desktopových verzích Microsoft Office, konkrétně v aplikacích Word, Outlook, PowerPoint a OneNote. Pro zprovoznění tohoto TTS je potřeba instalace přídavného modulu Microsoft Speech Platform a následného přidání příkazu „Přečíst“ na pás karet nebo panel nástrojů „Rychlý přístup“ (příkaz „Přečíst“ viz Obrázek č. 16) (Microsoft, 2017a).

Obrázek č. 16 - Výstřižek z programu Microsoft Office Word 2016. Zdroj: Snímek obrazovky Microsoft Office – Word 2016.

5.1.3 SpeechTech

Obrázek č. 17 - Logo společnosti SpeechTech. Zdroj: SpeechTech, (2017a).

O společnosti:

Společnost (logo společnosti viz Obrázek č. 17) byla založena roku 2000 a pomáhá se zaváděním hlasových technologií jako jsou TTS a STT. Stejně tak se věnuje analýze hovorů. Pro tyto účely nabízí základní moduly pro řečové technologie i komplexnější řešení, která jsou přizpůsobena podle potřeb zákazníka. Společnost využívá zejména technologii automatického přepisu mluveného diktátu, která slouží uživatelům v různých oblastech. Tato technologie je používána zejména ve zdravotnictví, advokacii, či při tvorbě firemních zápisů. Práce společnosti byla dokonce oceněna Křišťálovým diskem na posledním ročníku počítačového veletrhu INVEX a Českou hlavou v roce 2008 (SpeechTech, 2017a).

Obecně se společnost zabývá vývojem řady zajímavých programů pro:

- Titulkování v reálném čase živě vysílaných televizních pořadů

- TTS pro nevidomé při posílání hlasových SMS zpráv
- Interaktivní hlasový dialogový systém – pro oblast telefonie
- Možnost přepisu řečových nahrávek do textu, hlasová biometrie a pokročilá analýza nestrukturovaných dat
- Možnost komplexní analýzy (SpeechTech, 2017a).

O programu:

Společnost SpeechTech vyvíjí mnoho druhů softwaru, jak je uvedeno na jejich webových stránkách. Níže je uveden jejich přehled včetně základních funkcí, které nabízejí:

SpeechTech Megaword: Automatické přepisování mluveného diktátu do textové podoby.

SpeechTech Archiv: Tvorba a vyhledávání audiovizuálních archívů na základě technologie automatického přepisování mluvené řeči.

SpeechTech Text-to-Speech: Generování kvalitní přirozené řeči ze zadaných textů.

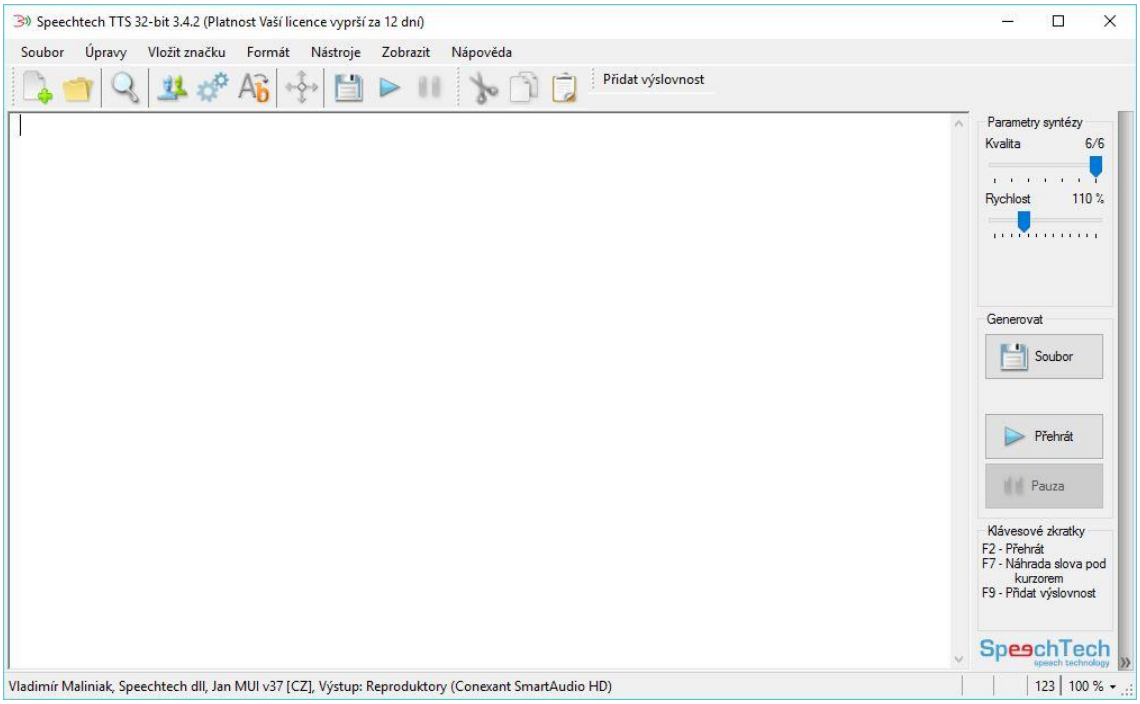
SpeechTech Hlasová biometrie: Ověření a verifikace osob na základě jedinečných vlastností jejich hlasu.

SpeechTech Analytics: Analýza hovorů v call centrech.

SpeechTech Titulkování: Vhodnost převážně pro sluchově handicapované osoby pro zpřístupnění zvukové informace multimediálního obsahu živě vysílaných pořadů.

SpeechTech Hlasové dialogové systémy: Hlasová komunikace mezi počítačem a člověkem (SpeechTech, 2017b).

Pro analýzu v rámci této práce je použitý SpeechTech TTS verze 3.4.2, který je dostupný v desktopové verzi a díky zástupci z firmy CCNovaSoft je testováno na dočasně licencované verzi (viz Obrázek č. 18).



Obrázek č. 18 - Výstřižek z programu SpeechTech TTS verze 3.4.2. Zdroj: Snímek z obrazovky SpeechTech TTS verze 3.4.2.

5.1.4 Linguatec

Obrázek č. 19 - Logo společnosti Linguatec. Zdroj: Linguatec, (2017b).

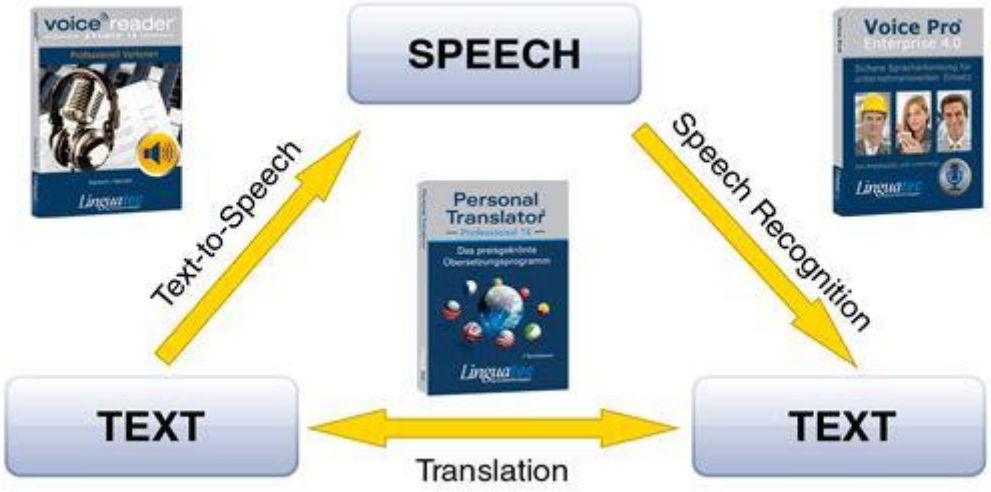
O společnosti:

Linguatec (logo společnosti viz Obrázek č. 19) je vedoucí společností poskytující jazykově orientovaný software pro kancelářské využití v Německu a zároveň je jedinou společností, která vyhrála třikrát cenu European Information Technology (volně přeloženo jako Evropská cena informačních technologií). V oblasti jazykově-technologické specializace se zaměřují na tři sekce, které lze prezentovat jako jazykově-technologický trojúhelník (viz Obrázek č. 20):

Voice Output (hlasový výstup): Převod textu na řeč (TTS).

Speech Recognition (rozeznávání hlasu): Rozeznávání hlasu, tedy reverzní proces k TTS, lze označit také jako STT.

Automatic Translation (automatický překlad): Automatické překládání textu. (Linguatec, 2017a).



Obrázek č. 20 – Linguatec jazykově-technologický trojúhelník. Zdroj: Linguatec, (2017c).

O programu:

Výhodou je, že společnost nabízí možnost vyzkoušet si jejich TTS demo online, ovšem způsobem, který imituje desktopovou verzi. Na základě této možnosti si lze program vyzkoušet v obdobné podobě jako u zakoupených verzí. Jedinou nevýhodou je, že mnoho funkcí není v demo verzi k dispozici, což lze požadovat za samozřejmé. Voice Reader 15 je vydáván ve čtyřech verzích a všechny obsahují více než 45 jazyků:

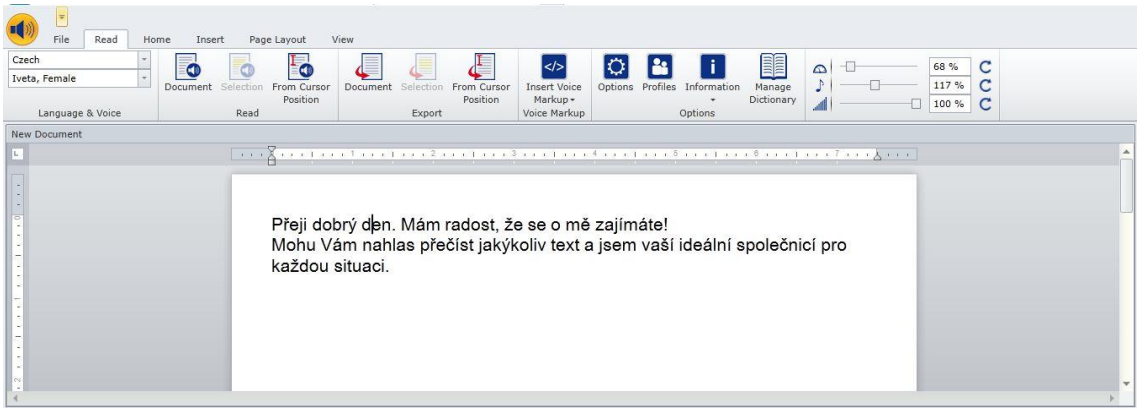
Voice Reader Home 15: pro soukromé účely

Voice Reader Studio 15: pro profesionální využití

Voice Reader Web: pro automatické předčítání webových stránek

Voice Reader server 15: pro komerční použití (Linguatec, 2017a).

A navíc je k dispozici webová demo verze „Voice Reader Studio Online Demo“ (viz Obrázek č. 21), která je použita pro testování.



Obrázek č. 21 - Výstřižek z programu Linguatec Voice Reader Studio Online Demo. Zdroj: Snímek obrazovky Voice Reader Studio Online Demo.

5.1.5 NaturalReader

Obrázek č. 22 - Logo společnosti NaturalReader. Zdroj: NaturalReader, (2017b).

O společnosti:

Společnost NaturalSoft Ltd. (logo společnosti viz Obrázek č. 22). Byla založena v Kanadě ve městě Vancouver. Cílem této společnosti je vývoj dokonalého řečového řešení pro osobní počítače a přenosná zařízení.

Společnost NaturalSoft pracuje nejen proto, aby dodržela kvalitu, ale aby předčila očekávání zákazníků. Proto se neustále snaží zlepšovat produkt NaturalReader a všechny jeho funkce tak, aby naplnil potřeby uživatelů.

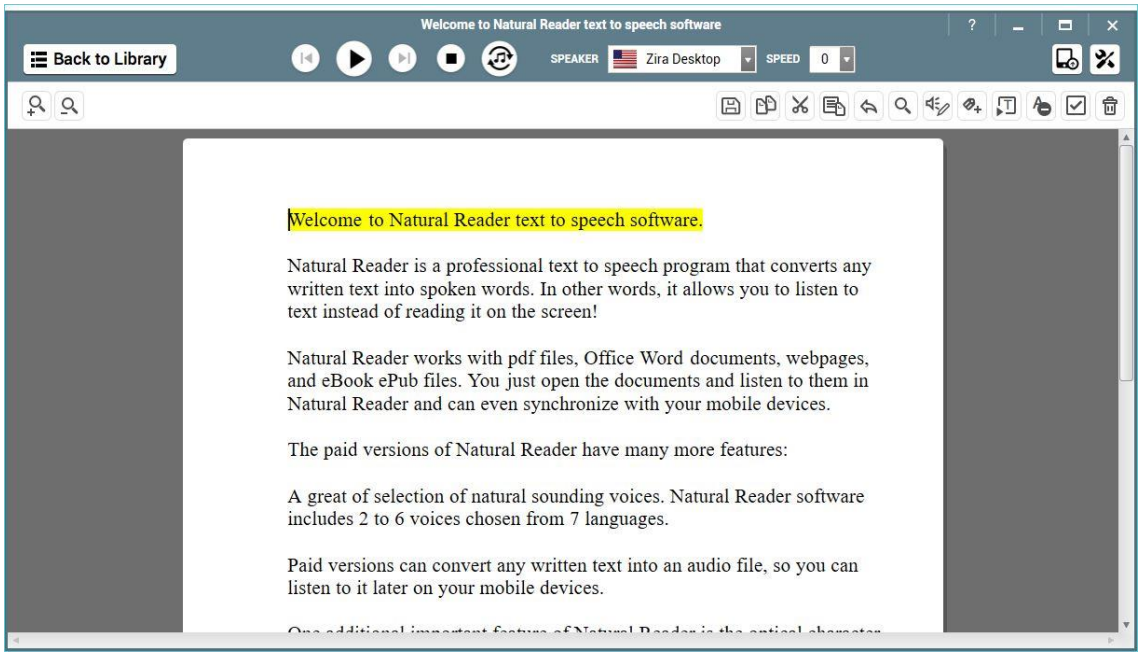
Momentálně se společnost zaměřuje primárně na TTS software. Všechny produkty jsou k dispozici online s možností platby za nadstandardní služby a jazyková rozhraní pomocí platební brány PayPal a 2CheckOut (NaturalReader, 2017a).

O programu:

Společnost NaturalSoft vyvinula TTS software s názvem NaturalReader 14 a nabízí jej ve čtyřech následujících verzích: **FREE, PERSONAL, PROFESSIONAL**

a **ULTIMATE**. Hlavní rozdíl mezi jednotlivými verzemi NaturalReader 14 je převážně počet dostupných přirozených hlasů a celkových funkcí.
Kromě zakoupení klasických samostatných verzí nabízí společnost nákup výhodných balíčků pro školní instituce a podobně, na kterých lze ušetřit nemalý obnos peněz.

Analýza v této práci je provedena na program NaturalReader 14 FREE (viz Obrázek č. 23), a to z důvodu možnosti bezplatného otestování. Proto jsou nedostupné funkce ověřeny na základě dokumentace z oficiálních webových stránek společnosti NaturalSoft. (NaturalReader, 2017a).



Obrázek č. 23 - Výstřižek z programu NaturalReader Free. Zdroj: Snímek obrazovky NaturalReader Free.

5.2 Kritéria hodnocení TTS softwaru

Všechna kritéria a váhy (viz Tabulka č. 1) jsou vypracována na základě odborných textů, ze kterých autor čerpal informace k teoretické části bakalářské práce a předchozích zkušeností z předmětu „Úvod do podnikové informatiky“. Výběr a také relevance níže uvedených kritérií a vah byla osobně prodiskutována s odborníkem (Ing. Ladislavem Daníčkem) ze společnosti CCNovasoft během

semistrukturovaného rozhovoru, který schválil na základě svých profesních zkušeností zvolená kritéria a jejich váhy.

Váhy kritérií se pohybují na škále od 1 do 10 bodů důležitosti.

Název	Váhy
Cena	10
Jazyková rozmanitost, nářečí	10
Plynulost řeči	9
Desktopová verze	8
Helpdesk	8
Převod na zvukový soubor	8
Školení	8
Aktualizace	7
Podpora OS	7
API	6
Lokalizace	6
Možnost úpravy hlasu	5
Náročnost na HW	5
Podpora textových souborů	5
Verze pro mobilní zařízení	5
Doplňky (efekty a jiné)	4
Uživatelská přívětivost	4

Tabulka č. 1 - Kritéria s váhami. Zdroj: Vlastní zpracování.

API *(Application Programming Interface, volně přeloženo jako rozhraní pro programování aplikací)* (Christensson, 2016):

Popis kritéria: Jedná se o kritérium, které nemá nejvyšší důležitost pro klasické užití, ale vzhledem k tomu, že se na Fakultě informatiky a managementu, pro kterou je volba TTS systému primárně zamýšlena, nabízí obor Aplikovaná Informatika, je třeba dbát i na to, aby bylo možné využít API TTS softwaru pro případné použití pro programování podpůrných aplikací.

Váha kritéria: Při stanovení váhy tohoto kritéria je přihlédnuto k možnosti dalšího využití pro jiné aplikace (jak je uvedeno výše). Jelikož není zcela podstatné pro samotné využití softwaru pro běžnou činnost, je váha pro API stanovena na **6 bodů.**

Cena

Popis kritéria: Pořizovací cena TTS softwaru je pro většinu institucí na prvním místě. I když by například program splňoval všechny ostatní požadavky, je třeba dbát na to, kolik prostředků může uživatel vynaložit. Porovnávána je cena pro nekomerční užití, konkrétně za jednu licenci pro jednoho uživatele. Toto kritérium je bráno všeobecně podle ceníků stanovených na oficiálních stránkách porovnávaných společností. V porovnávání není kladen ohled na možné smlouvy, či dohody ve vzdělávacích institucích umožňující čerpání TTS zdarma nebo za zvýhodněnou cenu.

Váha kritéria: Cena je zde považována za jedno z nejdůležitějších kritérií, proto je váha stanovena na nejvyšší hodnotu **10 bodů**.

Desktopová verze – verze pro PC (*osobní počítač)*

Popis kritéria: Kritérium, které je zaměřeno na dostupnost softwaru pro osobní počítače. V jistých situacích je verze pro osobní počítače nesmírnou výhodou. Například při práci na seminárních a jiných pracích, kdy není tedy potřeba mít k dispozici více zařízení, ale uživatel si vystačí s jedním – tedy PC. Toto kritérium je důležité také pro firemní uživatele, kteří zejména pracují na PC.

Váha kritéria: Vzhledem k velké důležitosti dostupnosti verze pro PC je tomuto kritériu přiřazena váha s hodnotou **8 bodů**.

Doplňky (efekty a jiné)

Popis kritéria: Toto kritérium sleduje, zda má daný software k dispozici navíc nějaké doplňky, které zlepšují běžnou funkci softwaru. Jako příklad doplňku lze uvést možnost úpravy hlasu (akustika), zabarvení již přečteného textu, grafická simulace postavy a jiné.

Váha kritéria: Jelikož doplňky jsou bonus ke standardním činnostem softwaru, které má splňovat, je váha tohoto kritéria stanovena na **4 body**.

Aktualizace (z hlediska ceny a frekvence)

Popis kritéria: Velmi důležité kritérium je zastoupeno aktualizací. Je žádoucí, aby byl software co možná nejčastěji aktualizován, aby se tak neustále opravovaly případné chyby a nedostatky a zlepšovaly se funkcionality.

Váha kritéria: Potřeba aktualizací je samozřejmá, je to hlavní požadavek pro stabilitu a správnou funkcionalitu softwaru, proto je váha stanovena na **7 bodů**.

Helpdesk (podpora)

Popis kritéria: V případě jakýchkoliv problémů je zapotřebí podpory ze strany společnosti, která software vyvinula. Uživateli se mnohokrát může stát, že se dostane do situace, u které si neví rady (chybová hláška, postup při syntéze). Hodnocena je dostupná forma komunikace se společností.

Váha kritéria: Přihlédnuto je k důležitosti podpory ze strany společnosti poskytující software. Proto je váha tohoto kritéria stanovena na **8 bodů**.

Jazyková rozmanitost a nářečí

Popis kritéria: Dalším důležitým kritériem je podpora jazyků a nářečí. Na půdě univerzity je možné setkat se s publikacemi v množství rozmanitých jazyků, a také výměnnými studenty ze zahraničních univerzit. Proto je potřeba, aby byl TTS software schopen předčítat v mnoha jazycích a nářečích (čím více, tím samozřejmě lépe).

Váha kritéria: Jedná se zde o kritérium, které je podstatou samotného TTS, proto je váha stanovena na maximální hodnotu **10 bodů**.

Lokalizace (*význam: dostupnost softwaru v určitém jazyce*)

Popis kritéria: Je důležité, aby byl program dostupný ve více jazycích, jelikož se na fakultě pohybují lidé z různých zemí (v rámci programu Erasmus a jiných aktivit). Toto kritérium uvažuje dostupnost jazyka, ve kterém je samotný program popsán (tzn.: menu, popis programu a podobně).

Váha kritéria: Nejdůležitější je zajisté podpora českého jazyka, ale jak je uvedeno výše, jsou situace, u kterých je potřeba dostupnosti více jazyků. Z tohoto důvodu je váha stanovena na průměrných **6 bodů**.

Možnosti úpravu hlasu

Popis kritéria: V některých situacích je potřeba, aby byl hlas mírně upraven, ať již pro lepší poslech, nebo pro urychlení předčítání. V rámci tohoto kritéria jsou brány v úvahu možnosti jako zrychlení předčítání, zvýšení a snížení hloubky hlasu.

Váha kritéria: Nejedná se o příliš důležité kritérium, přestože někdy může být nabízena zajímavá úprava hlasu. Váha je proto stanovena na průměrných **5 bodů**.

Náročnost na hardware

Popis kritéria: TTS software je na rozdíl od STT softwaru poměrně nenáročný. Zpravidla je jeho velikost okolo 30–400 Mb (megabyte).

Váha kritéria: Z důvodu nenáročnosti TTS softwaru na zařízení je tomuto kritériu přiřazena váha **5 bodů**.

Plynulost řeči

Popis kritéria: Plynulost lze chápat jako schopnost softwaru napodobit reálný hlas do podobné nebo stejné míry. Čím je řeč plynulejší, tím lépe se jí naslouchá a celkový výstup je pro uživatele přínosnější.

Váha kritéria: Toto kritérium patří mezi důležitá kritéria, jelikož úzce souvisí s hlavní funkcí softwaru, proto je u něj váha stanovena na hodnotu **9 bodů** (*je hodnoceno na základě subjektivního názoru autora*).

Podpora operačního systému

Popis kritéria: Je důležité sledovat, na jaký operační systém je možné daný software nainstalovat a na jakém jej lze využívat. U desktopového softwaru je důležitá kompatibilita s operačními systémy Windows, Macintosh a Linux. U mobilních zařízení zejména Android, Windows Phone, iOS (iPhone Operating System, volně přeloženo jako mobilní operační systém běžící na iPhonu (iPadu a iPod Touch)) (Rouse, 2017).

Váha kritéria: Každý uživatel může využívat zcela odlišný OS, proto je váha stanovena na **7 bodů**.

Podpora textových souborů

Popis kritéria: Toto kritérium řeší o možnost otevření textových souborů, které obsahují nějaký text k syntéze. Kritérium je zavedeno proto, aby se sledovalo, jaké formáty souborů daný TTS software podporuje.

Váha kritéria: Z důvodu všeobecné možnosti vkládat „surový“ text přímo do TTS softwaru je váha stanovena na **5 bodů**.

Převod na zvukový soubor

Popis kritéria: Jednou z důležitých funkcí je i možnost uložit si syntézu jako zvukový soubor pro pozdější přehrání. Jakýkoliv text lze nahrát (například knihu) a výsledný přečtený text převést (uložit) do zvoleného zvukového souboru (například MP3, WAV nebo Ogg).

Váha kritéria: Toto je poměrně důležité kritérium, jelikož se jedná o výstup funkce TTS softwaru, který je tak možné využít později. Váha je zde stanovena na **8 bodů**.

Školení

Popis kritéria: Při zakoupení softwaru je důležitá podpora při implementaci (proces zavedení do provozu). Zde je sledováno, jakou podporu poskytuje daná společnost, zda je ochotna asistovat při zavádění softwaru a zaškolení uživatele v ovládání.

Váha kritéria: I když složitost TTS softwaru není zdaleka vysoká, je výpomoc při implementaci a zaškolení odpovědného personálu (uživatelů) velmi důležitým faktorem. Z tohoto důvodu je váha stanovena na **8 bodů**.

Uživatelská přívětivost

Popis kritéria: Uživatelská přívětivost znamená, jakým způsobem je pro uživatele daný software srozumitelný. Sleduje se, aby jak snadno se uživatel vyzná v ovládání a zda mu je celkový design prostředí příjemný.

Váha kritéria: I když se nejedná o příliš důležité kritérium, není dobré jej podceňovat. Váha je zde proto stanovena na **4 body** (*je hodnoceno na základě subjektivního názoru autora na základě práce s jednotlivými softwary*).

Verze pro mobilní zařízení

Popis kritéria: Vzhledem k neuvěřitelnému nárůstu užívání mobilních zařízení je zapotřebí sledovat, zda je k dispozici i aplikace pro mobilní telefony, tablety a čtečky.

Váha kritéria: Primární je desktopová verze TTS softwaru, proto je zde váha stanovena na **5 bodů**.

5.3 Hodnocení kritérií u konkrétních firem

V této kapitole jsou rozebrána kritéria u konkrétních firem. Vybrané informace jsou uvedeny v tabulkách pro lepší přehlednost.

5.3.1 Acapela Group

Jak již bylo zmíněno výše, Acapela Group je převážně firma, která sice vyvíjí vlastní software, ale zároveň své know-how dává k dispozici jiným firmám pro využití TTS. Princip Acapela Group lze tedy vidět u mnoha jiných programů, které využívají právě „jádro“ tohoto systému.

Jednotlivá kritéria proto budou velmi často posuzována z informací dostupných od české společnosti spektra.eu (společnost poskytující širokou škálu programů zejména pro postižené osoby) distribuující Acapela Group software. (Spektra.eu, 2017). Všechna kritéria TTS společnosti Acapela Group jsou popsána v Tabulce č. 2 níže.

Název kritéria	Slovní popis
API	Pro vývojáře je k dispozici mnoho různých verzí API. Konkrétně se jedná o tato API: Cloud API, Server, Desktop, Mobile, X, 10, Android, Linux server, Linux Embedded. Díky

	tomuto je Acapela Group velmi dobrá pro vývojáře, jelikož není nutné striktně se držet jedné platformy.
Cena	Cena za tento TTS je 9 917,36 KČ vč. aplikace (cena stanovena na základě ceníku uvedeného na webových stránkách „spektra.eu“.).
Desktopová verze	Je k dispozici jako desktopová verze, konkrétně je možné používat TTS až na 3 PC v rámci jedné licence. Výhodou je, že není potřeba instalace, software je spustitelný z přenosného USB.
Doplňky	Podpora SAPI4 a SAPI5 (rozhraní, přes které mluvící aplikace syntézu používají), NVDA (odečítač obrazovky), Pronunciation Editor (úprava výslovnosti jmen a jiných slov), Voice Manager (možnost vytváření a úpravy hlasu podle vlastních požadavků) (Spektra.eu, 2017).
Aktualizace	Společnost Acapela Group neustále přidává nové jazyky a hlasy do celkové nabídky svých programů.
Helpdesk	Pro pomoc lze kontaktovat českou společnost spektra.eu pomocí emailu nebo přímo acapela-group.com pomocí kontaktního formuláře dostupného na webových stránkách.
Jazyková rozmanitost	K dispozici je český hlas Eliška a dalších 100 hlasů v 31 jazycích.
Lokalizace	Arabština, Holandština, Angličtina, Finština, Francouzština, Němčina, Italština, Norština, Švédština.
Možnost úpravy hlasu	Je možné použít jisté mutace hlasu, nastavit hlasitost, rychlost, nastavit způsob předčítání a jiné.
Náročnost na HW	Nenáročné na HW, lze spustit pomocí USB. Soubory je možné uložit na PC a spouštět i bez potřeby USB až na 3 PC. Za pomoci modulu Colibri jsou jednotlivé hlasy zmenšené a nezabírají tedy velké množství místa na disku. V opačném případě je možné využívat vysoce kvalitní hlasy o velikosti do 250 Mb.

Plynulost řeči	Velice kvalitní hlasy s plynulým předčítáním. Nejplynulejší a nejkvalitnější z dostupných hlasů je anglický, který je velice podobný skutečnému hlasu. Hlas Eliška je také plynulý, ale lze rozeznat, že se jedná o hlas strojový.
Podpora OS	Windows 7, Windows 8 a Windows 10 s plnou podporou 32 a 64 bitových verzí Windows.
Podpora textových souborů	Není přímo vázáno na textové soubory, dokáže totiž přečíst veškerý text na monitoru.
Převod na zvukový soubor	Za pomoci modulu Speech Creator (*volně přeloženo jako Tvůrce řeči*), lze přečíst jakýkoliv text a uložit jej jako MP3.
Školení	Školení není přímo zahrnuté v licenci programu, ale je možné navštívit společnost spektra.eu v předváděcím centru po předchozí domluvě a zjistit případné informace týkající se chodu TTS. V případě zájmu pro větší skupinu lidí, či za příplatek je možné individuální školení.
Uživatelská přívětivost	Program je velmi jednoduchý na ovládání a má příjemné rozhraní.
Verze pro mobilní zařízení	Acapela Group poskytuje aplikaci pro Android, která funguje na podobném principu jako desktopová verze s tím rozdílem, že je zapotřebí zakoupit jednotlivé hlasy zvlášť. Tato aplikace není součástí licence desktopové verze.

Tabulka č. 2 – Kritéria TTS společnosti Acapela Group. Zdroj: Vlastní zpracování.

5.3.2 Microsoft

U společnosti Microsoft je popsán Microsoft Office 365, který obsahuje možnost předčítání obsahu textu, i když ve velmi zjednodušené podobě. Jak bylo popsáno výše, jedná se o přídavný modul (Imerzivní čtení) k již exitujícímu programu. Popisován je tedy tento samostatný modul. Velká výhoda spočívá zejména v tom, že Microsoft Word je jedním z nejpoužívanějších textových editorů, čímž by byla velmi jednoduchá a levná implementace. Všechna kritéria TTS společnosti Microsoft jsou popsána v Tabulce č. 3 níže.

Název kritéria	Slovní popis
API	Není k dispozici, je pouze přiřazen k Office 365.
Cena	Samostatně neprodejný, je součástí Office 365, který je v rámci univerzity pro studenty zdarma. Jinak se tržní cena pohybuje okolo 1 800 Kč/rok pro nekomerčního uživatele.
Desktopová verze	Tento software je dostupný jako desktopová verze. Office 365 lze spustit prakticky na jakémkoli zařízení s připojením k internetu.
Doplňky	Změna velikosti textu a mezer mezi textem, změna druhu písma, možnost navolení pozadí, rozparcelování textu (podle slabik, podstatných jmen, sloves a přídavných jmen) a úprava rychlosti předčítání.
Aktualizace	Pravidelné aktualizace od společnosti Microsoft.
Helpdesk	Nepřetržitá pomoc od vyškolených profesionálů po celém světě v každou denní dobu, a to jak mailem, tak telefonicky.
Jazyková rozmanitost	Obsahuje velmi kvalitní český jazyk (mužský hlas). Možnost přepnout na jakýkoliv jazyk a tím změnit jazyk předčítání.
Lokalizace	Téměř ve všech světových jazycích.
Možnost úpravy hlasu	Možnost nastavení rychlosti předčítání.

Náročnost na HW	Nároky jsou minimální, jelikož vše funguje online v internetovém prohlížeči, záleží tedy pouze na rychlosti internetového připojení.
Plynulost řeči	Předčítání je velice plynulé, téměř k nerozeznání od skutečné lidské řeči.
Podpora OS	Podpora všech verzí Windows a Mac.
Podpora textových souborů	Všechny textové dokumenty Microsoft Office. Případně lze text z ostatních souborů přesunout do souboru s podporovaným formátem.
Převod na zvukový soubor	Není možný.
Školení	Probíhají různá školení po celém světě, a to jak oficiální od společnosti Microsoft, tak i neoficiální (soukromé).
Uživatelská přívětivost	Léty osvědčený uživatelsky přívětivý design, na který je naučena většina uživatelů PC v České republice a v Evropě.
Verze pro mobilní zařízení	Jelikož je dostupný jako přídavný modul v Office 365, lze nainstalovat aplikaci Office 365 do mobilního zařízení mající (iOS, Android nebo Windows).

Tabulka č. 3 - Kritéria TTS společnosti Microsoft. Zdroj: Vlastní zpracování.

5.3.3 SpeechTech

Pro otestování jednotlivých kritérií je použit Speechtech TTS 32-bit 3.4.6. Kritéria TTS společnosti SpeechTech jsou popsána v Tabulce č. 4 níže.

Název kritéria	Slovní popis
API	Není dostupný.
Cena	Cena se odvíjí na základě počtu hlasů v objednávce. Nejlevnější řešení je 1 český hlas za 2500 Kč. Nejdražší nabídka je s 5ti českými hlasy za 9990 Kč. Případně lze zakoupit hlasy samostatně za 2500 Kč.
Desktopová verze	Celý program je ve verzi pro PC.
Doplňky	Nastavení speciální výslovnosti pro konkrétní slova či zkratky (např.: „CZ“ se přečte jako „Česká Republika“).
Aktualizace	Po zakoupení občasné aktualizace od společnosti SpeechTech. Jediný rozdíl spočívá ve verzi pro domácnosti a firmy. Pro domácnosti je dostupné neomezeně, u firem jen 1 rok, a poté je nutné zakoupit novou licenci.
Helpdesk	Možnost kontaktování pomocí mailu, či telefonu. Zaměstnanci společnosti jsou ochotní a vždy rádi poradí.
Jazyková rozmanitost	Český, slovenský a anglický jazyk.
Lokalizace	Program je kompletně v českém, slovenském a anglickém jazyce.
Možnost úpravy hlasu	Možnost úpravy rychlosti a hlasitosti předčítání.
Náročnost na HW	Program je poměrně nenáročný, avšak jako doporučené je uvedeno 4GB RAM, 4 jádrové CPU, a OS Windows 7 a vyšší. Pro některé uživatele by mohl být problém s velikostí jednotlivých hlasů (okolo 400 Mb jeden).

Plynulost řeči	Řeč je velice kvalitní a srovnatelná se skutečnou lidskou řečí.
Podpora OS	Microsoft Windows 10, 10PRO, Microsoft Windows 7 a 8, 32-bit a 64-bit, Microsoft Windows Vista 32-bit a 64-bit, Microsoft Windows XP 32-bit.
Podpora textových souborů	Podporuje textové soubory formátu Microsoft Office Word a základní implementované formáty Microsoft Windows.
Převod na zvukový soubor	Přečtený text lze převést na zvukový soubor ve formátu WAV a MP3.
Školení	Školení není přímo součástí licence, avšak je možné domluvit se se společností o případné možnosti školení.
Uživatelská přívětivost	Jednoduché ovládání s obrázkovými tlačítky a jasnými příkazy.
Verze pro mobilní zařízení	Je dostupná verze pro Android.

Tabulka č. 4 - Kritéria TTS společnosti SpeechTech. Zdroj: Vlastní zpracování.

5.3.4 Linguatec

Jednotlivá kritéria jsou pozorována na dostupné demo verzi, která je k dispozici z oficiálních webových stránek společnosti Linguatec. Všechna kritéria TTS společnosti Linguatec jsou popsána v Tabulce č. 5 níže.

Název kritéria	Slovní popis
API	Není k dispozici.
Cena	Cena je stanovena vždy na jeden hlas s cenou 1 275 Kč s možností dokoupení dalších hlasů za stejnou cenu.
Desktopová verze	Licence se vztahuje pouze na desktopovou verzi.
Doplňky	Umístění značek do textu, čtení konkrétních bloků textu, čtení z pozice kurzoru, čtení z dokumentu.
Aktualizace	Možnost získání aktualizované verze pomocí webového formuláře. Nutné zadat licenční klíč, nebo uhradit požadovanou částku.
Helpdesk	Možnost kontaktování podpory přes webový formulář.
Jazyková rozmanitost	Možnost zakoupení až 45 jazyků s celkově 67 hlasy.
Lokalizace	Program je dostupný v anglickém a německém jazyce.
Možnost úpravy hlasu	Nastavitelnost hlasitosti, rychlosti a jemnosti (tónu) hlasu.
Náročnost na HW	Alespoň 1 GB na 1 hlas, 2 GB RAM paměti.
Plynulost řeči	V anglickém jazyce kvalitní strojová řeč, která je u českého hlasu bohužel rozeznatelná od skutečného hlasu.
Podpora OS	Operační systém Windows Vista, Windows 7, Windows 8 a Windows 10 (vždy ve verzi 32- a 64- bitů).
Podpora textových souborů	Jsou podporovány textové soubory ve formátu .txt, .rtf, .docx, .doc, .html, .epub, .htm, .pdf, .odt a .mht.

Převod na zvukový soubor	Lze ukládat zvukový záznam přečteného textu na zvukový soubor, a to ve formátu MP3 a WAV.
Školení	Není k dispozici. Je ovšem možnost kontaktovat helpdesk pro případnou pomoc.
Uživatelská přívětivost	Velmi jednoduché ovládání podobné textovým editorům od společnosti Microsoft.
Verze pro mobilní zařízení	Není k dispozici.

Tabulka č. 5 - Kritéria TTS společnosti Linguatec. Zdroj: Vlastní zpracování.

5.3.5 NaturalReader

Jednotlivá kritéria jsou pozorována na volně stažitelné desktopové verzi, která je dostupná z oficiálních webových stránek společnosti NaturalReader. Všechna kritéria TTS společnosti NaturalReader jsou popsána v Tabulce č. 6 níže.

Název kritéria	Slovní popis
API	Není k dispozici.
Cena	Možnost využití omezené verze zdarma, nebo zakoupení plné verze programu za 1 514 Kč nebo verzi s funkcí OCR (Optical Character Recognition, volně přeloženo jako „Optické rozpoznávání znaků“) za 2 820 Kč. K dispozici je i možnost zakoupení prémiové webové verze za 108 Kč/měsíc.
Desktopová verze	K dispozici je plně funkční online a PC verze.
Doplňky	Možnost přesunutí textového souboru do okna programu pro okamžité předčítání, za pomoci funkce OCR dokáže předčítat text z obrázků, synchronizování řeči pro mobilní zařízení. Úprava výslovnosti u konkrétních slov a zkratek.
Aktualizace	Vývojáři stále rozvíjejí a vylepšují stávající verze TTS v pravidelných intervalech. Aktualizování na nejnovější verzi stojí 1 078 Kč.
Helpdesk	Možnost kontaktovat podporu pomocí emailové adresy nebo kontaktního formuláře dostupného na oficiálních webových stránkách společnosti.
Jazyková rozmanitost	K dispozici je 11 jazyků s celkově 57 hlasy.
Lokalizace	Anglický jazyk.
Možnost úpravy hlasu	Úprava rychlosti předčítání.
Náročnost na HW	Velmi nenáročný program. Min 500 Mb volného místa, Procesor 500 Mhz, RAM minimálně 64 Mb.

Plynulost řeči	Velmi kvalitní a plynulá řeč podobná opravdové lidské řeči.
Podpora OS	Windows 7, Windows 8, Windows 8.1, Windows 10, Mac OS X 10.10 a vyšší.
Podpora textových souborů	Podpora textových souborů ve formátu .pdf, .doc, .txt, .rtf a .epub.
Převod na zvukový soubor	Převod na zvukový soubor je možný, a to do formátu MP3, Ogg a WAV.
Školení	Není k dispozici.
Uživatelská přívětivost	Zajímavý a hravý design s příjemnými barvami a jednoduchým ovládáním. Bohužel pouze v anglickém jazyce.
Verze pro mobilní zařízení	Není, je zde ovšem možnost synchronizování (převodu) textu a předčítání z desktopové verze do mobilního zařízení.

Tabulka č. 6 - Kritéria TTS společnosti NaturalReader. Zdroj: Vlastní zpracování.

5.4 Porovnání jednotlivých kritérií TTS systémů

Níže je u každého kritéria přiřazen graf nebo tabulka, ve kterých jsou porovnány jednotlivé TTS mezi sebou. Lze tak velmi zřetelně vidět, který TTS je v daném kritériu lepší. Ke každému grafu a každé tabulce je doplněn popis, který lépe vysvětluje důvody výsledků.

API

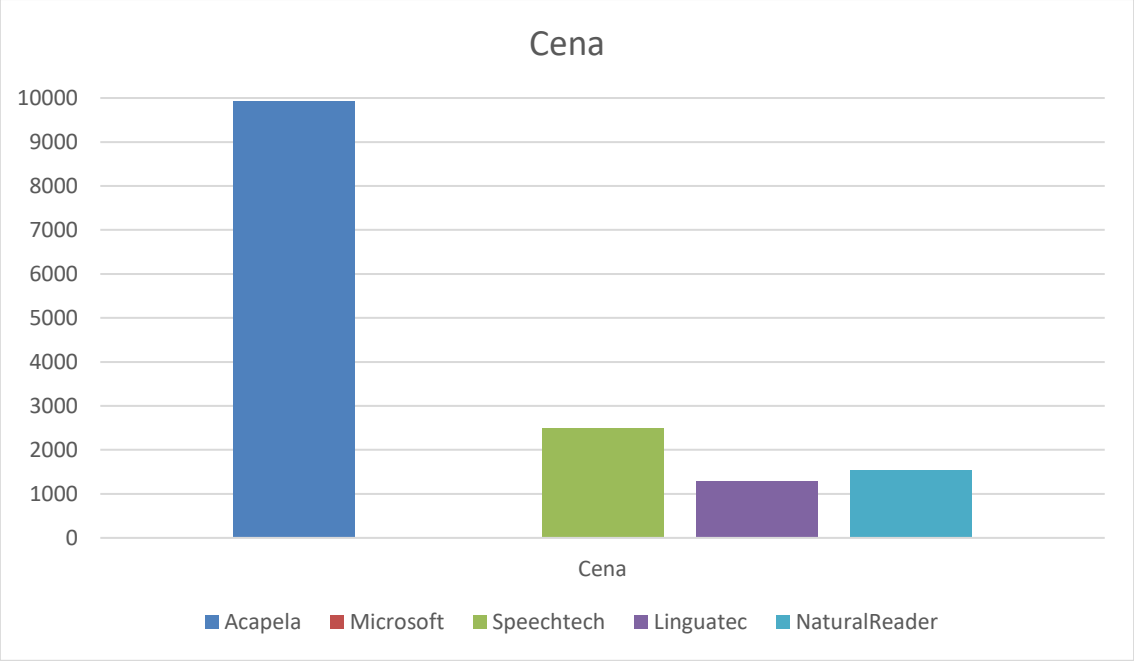
V následující Tabulce č. 7 níže je zobrazeno, zda daný TTS umožňuje API pro případné vývojáře.

Název společnosti	Umožňuje API	Neumožňuje API
Acapela Group	☒	☐
Microsoft	☐	☒
SpeechTech	☐	☒
LinguaTec	☐	☒
NaturalReader	☐	☒

Tabulka č. 7 - Tabulka srovnání TTS systémů z hlediska API. Zdroj: Vlastní zpracování.

Cena

V Grafu č. 1 níže je zobrazen minimální obnos Kč potřebný k zakoupení daného programu. Čím více je nutné investovat, tím méně vhodný je daný systém pro nekomerčního uživatele.



Graf č. 1 – Graf srovnání TTS systémů z hlediska ceny. Zdroj: Vlastní zpracování.

Desktopová verze

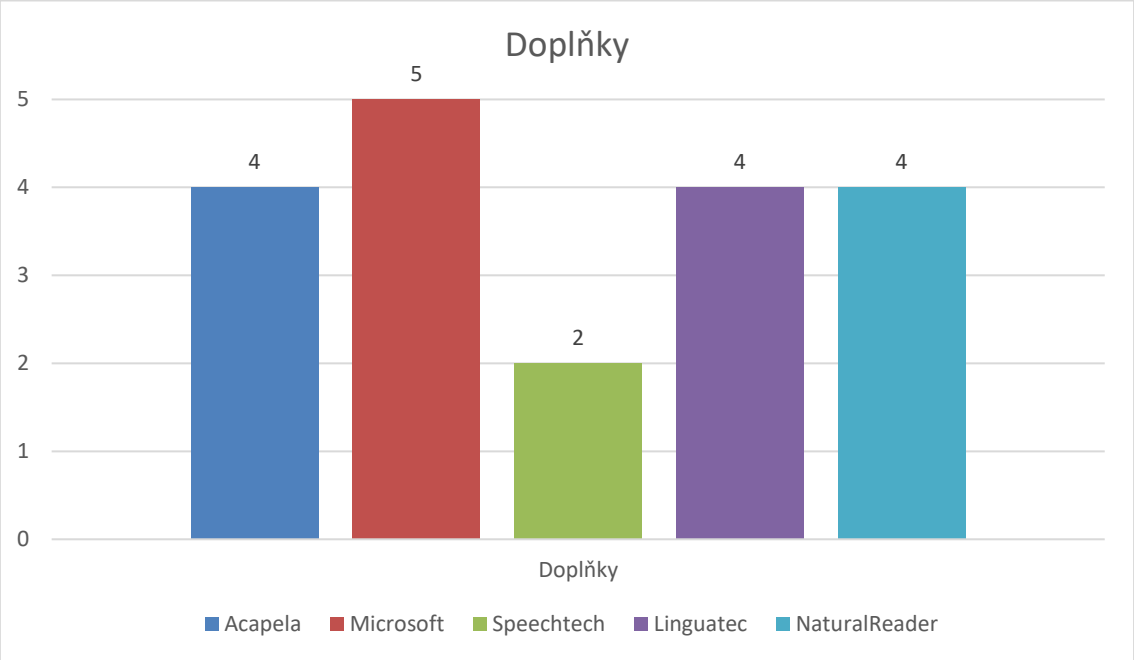
V následující Tabulce č. 8 níže je zobrazeno, zda je daný TTS dostupný v desktopové verzi. Toto kritérium sice neovlivňuje výsledky porovnání, ale je natolik důležité, že je do porovnání zahrnuto.

Název společnosti	Desktopová verze	
	ANO	NE
Acapela Group	☒	☐
Microsoft	☒	☐
SpeechTech	☒	☐
LinguaTec	☒	☐
NaturalReader	☒	☐

Tabulka č. 8 - Tabulka srovnání TTS systémů z hlediska dostupnosti desktopové verze. Zdroj: Vlastní zpracování.

Doplňky (efekty a jiné)

V Grafu č. 2 níže je zobrazeno, jaký počet doplňků jednotlivý TTS obsahuje.



Graf č. 2 - Graf srovnání TTS systémů z hlediska doplňků. Zdroj: Vlastní zpracování.

Aktualizace

V následující Tabulce č. 9 je zobrazeno, jak často jsou aktualizace u jednotlivých TTS poskytovány.

Název společnosti	Aktualizace		
	Vůbec	Občasně	Pravidelně
Acapela Group	☐	☒	☐
Microsoft	☐	☐	☒
SpeechTech	☐	☒	☐
LinguaTec	☒	☐	☐
NaturalReader	☒	☐	☐

Tabulka č. 9 - Tabulka srovnání TTS systémů z hlediska frekvence aktualizace. Zdroj: Vlastní zpracování.

Helpdesk

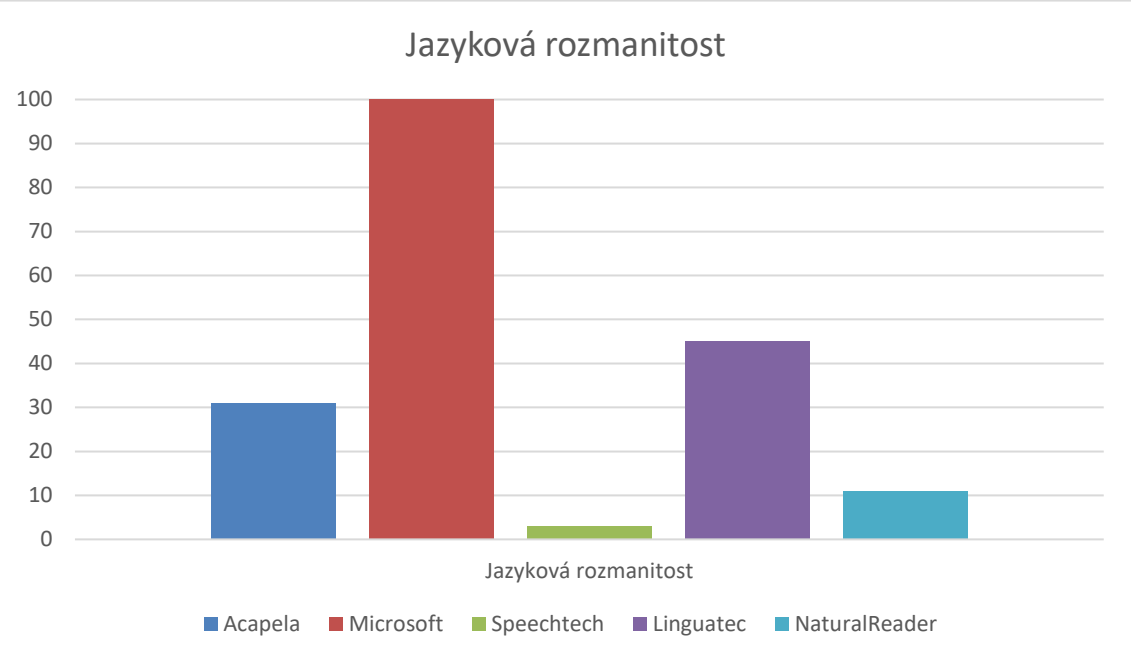
V Tabulce č. 10 níže je zaznamenána forma kontaktování helpdesk.

Název společnosti	Helpdesk	
	web/mail	tel.
Acapela Group	☒	☐
Microsoft	☒	☒
SpeechTech	☒	☐
LinguaTec	☒	☐
NaturalReader	☒	☐

Tabulka č. 10 - Tabulka srovnání TTS systémů z hlediska helpdesk. Zdroj: Vlastní zpracování.

Jazyková rozmanitost, nářečí

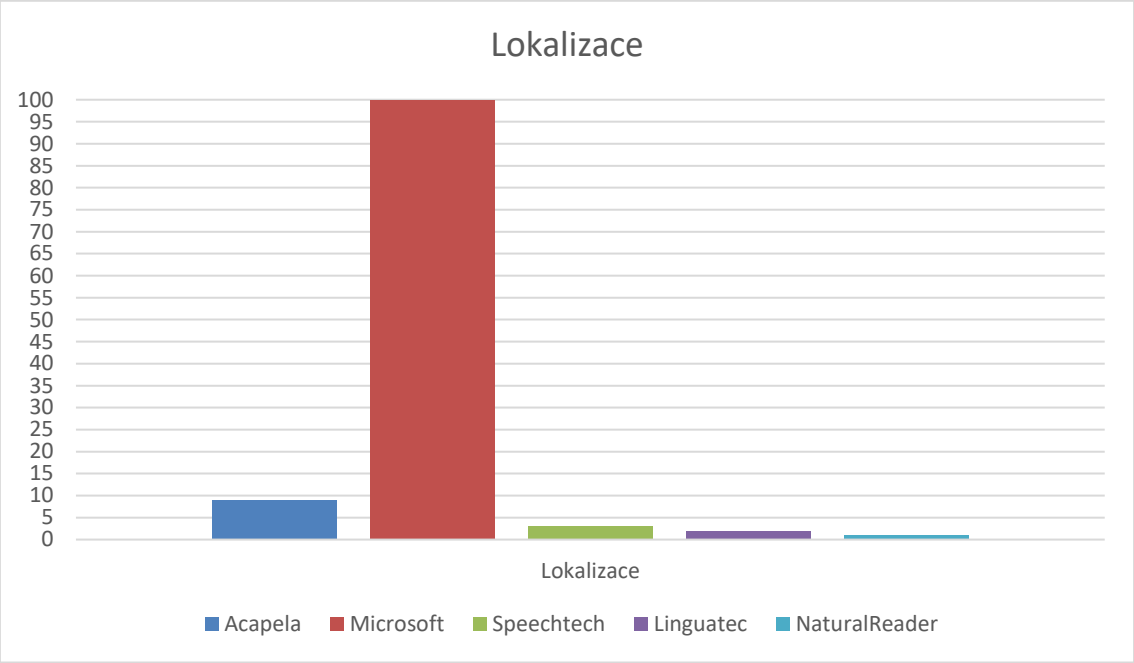
V Grafu č. 3 níže jsou zobrazeny jednotlivé TTS podle počtu dostupných jazyků předčítání.



Graf č. 3 - Graf srovnání TTS systémů z hlediska jazykové rozmanitosti. Zdroj: Vlastní zpracování.

Lokalizace

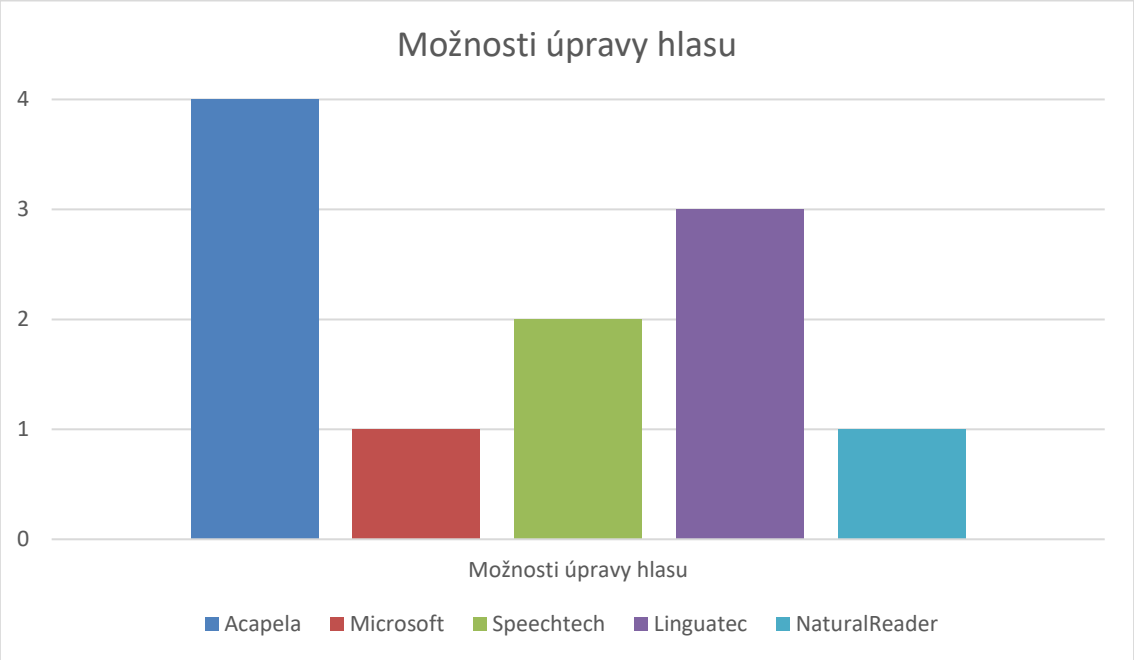
V Grafu č. 4 níže jsou zobrazeny jednotlivé TTS podle počtu jazyků, do kterých je TTS přeložen.



Graf č. 4 - Graf srovnání TTS systémů z hlediska lokalizace. Zdroj: Vlastní zpracování.

Možnosti úpravy hlasu

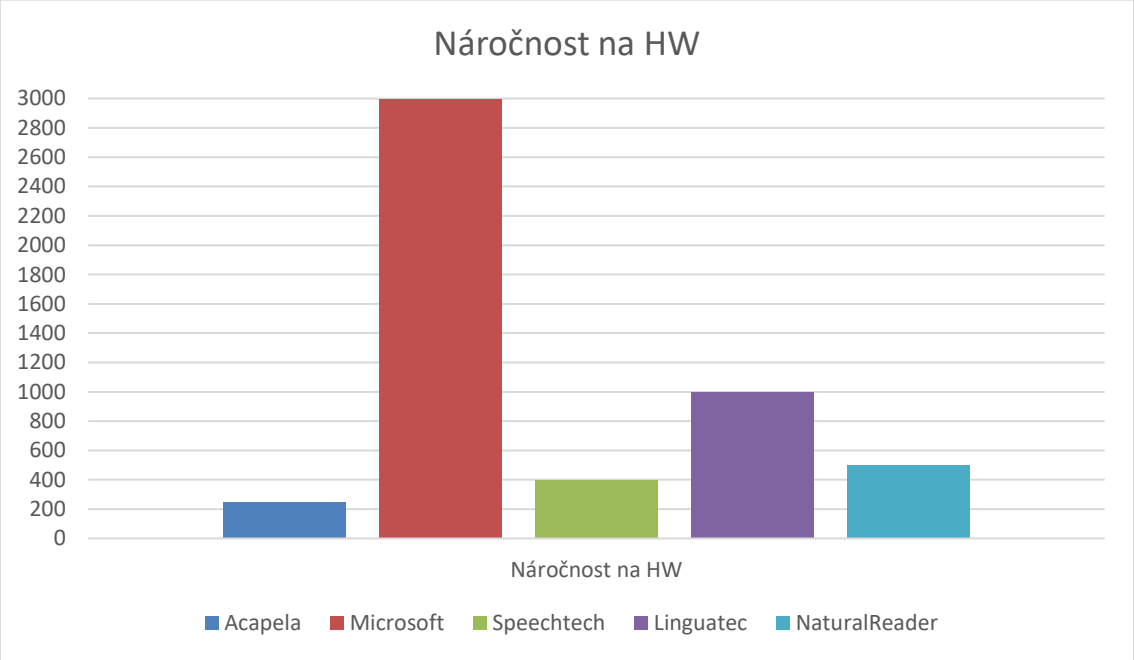
V Grafu č. 5 níže jsou hodnoty určující počet dostupných funkcí úpravy hlasu.



Graf č. 5 - Graf srovnání TTS systémů z hlediska možnosti úpravy hlasu. Zdroj: Vlastní zpracování.

Náročnost na HW

V Grafu č. 6 níže je zobrazena potřeba místa na disku v Mb.



Graf č. 6 - Graf srovnání TTS systémů z hlediska náročnosti na HW. Zdroj: Vlastní zpracování.

Plynulost řeči

V následující Tabulce č. 11 níže je hodnocena kvalita předčítaného hlasu na škále od 1 do 10 (hodnoceno na základě subjektivního názoru autora).

Název společnosti	Plynulost řeči
	Kvalita na škále od 1 do 10
Acapela Group	6
Microsoft	8
SpeechTech	6
LinguaTec	7
NaturalReader	8

Tabulka č. 11 - Tabulka srovnání TTS systémů z hlediska plynulosti řeči. Zdroj: Vlastní zpracování.

Podpora OS

V následující Tabulce č. 12 jsou zobrazeny podporované OS.

Název společnosti	Podpora OS	
	Windows	Mac
Acapela Group	☒	☐
Microsoft	☒	☒
SpeechTech	☒	☐
LinguaTec	☒	☐
NaturalReader	☒	☒

Tabulka č. 12 - Tabulka srovnání TTS systémů z hlediska podpory OS. Zdroj: Vlastní zpracování.

Podpora textových souborů

V následující Tabulce č. 13 je zobrazeno, jaká je dostupná podpora textových souborů.

Název společnosti	Podpora textových souborů		
	Nepodporuje	Všechna .doc	Další přípony
Acapela Group	☒	☐	☐
Microsoft	☐	☒	☐
SpeechTech	☐	☒	☐
LinguaTec	☐	☒	☒
NaturalReader	☐	☒	☒

Tabulka č. 13 - Tabulka srovnání TTS systémů z hlediska podpory textových souborů. Zdroj: Vlastní zpracování.

Převod na zvukový soubor

V následující Tabulce č. 14 je zobrazeno, do kolika zvukových formátů je možné převést předčítanou řeč.

Název společnosti	Převod na zvukový soubor			
	Nepodporuje	MP3	WAV	Ogg
Acapela Group	☐	☒	☐	☐
Microsoft	☒	☐	☐	☐
SpeechTech	☐	☒	☒	☐
LinguaTec	☐	☒	☒	☐
NaturalReader	☐	☒	☒	☒

Tabulka č. 14 - Tabulka srovnání TTS systémů z hlediska převodu na zvukový soubor. Zdroj: Vlastní zpracování.

Školení

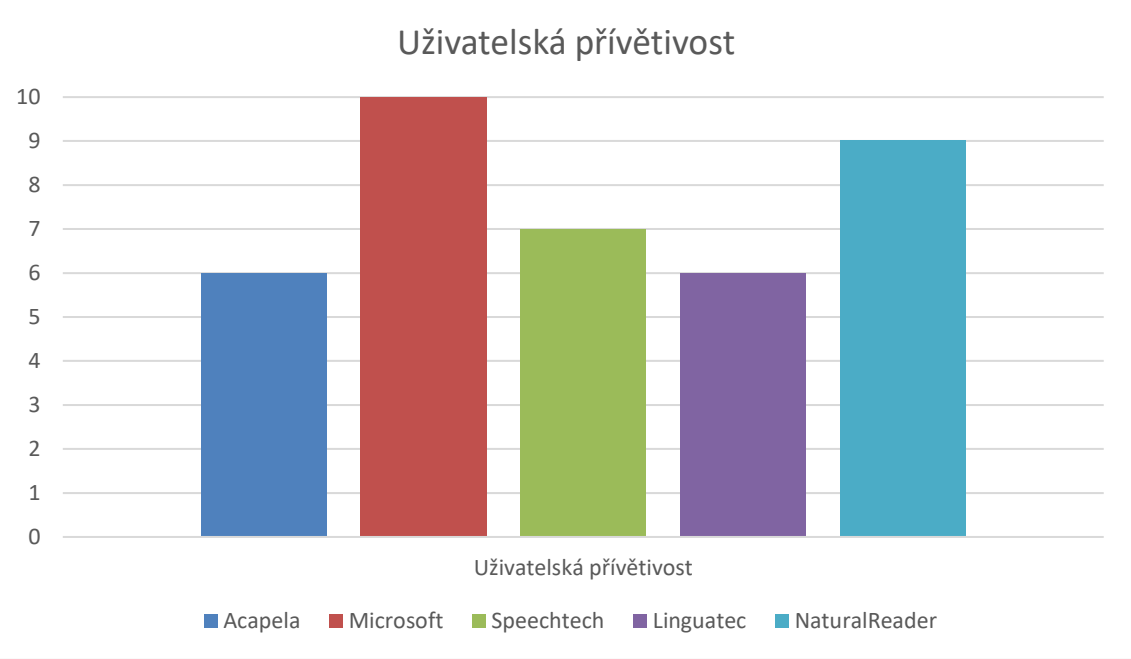
V Tabulce č. 15 níže je zobrazeno, zda je dostupné školení či nikoliv.

Název společnosti	Školení		
	Není možné	Je možné	Je možné po individuální domluvě
Acapela Group	☒	☐	☐
Microsoft	☐	☒	☐
SpeechTech	☐	☐	☒
LinguaTec	☒	☐	☐
NaturalReader	☒	☐	☐

Tabulka č. 15 - Tabulka srovnání TTS systémů z hlediska školení. Zdroj: Vlastní zpracování.

Uživatelská přívětivost

V Grafu č. 7 níže je hodnocena uživatelská přívětivost na škále od 1 do 10 (uvedeno na základě subjektivního názoru autora).



Graf č. 7 - Graf srovnání TTS systémů z hlediska uživatelské přívětivosti. Zdroj: Vlastní zpracování.

Verze pro mobilní zařízení

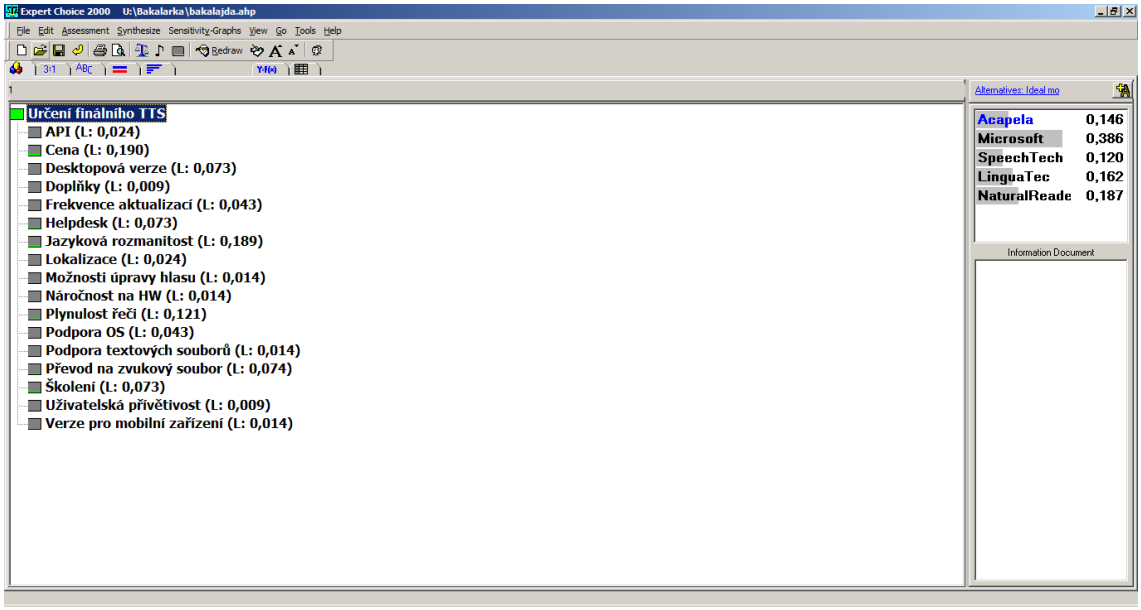
V Tabulce č. 16 níže je zobrazena dostupnost verze pro mobilní zařízení.

Název společnosti	Verze pro mobilní zařízení		
	Nepodporuje	Android	iOS
Acapela Group	☐	☒	☐
Microsoft	☐	☒	☒
SpeechTech	☐	☒	☐
LinguaTec	☒	☐	☐
NaturalReader	☒	☐	☐

Tabulka č. 16 - Tabulka srovnání TTS systémů z hlediska verze pro mobilní zařízení. Zdroj: Vlastní zpracování.

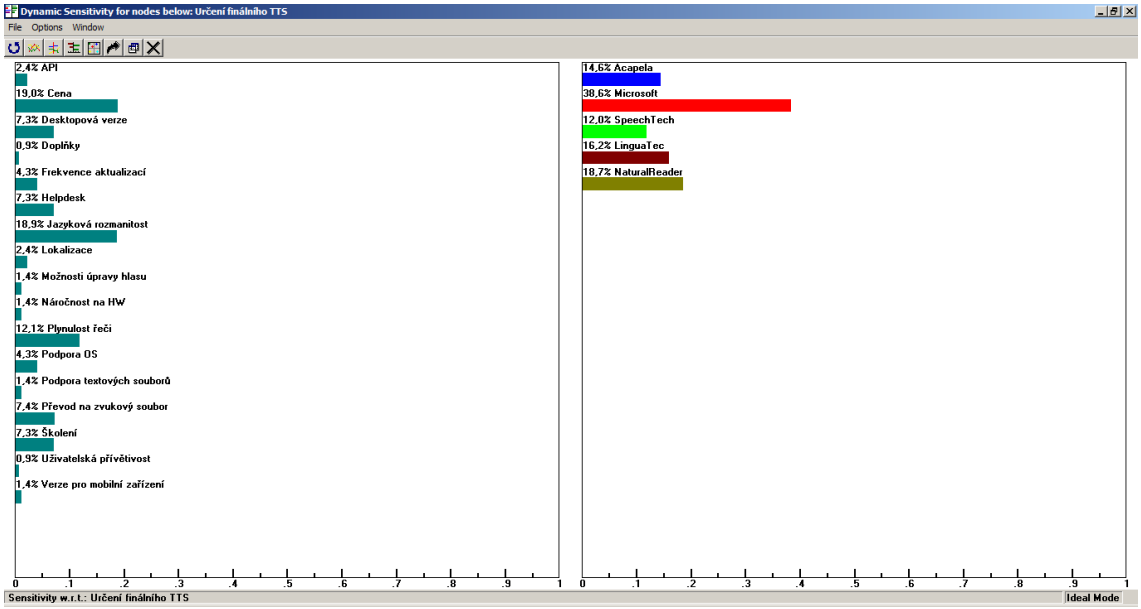
6 Shrnutí výsledků práce

V této kapitole jsou mezi sebou pomocí programu Expert Choice (bližší popis výše) porovnány TTS s důrazem na důležitost všech kritérií (viz Obrázek č. 24). V hodnocení je započítáno i do jaké míry splňuje dané kritérium každý TTS.



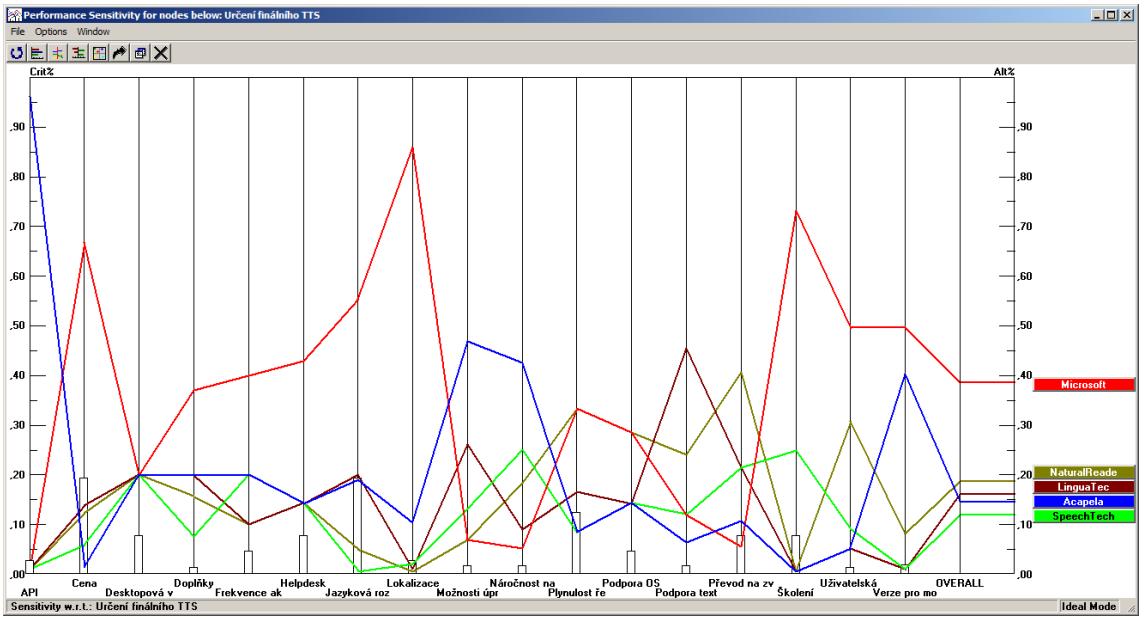
Obrázek č. 24 - Zadaná kritéria a hodnoty – Výstřižek z programu Expert Choice 2000. Zdroj: Vlastní zpracování.

V Grafu č. 8 lze pozorovat na levé straně procentuální zastoupení důležitosti jednotlivých kritérií (součet zadaných vah všech kritérií je 100 %) a na pravé straně jednotlivé TTS programy s počtem procent určujícím jejich vyhovění zadaným kritériím.



Graf č. 8 - Dynamický graf citlivosti – Výstřižek z programu Expert Choice 2000. Zdroj: Vlastní zpracování.

V Grafu č. 9 lze pozorovat jednotlivé rozdíly u konkrétních TTS programů, které jsou zde porovnávány. Každý program je reprezentován odlišnou barvou. Na spodní části grafu jsou uvedena jednotlivá kritéria (označená bílým obdélníkem vystupujícím po vertikální lince určující jejich důležitost pro vyhodnocení výsledků). Jednotlivé křivky (odlišené barvami podle TTS) na těchto vertikálních linkách určují hodnotu programu v daném kritériu (podle výše, které zde křivka dosahuje). Konečný součet všech křivek je zobrazen na pravé straně a udává tak celkové pořadí jednotlivých TTS.



Graf č. 9 - Performační graf citlivosti – Výstřižek z programu Expert Choice 2000. Zdroj: Vlastní zpracování.

Z grafu je patrné, že nejlépe vyhovující TTS podle stanovených kritérií je právě Microsoft (tedy Office 365), druhým je NaturalReader, třetí LinguaTec, čtvrtý Acapela Group a poslední SpeechTech.

Důvodem velkého rozdílu mezi programem od společnosti Microsoft a ostatními společnostmi jsou nízké pořizovací náklady, velké množství jazyků, známé prostředí a možnost školení. Tato kritéria měla největší vliv na celkové výsledky, jak je možné pozorovat z Grafu č. 9. V případě, že by nebyl kladen tak velký důraz na cenu, došlo by k posunu ostatních společností vzhůru a rozdíl v závěrečných výsledcích by se značně snížil. Rovněž lze uvažovat i o lokalizaci a jazykové rozmanitosti, kde by uživatel mohl požadovat pouze český a anglický jazyk. Tím by se ostatní TTS systémy vyrovnaly společnosti Microsoft, jelikož i ony z většiny tyto dva jazyky podporují. Křivky společnosti Microsoft viditelné na grafu by se následkem takové úpravy vyrovnaly s ostatními společnostmi, čímž by došlo ke změně výsledného pořadí vyhovujících TTS systémů.

7 Závěry a doporučení

Tato bakalářská práce s názvem: „Analýza a využití TTS (Text-to-Speech) softwaru pro studijní účely“, má za cíl informovat o existenci TTS a vybrat jeden konkrétní pro případnou implementaci ve vzdělávací instituci. Hlavním důvodem je usnadnění studia a práce v rámci vzdělávacích institucí.

Jak je popsáno v kapitole Shrnutí výsledků práce, nejlépe vyhovujícím TTS je program od společnosti Microsoft – tedy Office 365 se svou funkcí „Imerzivní čtení“. Výsledek je bezpochyby ovlivněn skutečností, že největší důraz byl kladen na cenu, množství dostupných jazyků pro předčítání, možnost školení a znalost daného prostředí. Při uvažování využití TTS systému na Univerzitě Hradec Králové, je velkou výhodou tohoto TTS to, že je plně dostupný pro všechny studenty a pracovníky univerzity zdarma, a to i pro domácí použití, čímž jsou pořizovací náklady prakticky nulové (nulové náklady nebyly započítány v grafickém porovnávání v programu Expert Choice (viz Graf č. 9), kde byly srovnávány pořizovací náklady na jednu licenci pro nekomerční uživatele jedná se pouze o bonusovou informaci spojenou s Univerzitou Hradec Králové). Důležité je i to, že prostředí Office 365 je známé široké veřejnosti a velké množství lidí s ním již pracovalo. Jelikož je Microsoft velká společnost s nemalým kapitálem, má jejich program náskok v počtu dostupné lokalizace a jazyků pro předčítání. To je také důvodem, proč je k dispozici celá řada školení od externích firem. Mezi nevýhodami lze zmínit to, že u Imerzivního čtení Office 365 není možnost převodu na zvukový soubor, vkládání jiných souborů, než od společnosti Microsoft a jiných doplňků jako jsou analyzování a předčítání textu v obrázcích, automatické předčítání webových stránek a podobně.

V případě, že by nebyl kladen důraz na cenu, známé prostředí a byla by dostačující nabídka světových jazyků, byl by s největší pravděpodobností lépe vyhovující některý z ostatních analyzovaných programů. Jako příklad lze uvést TTS od společnosti Acapela Group, který svými funkcemi znatelně převyšuje Imerzivní čtení Office 365.

V návaznosti na výsledky práce a možnosti jejího rozšíření autor zamýšlí zaměřit se na testování TTS a jeho vlivů na učení v porovnání s běžnými učebními

metodami (a to konkrétně pod záštitou společnosti CCNovasoft). Díky výsledkům tohoto zkoumání by bylo možné lépe podložit užitečnost TTS a zvýšit tak povědomí o něm a jeho užívání mezi laickou i odbornou veřejností, a tedy i ve společnostech nebo ve vzdělávacích institucích.

V dnešní době je TTS velmi dobrým řešením v otázkách rovnoprávnosti osob ve společnosti. Žijeme v moderní době, ve které se vyskytují osoby s různým zdravotním postižením nebo znevýhodněním. Právě tento program může pomoci snazší integraci těchto osob ve vzdělávacích (a jiných) institucích i ve zvládání každodenního života mnohem lépe.

TTS je samozřejmě velmi užitečným nástrojem i pro osoby bez zdravotního postižení či jiného znevýhodnění. Za pomoci tohoto chytrého programu lze vytvořit audioknihu z jakékoli publikace. Je tak možné poslouchat studijní materiály či jiné texty na cestách, a to bez nutnosti namáhavého soustředění se na čtení. Tento program dokáže usnadnit práci studentům se studijními materiály, číst skripta, přednášky a učebnice, což vyhovuje zejména osobám, které si lépe pamatují učivo při vnímání sluchem. TTS je ovšem užitečné i pro řidiče při předčítání knih při jízdě autem nebo například při hlášení zpoždění vlaků na vlakových nádražích, a v mnoha dalších oblastech.

8 Seznam použité literatury

[1] ACAPELA GROUP. (2017a) *Profile*. © 2017 Acapela Group. [Online] Dostupné z: http://www.acapela-group.com/company/profile/, [Citace: 12.7.2017]

[2] ACAPELA GROUP. (2017b) *Timeline*. © 2017 Acapela Group. [Online] Dostupné z: http://www.acapela-group.com/company/timeline/, [Citace: 12.7.2017]

[3] ACAPELABOX. (2017) *Logo Acapela Group*. © Acapela Group 2010-2014. [Online] Dostupné z: https://acapela-box.com/AcaBox/Elements/logoAG_XL_transp.png, [Citace: 12.7.2017]

[4] COLLINSDICTIONARY.COM. (2017a) *Definition of 'diphone'.* HarperCollins Publishers, © Collins 2017. [Online] Dostupné z: https://www.collinsdictionary.com/dictionary/english/diphone, [Citace: 12.6.2017]

[5] COLLINSDICTIONARY.COM. (2017b) *Definition of 'formant'*. HarperCollins Publishers, © Collins 2017. [Online] Dostupné z: https://www.collinsdictionary.com/dictionary/english/formant, [Citace: 12.6.2017]

[6] COLLINSDICTIONARY.COM. (2017c) *Definition of 'phoneme'*. HarperCollins Publishers, © Collins 2017. [Online] Dostupné z: https://www.collinsdictionary.com/dictionary/english/phoneme, [Citace: 12.6.2017]

[7] DAGBA, Théophile K., AOGA, John Oscar Raoul, FANOU, Codjo C. (2016) *Design of a Yoruba Language Speech Corpus for the Purposes of Text-to-Speech (TTS) Synthesis*. In: Nguyen Nam-Trung., Trawiński Bogdan., Fujita H., Hong TP. (eds) Intelligent Information and Database Systems. ACIIDS 2016. Lecture Notes in Computer Science, vol 9621. Springer, Berlin, Heidelberg

[8] DUTOIT, Thierry. (1997) *An introduction to text-to-speech synthesis*. Boston: Kluwer Academic Publishers. ISBN 0792344987.

[9] EXPERT CHOICE. (2017a) *Expert Choice – About Us*. © 2017 Expert Choice. [Online] Dostupné z: http://expertchoice.com/about-us/ [Citace: 22.6.2017]

[10] GOLDMANOVÁ, Anna. (2015) *Prozódie.* © Eva Heřmanová, Jitka Černá 2015. [Online] Dostupné z: http://www.artslexikon.cz/index.php?title=Proz%C3%B3die, [Citace: 14.6.2017]

[11] HAWKING.ORG.UK. (2017) *Stephen Hawking Brief Biography*. [Online] Dostupné z: http://www.hawking.org.uk/about-stephen.html, [Citace: 20.6.2017]

[12] HORÁK, Petr. (2009). Ústav fotoniky a elektroniky, v.v.i. AV ČR, Praha. [Online]. Dostupné z: http://www.fit.vutbr.cz/study/courses/ZRE/public/pred/12_synteza_horak/synteza_reci.pdf, [Citace: 18.5.2017]

[13] CHRISTENSSON, Petr. (2016). *API Definition*. TechTerms, © 2017 Sharpened Productions. [Online] Dostupné z: https://techterms.com/definition/api, [Citace: 20.6.2017]

[14] KROTHAPALLI, Sreenivasa Rao. (2012) *Predicting prosody from text for text-to-speech synthesis*. New York, NY: Springer, 2012. ISBN 9781461413387.

[15] KURZY.CZ. (2017) *Kurzy měn – kurzovní lístek ČNB*. © 2000–2017 Kurzy.cz, spol. s r.o., AliaWeb, spol. s r.o. [Online] Dostupné z: http://www.kurzy.cz/kurzy-men/, [Citace: 25.8.2017]

[16] LEMMETTY, Sami. (1999). *Review of Speech Synthesis Technology*. HELSINKI UNIVERSITY OF TECHNOLOGY. [Online]. Dostupné z: http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html, [Citace: 21.5.2017]

[17] LINGUATEC. (2017a) *Linguatec Company Profile*. © 2016 Linguatec. [Online] Dostupné z: https://www.linguatec.de/en/about-us/company-profile/, [Citace: 12.7.2017]

[18] LINGUATEC. (2017b) *Linguatec Logo*. © 2016 Linguatec. [Online] Dostupné z: https://www.linguatec.de/wp-content/uploads/2016/01/linguatec-logo-en-e1435148359145.png, [Citace: 12.7.2017]

[19] LINGUATEC. (2017c) *The Linguatec triangle of language technology*. © 2016 Linguatec. [Online] Dostupné z: https://www.linguatec.de/wp-content/uploads/2016/03/Firmendreieck_2016_en.png, [Citace: 12.7.2017]

[20] MICROSOFT. (2017a) *Profil společnosti Microsoft Česká republika*. Microsoft, © Microsoft 2017. [Online] Dostupné z: https://news.microsoft.com/cs-cz/profil-spolecnosti-microsoft-ceska-republika/#sm.000166wwdd3t2d0xztu20a93fm1kb#EbwZ0UXh58lo2UJi.97, [Citace: 26.6.2017]

[21] MICROSOFT. (2017b) *Použití funkce Přečíst s moduly pro převod textu na řeč pro více jazyků*. Microsoft, © Microsoft 2017. [Online] Dostupné z: https://support.office.com/cs-cz/article/Pou%C5%BEit%C3%AD-funkce-P%C5%99e%C4%8D%C3%ADst-s-moduly-pro-p%C5%99evod-textu-na-%C5%99e%C4%8D-pro-v%C3%ADce-jazyk%C5%AF-e522a4f2-37cb-492b-be6a-8997d23dfe70#__toc272757143, [Citace: 26.6.2017]

[22] MICROSOFT. (2017c) *Logo společnosti Microsoft. Microsoft*, © Microsoft 2017. [Online] Dostupné z: https://blogs.microsoft.com/wp-

content/uploads/2012/08/8867.Microsoft_5F00_Logo_2D00_for_2D00_screen.jpg, [Citace: 28.6.2017]

[23] MICROSOFT. (2017d) *Microsoft Office 365 Logo*. Microsoft, © Microsoft 2017. [Online] Dostupné z: https://ncmedia.azureedge.net/ncmedia/2016/06/microsoft-office-365-logo.jpg, [Citace: 28.6.2017]

[24] NATURALREADER. (2017a) *About us*. © Naturalsoft Ltd. [Online] Dostupné z: https://www.naturalreaders.com/exploring.html?index=2&info=contactus, [Citace: 18.7.2017]

[25] NATURALREADER. (2017b) *Logo NaturalReader*. © Naturalsoft Ltd. [Online] Dostupné z: https://www.naturalreaders.com/images/LOGObig.png, [Citace: 18.7.2017]

[26] OXFORDDICTIONARIES.COM. (2017). *Definition of implementation in English*. Oxford univerzity press, © 2017 Oxford University Press. [Online] Dostupné z: https://en.oxforddictionaries.com/definition/implementation, [Citace: 22.6.2017]

[27] PSUTKA, Josef. (2006) *Mluvíme s počítačem česky*. Praha: Academia, Česká matice technická (Academia). ISBN 80-200-1309-1.

[28] ROUSE, Margaret. (2007) *DEFINITION text-to-speech (TTS)*. © 2003–2017, TechTarget. [Online] Dostupné z: http://searchmobilecomputing.techtarget.com/definition/text-to-speech, [Citace: 14.6.2017]

[29] ROUSE, Margaret. (2009) *DEFINITION voice-to-text*. © 2008–2017, TechTarget. [Online] Dostupné z: http://searchunifiedcommunications.techtarget.com/definition/voice-to-text, [Citace: 20.6.2017]

[30] ROUSE, Margaret. (2017) *DEFINITION Apple iOS*. © 2003–2017, TechTarget. [Online] Dostupné z: http://searchmobilecomputing.techtarget.com/definition/iOS, [Citace: 20.6.2017]

[31] SPEECHTECH. (2017a) *O společnosti*. [Online] Dostupné z: http://www.speechtech.cz/cz/o-speechtech/o-speechtech/, [Citace: 10.7.2017]

[32] SPEECHTECH. (2017b) *Produkty SpeechTech.* [Online] Dostupné z: http://www.speechtech.cz/cz/, [Citace: 10.7.2017]

[33] SPEKTRA.EU. (2017) *Acapela Infovox 4 (Eliška) - hlasový syntetizér*. Spektra.eu, Infovox 4 © Acapela Group. [Online] Dostupné z: http://www.spektra.eu/cs/zrakove-vady/programy/syntezatory/acapela, [Citace: 26.6.2017]

Oskenované zadání práce

Univerzita Hradec Králové
Fakulta informatiky a managementu
Akademický rok: 2016/2017

Studijní program: Systémové inženýrství a informatika
Forma: Prezenční
Obor/komb.: Informační management (im3-p)

Podklad pro zadání BAKALÁŘSKÉ práce studenta

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
Maliniak Vladimír	Kuttelwascherova 926/11, Praha - Černý Most	I14381

TÉMA ČESKY:

Analýza a využití TTS (Text to Speech) softwaru pro studijní účely

TÉMA ANGLICKY:

Analysis and Usage of TTS (Text to Speech) Software for Study Purposes

VEDOUCÍ PRÁCE:

Ing. Tereza Otčenášková, B.A. - KIT

ZÁSADY PRO VYPRACOVÁNÍ:

Cíl práce:
Cílem práce je představit software Text to Speech a jeho možné využití jak pro osobní, tak pro studijní účely.

Osnova:
1. Úvod
2. Cíl práce a metodika
3. Teoretická část
4. Praktická část
5. Shrnutí výsledků a diskuse
6. Závěr a doporučení
7. Zdroje
8. Přílohy

SEZNAM DOPORUČENÉ LITERATURY:

- SPROAT, Richard William (ed.). Multilingual text-to speech synthesis: the bell labs approach. Dordrecht: Kluwer Academic Publishers, 2001. ISBN 0-7923-8027-4.

- TAYLOR, Paul. Text-to-speech synthesis. Cambridge, UK: Cambridge University Press, 2009. ISBN 0521899273.

- HOLMES, J. N. a Wendy HOLMES. Speech synthesis and recognition. 2nd ed. New York: Taylor & Francis, 2001. ISBN 0748408576.

- BY THIERRY DUTOIT. An introduction to text-to-speech synthesis. Dordrecht [u.a.]: Kluwer, 2001. ISBN 9781402003691.

- Zhao, Y. (2007). Speech technology and its potential for special education. Journal of Special Education Technology, 22(3), 35-41. Retrieved from http://search.proquest.com/docview/228523470?accountid=14623

- Pathak, P. (2010). Speech recognition technology: Applications & future.International Journal of Advanced Research in Computer Science, 1(4) Retrieved from http://search.proquest.com/docview/1443704002?accountid=14623

- Flood, J. (2007). NaturalReader: A new generation text reader. Developmental Disabilities Bulletin, 35(1), 44-55. Retrieved from http://search.proquest.com/docview/232466316?accountid=14623

- Lange, A. A., McPhillips, M., Mulhern, G., & Wylie, J. (2006). Assistive software tools for secondary-level students with literacy difficulties. Journal of Special Education Technology, 21(3), 13-22. Retrieved from http://search.proquest.com/docview/228487206?accountid=14623

(c) IS/STAG , Portál - Podklad kvalifikační práce , I14381 , 13.10.2016 10:29

Podpis studenta: ..... Datum: 13.10.2016

Podpis vedoucího práce: ..... Datum: 13.10.2016

(c) IS/STAG , Portál - Podklad kvalifikační práce , II4381 , 13.10.2016 10:29