Univerzita Pardubice
Fakulta-ekonomicko správní

Testování softwarových nástrojů pro převod textu na mluvenou řeč

Ivana Linhartová

Bakalářská práce

2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: **2008/2009**

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení:	**Ivana LINHARTOVÁ**
Studijní program:	**B6209 Systémové inženýrství a informatika**
Studijní obor:	**Regionální a informační management**
Název tématu:	**Testování softwarových nástrojů pro převod textu na mluvenou řeč**

Zásady pro vypracování:

1. Popis technologie pro převod textu na řeč
2. Softwarové nástroje pro převod textu na řeč
3. Návrh vhodného způsobu testování
4. Realizace testování a vyhodnocení výsledků

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

[1] Text to Speech : Alive Text to Speech reads text and converts text to MP3, WAV, OGG or VOX files [online]. 2003-2008 [cit. 2008-06-29]. Dostupný z WWW: <http://www.text-speech.com/>.

[2] ReadPlease : Text-to-speech software that lets your computer talk [online]. 1999-2005 [cit. 2008-06-29]. Dostupný z WWW: <http://www.readplease.com/>.

[3] CS-VOICE 97 : Stránky FROG Systems [online]. 1995-2003 [cit. 2008-06-29]. Dostupný z WWW: <http://www.frog.cz/prod04.htm>.

[4] Text to Speech Software with AT&T Natural Voices, NeoSpeech, Acapela, Nuance, SAPI Voices [online]. 2000-2008 [cit. 2008-06-29]. Dostupný z WWW: <http://www.nextup.com/>.

[5] Download Free Text-to-Speech Software [online]. 1995-2007 , 10. 05. 2008 [cit. 2008-06-29]. Dostupný z WWW: <http://www.dyslexia.com/helpread.htm>.

[6] Encyklopedie - NAVAJO : Syntéza řeči [online]. 2006 [cit. 2008-06-29]. Dostupný z WWW: <http://synteza-reci.navajo.cz/>.

[7] Wikipedia - the free encyklopedia : Speech synthesis [online]. [2000] , 24. 06. 2008 [cit. 2008-06-29]. Dostupný z WWW: <en.wikipedia.org/wiki/Speech_synthesis>.

[8] FEKT VUT - Fakulta - Magisterské projekty : TEXT TO SPEECH SYSTEM [online]. 2006 [cit. 2007-06-29]. Dostupný z WWW: <http://www.feec.vutbr.cz/EEICT/2006/sbornik/02-Magisterske_projekty/08-Grafika_a_multimedia/04-xoczko00.pdf>.

Vedoucí bakalářské práce: **Ing. Milan Tomeš**
Ústav systémového inženýrství a informatiky

Datum zadání bakalářské práce: **6. října 2008**

Termín odevzdání bakalářské práce: **1. května 2009**

doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.

doc. Ing. Jiří Křupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 27. 04. 2009

Ivana Linhartová

SOUHRN

Bakalářská práce se zabývá převodem textu na mluvenou řeč (tzv. syntézou řeči). První část práce je věnována popisu technologie převodu textu na řeč. Popsány jsou zde i jednotlivé základní přístupy při modelování řeči a představení současných českých i zahraničních TTS systémů. Ve druhé části, která je zaměřena hlavně na praktické činnosti, je navržen vhodný způsob testování jednotlivých TTS systémů a tento způsob je následně zrealizován. Závěr práce se zabývá vyhodnocením získaných výsledků.

KLÍČOVÁ SLOVA

komunikační proces, syntéza řeči, formantová syntéza, konkatenační syntéza, artikulační syntéza, řečový syntetizér, Text-to-Speech (TTS) systém

TITLE

Testing software tools for text to speech conversion

ABSTRACT

The Bachelor work analyses text transforming to speech (so called speech synthesis). The first part describes technology of text transforming to speech. The second part deal with practical terms experience approaches and speech modelling and introduction of contemporary Czech and foreign TTS systems have been described. Appropriate way of TTS system testing is which is additionally realised. Evaluated results gathered gradually are shown in the conclusion.

KEYWORDS

Communication process, speech synthesis, formant synthesis, concatenate synthesis, articulated synthesis, speech synthesis, Text-to Speech (TTS) system.

Obsah:

Seznam obrázků:

Seznam tabulek:

1 Úvod

V dnešním světě se počítače stávají nedílnou součástí každodenního života. Obklopují nás snad všude, proto není divu, že se člověk snaží o zdokonalování komunikace s ním tou nejpřirozenější cestou - komunikace řečí. Tato komunikace probíhá dvěma směry, prvním je komunikace člověka s počítačem a druhým komunikace počítače s člověkem. Tato práce se zabývá zejména komunikací počítače s člověkem, konkrétně převodem textu na mluvenou řeč (syntézu řeči). Cílem je popsání technologie syntézy řeči, následné představení, testování a vyhodnocení současných softwarových nástrojů pro převod textu na řeč (syntetizérů řeči). Důvodem výběru tohoto tématu je upozornění na problematiku, která je nedílnou součástí každodenního života určité skupiny lidí (zejména nevidomých).

V první kapitole je obecně popsán komunikační proces a s ním spojené pojmy artikulace a percepce řečového signálu. Dále kapitola obsahuje základní dělení řeči a jsou zde zmíněny jednotlivé vědy, které se zabývají zpracováním řeči.

Druhá kapitola obsahuje historii syntézy řeči, samotnou technologii převodu textu na řeč a jednotlivé základní přístupy syntézy řeči, mezi které je řazena formantova syntéza, konkatenační syntéza a artikulační syntéza. Dále tato kapitola popisuje jednotlivé řečové syntetizéry a to jak české, tak i zahraniční.

Třetí kapitola se zabývá realizací testování řečových syntetizérů. V první části této kapitoly je navrhnut způsob testování TTS systémů. Ve druhé části jsou popsány použité testy a ve třetí části jsou vyhodnoceny výsledky testování.

Závěrečná část práce vychází ze získaných poznatků a ze zkušeností z předchozího zkoumání.

2 Komunikační proces

Život v lidské společnosti je závislý na schopnosti jednotlivců komunikovat a vzájemně si sdělovat informace. Informace v abstraktní formě zpracovává lidský mozek a v okamžicích, kdy dospěje k potřebě informace sdílet s jinými lidmi, formuluje s využitím jazykových prostředků sdělení adresované jinému člověku, nebo skupině lidí – dochází ke komunikaci. Prostředky vzájemné komunikace můžeme v současnosti charakterizovat dvěma cestami – artikulovanou lidskou řečí a písemnou komunikací. Obě tyto cesty jsou vázány na určitý (národní) jazyk. Komunikaci umožňují i další prostředky neverbální a mimojazykové, jako jsou neartikulované hlasové projevy a motorika. Můžeme sem zahrnout i kódování informace do vizuální podoby jako jsou kouřové signály, znaková řeč neslyšících a indikátory na panelech přístrojů. Komunikace s přístroji a technologickými zařízeními je dosud většinou omezena na mechanické ovládací prvky – páky, tlačítka, klávesnice, obrazové a textové displeje apod. Je přirozenou snahou konstruktérů technických zařízení zlepšovat podmínky pro jejich řízení, respektive komunikaci s nimi. Hlasová komunikace tedy není již jen záležitostí sdělování mezi lidmi, ale bude se stále více stávat součástí *komunikace člověk – stroj*. [1]

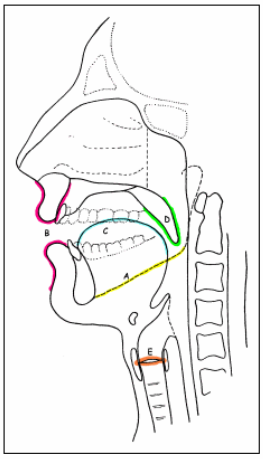
2.1 Artikulace a percepce

Pro úvahy o technickém řešení systémů podporujících hlasové komunikace je potřebné znát fyzikální podstatu artikulace (vytváření) a percepce (vnímání) řečového signálu.

- **Artikulace** je založena na změnách akustických vlastností hlasového traktu. Je řízena lidskou vůlí. V hlasovém traktu vzniká akustická vlna, která se od mluvčího šíří volným prostorem. Primárním zdrojem energie této akustické vlny je proud vzduchu vyháněný z plic. Veškeré artikulované projevy v češtině jsou vytvářeny výdechovým proudem vzduchu. Proud vzduchu vyháněný z plic vytváří slyšitelné artikulované zvuky dvěma základními principy, dle [1] tyto:
  - *Znělé* úseky promluvy jsou vytvářeny tak, že proud vzduchu prochází sevřenými hlasivkami, které vibrují a tak vytvářejí sled impulzů vstupujících do dutiny hrdelní, ústní a nosní. Tyto dutiny se chovají jako rezonátory s různými vlastními

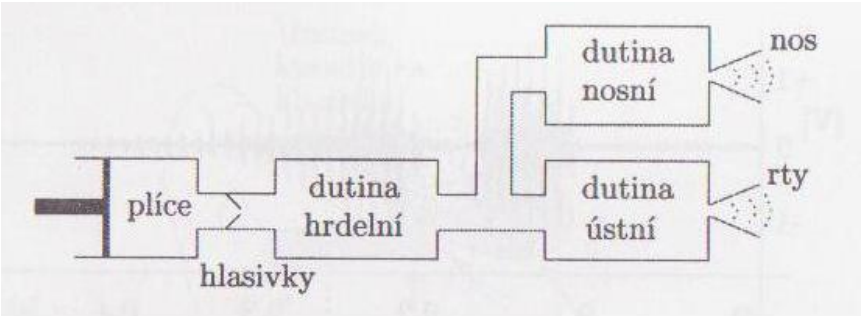
rezonančními vlastnostmi. Znělou promluvu tedy člověk artikuluje tím, že svalovou činností mění tvar uvedených rezonátorů.

– *Neznělé* úseky promluvy vznikají bez účasti hlasivek. Výdechový proud vzduchu je v hlasovém traktu ovládán tak, že v určitých místech nastavené překážky vytvoří slyšitelné turbulence, jejichž akustický projev lze charakterizovat jako širokopásmový šum, který může být více či méně ovlivněn ve svých výsledných vlastnostech průchodem uvedenými dutinami. Jiné segmenty řeči vznikají přerušováním hlasivkami modulovaného, či nemodulovaného vzduchového proudu jazykem nebo rty. Hlasové ústrojí člověka je znázorněno na obr. č. 1.



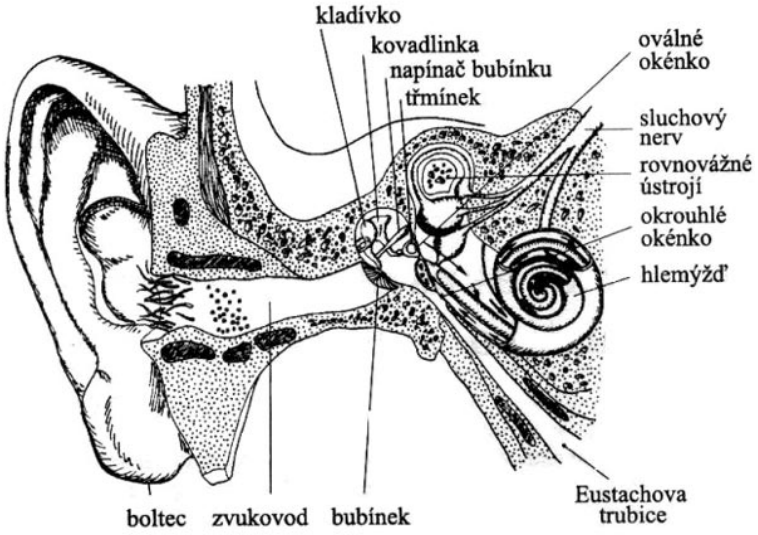
Obr. 1 Artikulační orgány, zdroj: [2]

Hlasovým ústrojím modulovaný proud vzduchu vystupuje přes rty a nosem do vnějšího prostředí. Ve vzduchu se vytvoří zvuková vlna, kterou zachytí sluchové ústrojí posluchače. Neméně významné je, že mluvčí svým sluchem kontroluje vlastní artikulaci a na principu zpětné vazby ji koriguje do podoby, kterou považuje za odpovídající zamýšlenému sdělení. Na dalším obrázku je znázorněn velmi zjednodušený akustický model artikulačního ústrojí. [1]



Obr. 2 Akustický model artikulačního aparátu, zdroj: [1]

- **Percepce** neboli zpracování zvukového vjemu je ve sluchovém ústrojí založena na převodu mechanických pohybů bubínku na podráždění nervových zakončení nervů vedoucích ze sluchového ústrojí do mozku, viz obr. č. 3. [1]



Obr. 3 Struktura ucha, zdroj: [3]

Převodní mechanizmus výchylek bubínku je „sestaven“ z *kladívka, kovadlinky* a *třmínku* tak, že sluchový orgán je schopen zprostředkovat poslech ve velmi širokém dynamickém rozpětí hlasitosti. Nepodstatnějším zjištěním je však to, že mechanické kmity jsou analyzovány v *hlemýždi*, který má na obvodu zužujícího se profilu své dutiny ohromné množství nervových zakončení. Z toho se usuzuje, že prvotní informace vedená do mozku je založena na vyhodnocení energie signálu ve frekvenčním spektru, protože stavba hlemýždě umožňuje rozlišit frekvenční složky a aktivovat v závislosti na rozkmitu v jeho jednotlivých místech příslušná nervová zakončení. Dostáváme tak argument z oblasti fyziologie sluchu proto, aby se při zpracování řečových signálů zabývalo kromě časového průběhu signálu také časovým průběhem krátkodobého (v čase se měnícího) spektra. Modelům percepce řeči jsou věnovány rozsáhlé publikace, které se zabývají vystižením nejrůznějších specifik sluchu. [1]

Zjednodušeně lze tedy říci, že artikulace představuje transformaci sdělení na řečový signál, který přenášenou informaci reprezentuje měnícím se rozložením amplitud složek krátkodobého spektra a že příjemce takto zakódovanou informaci interpretuje s využitím spektrální analýzy. [1]

2.2 Základní dělení řeči

Řeč je tvořena navazujícími segmenty, které vznikají měnícími se průřezy trubic a poloh překážek hlasového traktu. Tyto segmenty, jakožto nejmenší foneticky odlišné jednotky se nazývají *fonémy*. Z fonémů pak lze poskládat jednotlivé slabiky a z těch následně celá slova promluvy. Hlasovým projevem fonému je hláska. Čeština používá 44 různých fonémů. Pokud vezmeme v úvahu průběh vzniku řeči, je nutné si uvědomit, že jednotlivá slova vznikají změnou parametrů hlasového traktu. Tato změna ale není skoková, což je dáno silou svalů a setrvačností při přechodu traktu z jednoho stavu do druhého. To se odrazí na jednotlivých hláskách jako tzv. koartikulace, čili vzájemné ovlivnění předchozí a následující hlásky. Hláska tak díky této vazbě zní v závislosti na sousedních hláskách pokaždé jinak. Proto lze vedle fonémů využívat i jednotky, které umožňují tyto vlivy postihnout. Používají se především *difóny* (posloupnost samohláska-souhláska) a *trifóny* (foném v závislosti na levém a pravém sousedu). Jejich výhodou je právě postih koartikulace, nevýhodou pak je jejich velký počet (v případě trifónů v řádu tisíců). [4]

2.2.1 Vědy ve zpracování řeči

Zpracování řeči je pluri-disciplinární obor využívající poznatků věd přírodních, technických a humanitních, které vychází z [4, 5] a jsou to tyto:

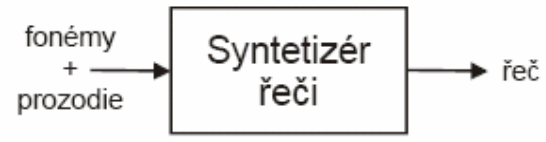
- *akustika* – věnuje se fyzikálním mechanismům tvorby a slyšení řeči
- *fonetika* – zabývá se tvořením a slyšením zvukové stránky řeči
- *fonologie* – věda o fonémech a jejich systému v daném jazyce
- *fysiologie* – zabývá se funkcí hlasového a sluchového ústrojí, pomoc při tvorbě různých modelů
- *gramatika* (mluvnice) – soubor pravidel o obměnách slov a jejich spojování ve věty, důležité pro syntézu
- *lexikologie* – věda o slovní zásobě jazyka, jejím vývoji a vztazích mezi slovy (synonyma, antonyma, homonyma), pomoc při tvorbě slovníků (i počítačových), důležité pro rozpoznávání řeči
- *morfologie* – zabývá se skladbou slov (odděleně od ostatních slov)
- *pragmatika* – zkoumá souvislost sdělení a záměru řečníka
- *prozodie* – věda o zvuková stránce jazyka (melodie, trvání hlásek, přízvuk ve slovech a ve větách)

- *sémantika* – věnuje se významu jazykových jednotek, vychází obvykle od slova, důležitá pro porozumění řeči
- *syntaxe* – nauka o větách a souvětích, část gramatiky

3 Syntéza řeči

Syntéza řeči představuje důležitou oblast problematiky zpracování řečového signálu. Je předmětem intenzivního zkoumání již po dlouhá léta. Jde o proces, při němž se uměle vytváří řeč. V případě počítačové syntézy se řeč uměle vytváří počítačem. Důvodem tak vysokého zájmu je skutečnost, že řeč je nejpřirozenější formou komunikace mezi lidmi. Umělé vytváření řeči počítačem si klade za cíl „zpřirozenit" komunikaci člověka s počítačem a stát se tak rovnocenným partnerem tradiční vizuální komunikace. Zařízení, které proces vytváření řeči provádí, nazýváme *syntetizér řeči*. Syntetizér je jádrem každého systému převodu textu na řeč, viz kap. 3.2. Konečným cílem syntézy řeči je vytvářet řeč v takové formě a kvalitě, aby nebyla rozpoznatelná od řeči člověka. Syntetická řeč by tedy neměla působit jednotvárně, měla by znít přirozeně a její poslech by neměl unavovat ani vyžadovat zvýšenou pozornost. [6]

Na syntetizér řeči, který je zjednodušeně znázorněn na obr. č. 4, se můžeme dívat jako na systém, který na základě vstupní informace vytváří řeč. Onou vstupní informací bývá společná fonetická a prozodická informace o promluvě, která se má generovat. Fonetická informace je přitom reprezentována posloupností hlásek a popisuje, jaká řeč se má vytvořit. Prozodická informace definuje průběhy základních prozodických charakteristik v rámci generované promluvy a popisuje, jak se má výsledná řeč vytvářet. Fonetické i prozodické vlastnosti řeči se nejčastěji odhadují z prostého textu, což si blíže popíšeme v kapitole 3.2. [6]



Obr. 4 Zjednodušené schéma typického syntetizéru, zdroj: [5]

Syntéza řeči může být realizovaná v buď hardwarovými zařízeními anebo softwarovou syntézou. Většina uživatelů dá v dnešní době nepochybně přednost softwarové syntéze, kterou lze pořídit a zprovoznit bez specializovaného hardwarového zařízení a stačí pro ni jakákoliv provozuschopná zvuková karta. [7]

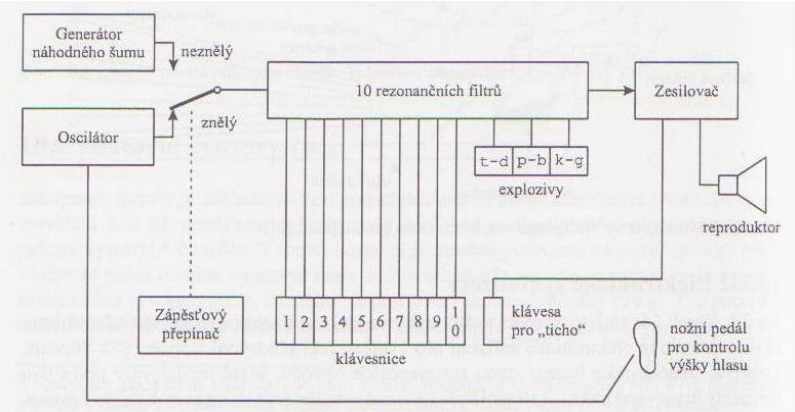
3.1 Historie

Schopnost mluvit je základním komunikačním prostředkem mezi lidmi. Není divu, že vytváření řeči již odedávna zajímalo naše předchůdce. Postupně se objevovaly první pokusy vytvářet řeč uměle, ovšem o této skutečnosti toho nebylo moc zjištěno. Důvod je nejspíše v tom, že až donedávna nebylo zcela jasné, k čemu by se přístroje syntetizující lidskou řeč používaly. První pokusy o syntézu řeči se objevují před dvěma sty lety, kdy ruský profesor *Christian Kartzenstein* na základě svého bádání v oblasti samohlásek sestrojil přístroj, který byl schopen tyto samohlásky uměle reprodukovat. Vytvořil jakési rezonátory, které se rozechvěly v důsledku proudění vzduchu. O několik let později přišel se svým vlastním vynálezem *Wolfgang von Kempelen*, který pracoval ve službách Marie Terezie. Zabýval se původně lidskou řečí, s níž souvisely léčebné kůry mysli a vědomí. Von Kempelenův přístroj byl zcela mechanický avšak první, který uměl vytvářet nejen části lidské řeči, ale i celá slova a někdy i věty, což na tu dobu dá se říci byla unikátní záležitost. Rekonstrukce tohoto „mluvícího stroje“ je znázorněna na obr. č. 5. Na přelomu osmnáctého a devatenáctého století přišel *Charles Wheatstone* se svojí verzí von Kempeleho zařízení. Tato verze byla dokonalejší a schopná produkovat samohlásky a většinu souhlásek. [8]

Obr. 5 Mluvící stroj Wolfganga von Kemplelena, zdroj: [9]

V průběhu dalších let žádné experimenty nepřinesly nijak závažné a zlomové výsledky. Prvním syntetizérem, který zaujal větší pozornost, byl v roce 1939 *VODER* od *Homera Dudleyho*, viz obr. č. 6. Tento syntetizér byl již elektronický a tudíž výsledný zvuk nebyl analogový. Jeho největší nevýhodou byla složitá obsluha. Jak léta běžela, snažili se různí technici a vynálezci o zdokonalení syntézy řeči a v důsledku toho se v rámci různých výzkumů přišlo na řadu dalších analýz. S trochou nadsázky se dá tvrdit, že až do počátku

počítačové éry neměly syntetizéry valnou šanci na úspěch. S nástupem výpočetní techniky se tato situace změnila. Složitá analýza řeči se v současnosti stává díky softwarovým nástrojům doslova hračkou, tedy alespoň co se porovnání s minulostí týče. Nutno ovšem podotknout, že ani moderní počítače problém nedokonalosti umělé řeči neřeší. Produktem softwarových syntetizérů je zatím bohužel řeč, která se pouze blíží mluvenému projevu školeného řečníka, což je od počátku cílem všech konstruktérů, vynálezců a techniků pracující v tomto oboru. [8]

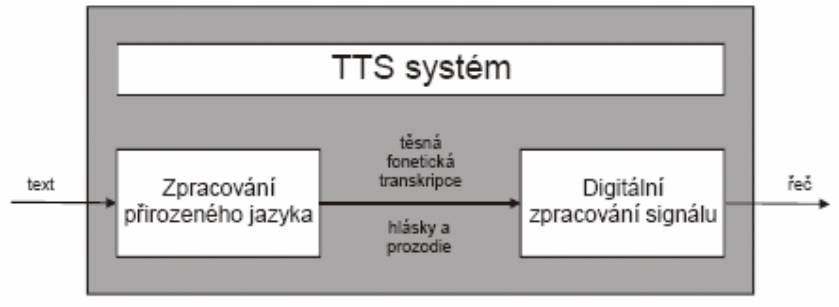


Obr. 6 Blokové schéma Voderu, zdroj: [6]

3.2 Syntéza řeči z textu

Tato kapitola se zabývá samotným převodem textu na řeč, kde již nemáme k dispozici úplnou informaci o tom, co a jak se má syntetizovat. Touto informací je nejčastěji posloupnost hlásek daného jazyka doplněná o prozodické značky (konturu výšky hlasu, trvání a hlasitost hlásek). Nejčastěji je nutno se na vstupu systému spokojit „pouze“ s informací ve formě textu. V takovém případě pak mluvíme o syntéze řeči z textu a systém převodu textu na řeč nazýváme syntetizérem řeči z textu. Syntéza řeči z textu nebo též konverze textu na řeč (text-to-speech, zkr. TTS) je nejobecnější a také nejkomplikovanější úlohou počítačové syntézy řeči. Na obr. č. 7 je znázorněno zjednodušené schéma obecného systému TTS. Skládá se ze dvou hlavních modulů: *zpracování přirozeného jazyka (natural language processing, zkr. NLP)* a *vlastního syntetizéru řeči*. Oba moduly jsou na sobě relativně nezávislé a oba mají svou důležitost. Modul NLP má za úkol zpracovat text na vstupu systému, analyzovat ho a získat pokud možno co nejvíce informací o tom, co se má syntetizovat. Tento modul provádí tzv. těsnou fonetickou transkripci textu. Na výstupu se potom často objevuje posloupnost hlásek příslušného jazyka doplněná o prozodické značky. Pak přichází na řadu syntetizér řeči,

který na základě této informace generuje řeč, tj. provádí vlastní syntézu řeči na signálové úrovni. [6]

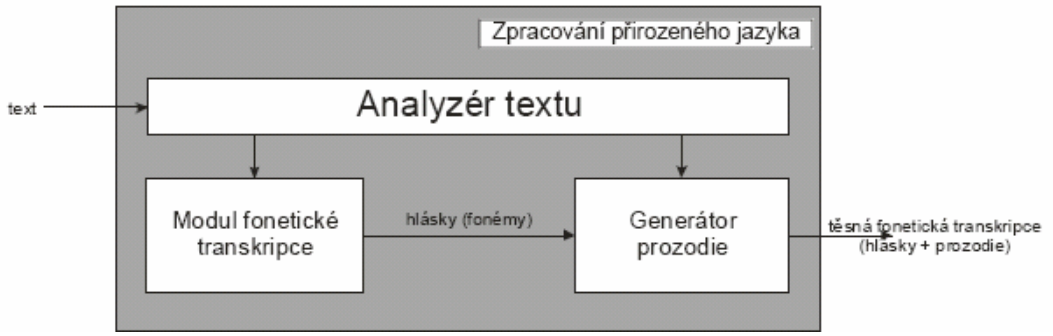


Obr. 7 Zjednodušené schéma systému konverze textu na řeč, zdroj: [5]

3.2.1 Zpracování přirozeného jazyka

Vlastním zpracování přirozeného jazyka se zabývá výpočetní lingvistika, která pracuje s takovými pojmy jako gramatika, inference (odvozování), rozbor nebo transdukce (přenos). Zpracováním českého jazyka se zabývají především v Ústavu formální a aplikované lingvistiky Univerzity Karlovy v Praze a na Fakultě informatiky Masarykovy univerzity v Brně. [6]

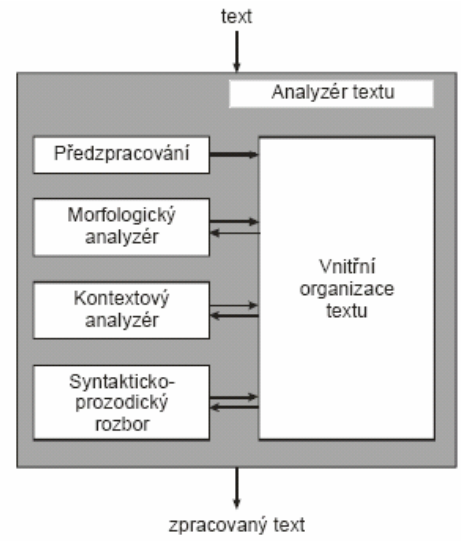
Tato kapitola se věnuje zpracováním přirozeného jazyka v kontextu zpracování vstupní textové informace v systému TTS. Syntetizér řeči potřebuje znát těsnou fonetickou transkripci textu, kterou má syntetizovat, tj. nejčastěji posloupnost hlásek a prozodických značek. Modul NLP obsahuje dva bloky, viz obr. č. 8: *modul fonetické transkripce (letter-to-sound, zkr. LTS)* a *generátor prozodie (prosody generator, zkr. GP)*. Do modulu NLP je vhodné doplnit další blok, *morfologicko-syntaktický analyzátor (morpho-syntactit analyzer, zkr. MSA),* který provádí morfologickou a syntaktickou analýzu vstupního textu. [6]



Obr. 8 Zjednodušené schéma modulu zpracování přirozeného jazyka, zdroj: [5]

- **Morfologicko-syntaktická analýza (analýza textu)**

  Kvalitní zpracování textu v systému TTS se neobejde bez důkladné morfologicko-syntaktické analýzy (MSA) vstupního textu. MSA odhaluje vnitřní strukturu každé věty – reprezentuje větu jako posloupnost mluvnických kategorií jednotlivých slov. Morfologicko-syntaktický analyzátor moderních systémů TTS je znázorněn na obr. č. 9. V současných systémech TTS nejsou jednotlivé komponenty řazeny sekvenčně, ale spíše paralelně. Každá z komponent totiž může do zpracovávaného textu v průběhu zpracování přidat novou informaci, ostatním komponentám do té doby nepřístupnou. Morfologicko-syntaktický analyzátor se obvykle skládá z komponent, které jsou znázorněny na zmiňovaném obr. č. 9. [6]



Obr. 9 Schéma morfologicko-syntaktického analyzátoru, zdroj: [5]

- **Fonetická transkripce**

  Úkolem fonetické transkripce je popsat výslovnostní podobu řeči. Fonetickou transkripci lze v zásadě provádět dvojím způsobem, a to ručně nebo automaticky. Ručně se vytváří na základě textu (zpracovaného pomocí MSA) nebo přímo pomocí řečového signálu promluvy. Jde však o pracný a časově náročný proces, který navíc mohou vykonávat pouze fonetičtí experti. Automaticky se fonetická transkripce provádí bez přímé účasti člověka-experta, a to v zásadě dvěma způsoby, dle [5, 6]:

  - na základě **fonetického slovníku**, který obsahuje ortografickou (textovou) a fonetickou podobu slova. Tento přístup je vhodný spíše pro analytické jazyky (např. angličtinu), ve kterých neexistuje mnoho tvarů odvozených od stejného slova. Díky tomu lze dosáhnout rozumné velikosti fonetického slovníku.

- pomocí **fonetických transkripčních pravidel**. Tento přístup je vhodný zejména pro flexivní jazyky (např. češtinu), v nichž existuje velké množství tvarů odvozených od stejného slova a z důvodu obrovského množství slov nebývá využití fonetického slovníku efektivní. V tom případě je třeba najít obecná pravidla, pomocí nichž je možné fonetický přepis promluvy provádět automaticky. Pravidla je možné přitom navrhnout analyticky (např. jakýsi expertní systém).

Oba přístupy se často využívají současně. V češtině se pro domácí slova většinou používají pravidla a slova cizího původu se pak uchovávají ve fonetickém slovníku výjimek. Vstupem modulu fonetické transkripce bývá „čistý" text zpracovaný modulem MSA, tj. text s rozepsanými zkratkami, číslicemi apod., a doplněný o morfologické, kontextové a syntakticko-prozodické informace. Pokud by některá informace chyběla, zpracovaný text nemusí být z hlediska automatické fonetické transkripce jednoznačný a výsledek automatické fonetické transkripce nemusí být úplně správný. [6]

3.2.2 Generování prozodie

Srozumitelnost a zejména přirozenost řeči ve značné míře závisí na prozodických charakteristikách obsažených v řeči. Mezi důležité prozodické charakteristiky řeči patří: intonace (melodie), časování a intenzita (hlasitost). Automatickému generování prozodie by tedy měla být při vývoji systému TTS věnována patřičná pozornost. Bohužel jde o extrémně složitou úlohu, neboť různorodost akustického vyjádření prozodie je částečně nezávislá na textové reprezentaci promluvy, což do značné míry ztěžuje úlohu odhadu prozodického vyjádření promluvy v úloze syntézy řeči z textu. V následujících odstavcích jsou velice stručně popsány metody generování prozodie, které vychází z [5, 6]:

- **Generování intonace**

  Intonace (melodie) promluvy souvisí s průběhem frekvence základního hlasivkového tónu (F_0), resp. výškou hlasu. Generování intonace v úloze syntézy řeči z textu je možné obecně popsat ve dvou krocích. První krok zahrnuje vytvoření symbolického popisu intonace (resp. kompletní prozodie) promluvy jako doplněk k ortografické a fonetické transkripci promluvy a v druhém kroku pak zvolený intonační model na základě tohoto popisu generuje výslednou intonační konturu (konturu F_0). Použitý symbolický popis prozodie přitom zpravidla předurčuje intonační modely vhodné pro generování intonace. Všechny symbolické popisy prozodie jsou charakteristické snahou co nejvěrněji popsat

průběh prozodie dané promluvy. Mezi nejznámější symbolické popisy intonace, tzv. *intonační fonologii* patří ToBI, INTSINT, Tilt a další. Často se také intonační průběhy popisují pomocí lingvisticky motivovaných pravidel, jelikož je to nejjednodušší generátor průběhu intonace. O generování výsledného průběhu základního hlasivkového tónu se starají intonační modely. Intonační modely používané v současných systémech TTS je možné rozdělit podle oblasti, ve které se s intonací pracuje, a to na:

- *Akustické modely intonace* – tyto modely jsou nejpoužívanější a vycházejí z akustické reprezentace intonace.
- *Percepční modely intonace* – tyto modely pracují s percepční reprezentací intonace (tj. na úrovni vnímání intonace posluchačem).
- *Lingvistické modely intonace* – jde o nejsložitější modely intonace, neboť vycházejí z obecné lingvistické reprezentace prozodie.

- **Generování intenzity**

  Základními jednotkami pro generování intenzity jsou především hlásky. Jejich intenzita se obvykle určuje pomocí statistik na základě rozsáhlých řečových korpusů, často ve spojení se specifickými informacemi o pozici hlásky ve slově, větném úseku či ve větě, přízvučnosti apod. V české řeči bývá intenzita považována za nejméně významnou prozodickou charakteristiku.

- **Generování časování**

  Problematika generování časování v systémech převodu textu na řeč zahrnuje odhad všech aspektů spojených s časováním řeči, tj. s trváním segmentů řeči, umísťováním pauz, rytmickým členěním promluv pomocí přízvuků atd.

  - **generování trvání** – základními jednotkami modulace trvání jsou v současných systémech TTS opět především segmenty na úrovni fonémů. Modely trvání těchto segmentů využívají znalostí o artikulačních a fonologických aspektech těchto segmentů ve formě pravidel nebo statistik spočtených z rozsáhlých řečových korpusů.
  - **generování pauz** – pro generování pauz se využívají informace o syntakticko-prozodické struktuře věty, neboť pauzy se umísťují na hranice mezi některými frázemi. Mohou být různého trvání: mezi dvěma úseky se slabší textovou koherencí se umísťuje delší pauza a naopak mezi dva úseky se silnější textovou koherencí se vkládá krátká pauza.

- **generování přízvuku** – přízvuk se obecně realizuje změnou všech tří základních prozodických charakteristik – frekvence základního hlasivkového tónu, trvání a intenzity. Hlavním úkolem je tedy nalézt pro každé slovo přízvučnou slabiku.

3.2.3 Hodnocení kvality syntetické řeči

V případě rekonstrukce původní promluvy lze poměrně jednoduše měřit rozdíly mezi původní a syntetickou řečí. U syntézy libovolné promluvy (tj. v případě syntézy TTS) však není v drtivé většině případů původní promluva k dispozici, a tak se musí postupovat jiným způsobem. Nejčastěji se provádějí *poslechové testy*, pomocí nichž skupiny lidí subjektivně hodnotí syntetickou řeč. Subjektivní hodnocení řeči může být do jisté míry zavádějící, objektivní testy však v současné době neexistují. Na druhou stranu je to nakonec vždy člověk-jednotlivec, který kvalitu řeči hodnotí. Při hodnocení syntetické řeči se hodnotí srozumitelnost a přirozenost. Toto testování upozorňuje na chyby, kterých se syntetizér dopouští, a pomáhá tyto chyby odstraňovat. Řadí se sem *testy srozumitelnosti*, které zjišťují, jak dobře posluchači rozumějí syntetické řeči a *testy přirozenosti***,** které na druhou stranu hodnotí řeč z celkového hlediska. Více se testováním budeme zabývat v kapitole č. 4. [6]

3.2.4 Aplikace syntézy řeči

Přestože počáteční představy o využití syntetizérů řečových signálů byly spojeny s investiční elektronikou (např. komunikace s počítači, automatické informační systémy), první nejvýznamnější impuls k rozvoji oboru přišel ze sféry spotřební elektroniky. Zde je uvedeno několik příkladů, dle [6, 10] tyto:

- **Spotřební elektronika pro handicapované lidi** – systémy TTS mají obrovský přínos pro handicapované lidi. TTS systémy jim mohou poskytnout nenahraditelné služby. Pomocí speciálně upravených klávesnic mohou zapsat svoji řeč a nechat ji generovat systémem TTS. Profitovat ze systémů TTS mohou samozřejmě i nevidomí, jimž slouží např. přístroje pro automatické čtení novin, knih, kapesní slovníky, elektronické hodinky, elektronické váhy apod.
- **Dopravní prostředky** - možnosti, které přináší syntéza řeči, jsou využity i u nejrůznějších kolejových a silničních vozidel, lodí a letadel. Důvodem je omezení vizuálně vnímaných informací, což dovoluje řidičům více se věnovat vlastnímu řízení vozidla a sledovat dopravní situace. Největší význam mají elektronické syntezátory řeči

pro lodní dopravu, kde automatické hlášení může zahrnovat informace z navigačních subsystémů, kompasů, měřičů rychlosti větru apod.

- **Výuka jazyků** – systémy TTS pak mohou být velice užitečné zvláště pro domácí samouky, kteří sice mají k dispozici kvalitní učebnice, ale chybí jim pro výuku nového jazyka velice důležité vnímání mluvené řeči. V budoucnu bude dokonce možné učit se formou dialogu s výukovým systémem.
- **Automatické čtení** – systémy TTS lze použít v řadě publikací, které převádějí vstupní textovou informaci na řeč, tedy všude tam, kde je z nějakého důvodu vhodné mít k dispozici zvukový výstup. Jako příklad uvedeme automatické čtení knih, úřední korespondence, faxových zpráv, elektronické pošty, webových stránek, SMS zpráv, automatické hlásiče odjezdu vlaků apod.
- **Multimédia, komunikace člověk-počítač, robotika** – je zřejmé, že pro kvalitní komunikaci člověk-počítač se bez kvalitních systémů TTS nelze obejít. Již nyní se hlasové výstupy stávají standardní výbavou počítačů.
- **Výzkum** – vedle praktických aplikací se systémy TTS intenzivně používají i pro výzkumné účely. Určité typy systémů TTS jsou velice oblíbené mezi fonetiky, protože pomocí nich lze studovat i dosud neobjasněné problémy spojené se vznikem a šířením řeči v hlasovém traktu člověka. Dále je možné systémy TTS využít ke studiu prozodie lidské řeči a k tvorbě prozodických modelů. Své uplatnění mohou systémy TTS najít i v lékařství, například při léčení vad poslechu, hluchnutí či poruch vnímání, nebo v psychologii a psychiatrii.
- **Zábava** – další rozsáhlou oblastí možného použití syntézy TTS jsou různé hry a hračky. Opět jde o rozmanité spektrum aplikací, jednoduchými mluvicími panenkami a knížkami počínaje a počítačovými hrami konče.

3.3 Základní přístupy

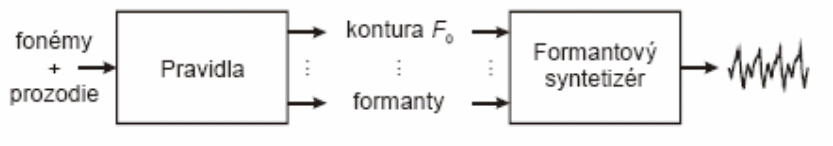
Přístupy k syntéze řeči se nejčastěji dělí podle použitého způsobu modelování při vytváření výsledné řeči na tři typy: formantovou, konkatenační a artikulační syntézu.

3.3.1 Formantová syntéza

Formantová syntéza byla po dlouhou dobu nejpoužívanější metodou syntézy řeči. Je založena na teorii zdroje a filtru a při vytváření řeči modeluje, dle [5, 6]:

- **zdroj buzení** – skládá se ze zdroje znělých zvuků, který simuluje periodickou vibraci hlasivek (neboli generátor impulsů s frekvencí F0), a ze zdroje neznělých zvuků, který generuje šum. Podle povahy výsledného zvuku se používá buď znělého, nebo neznělého zdroje buzení, v případě znělých šumových zvuků se oba typy buzení míchají.
- **vlastní hlasový trakt** – filtr modeluje hlasový trakt pomocí rezonátorů či antirezonátorů, které simulují formanty[1] a antiformanty[2] hlasového traktu člověka.

Parametry tohoto typu syntetizéru úzce souvisí zejména s formantovými charakteristikami řeči. Formantové syntetizéry se nejčastěji implementují podle pravidel – vývoj periody základního hlasivkového tónu, formantových frekvencí a dalších parametrů se tak odhaduje na základě manuálně nastavených pravidel. Pravidla odrážejí změny těchto parametrů v reálném systému vytváření řeči člověkem. Blokové schéma formantového syntetizéru podle pravidel je zobrazeno na obr. č. 10. [6]



Obr. 10 Blokové schéma formantového syntetizéru založeného na pravidlech, zdroj: [5]

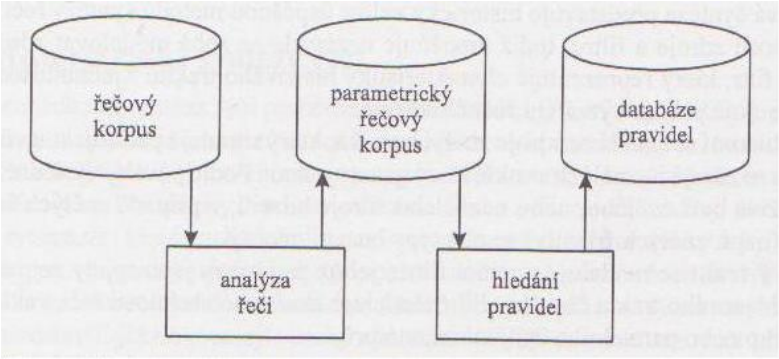
Pravidla:

Při definici již zmiňovaných důležitých pravidel se využívá *řečový korpus*, který se skládá z velkého množství digitálně uložených slov namluvených většinou profesionálním řečníkem. Pečlivě připravený korpus je pak podroben řečové analýze, která digitální data popisuje pomocí řečových parametrů. Pro parametrizaci je typické, že od sebe odděluje složku buzení a hlasového traktu, čímž usnadňuje další zpracování. Následuje časově velmi náročná, avšak

[1] Formanty - jazykově slovotvorné prvky bez lexikálního významu. Jsou charakterizovány jako frekvenční oblasti s vyšší koncentrací akustické energie přítomné v hlavních rezonančních oblastech hlasového traktu. Jsou to jakési „vrcholky" frekvenčního spektra daného zvuku (zejména samohlásky).

[2] Antiformanty lze definovat jako „údolí" ve frekvenčním spektru. Používají se, zapojí-li se do řeči i dutina nosní.

pro celkovou kvalitu výsledné syntetické řeči nesmírně důležitá fáze hledání pravidel v parametrických řečových datech. Tímto úkolem se většinou zabývají experti z oblasti akustiky, fonetiky a fonologie, kteří nejprve zkoumají řečová data od všech řečníků a jejich snahou je nalézt obecné tvary pravidel, která jsou často s výhodou přebírána již od existujících systémů. Poté se nalezená data ukládají do *databáze pravidel.* Konkrétní hodnoty parametrů pravidel (např. frekvence a šířka pásem formantů nebo doba trvání) se pak nastavují pro každou hlásku na hodnoty jednoho řečníka – hlasu, jímž by měl „mluvit" navrhovaný syntetizér. Úspěšnost nalezených pravidel a nastavení jejich parametrů drasticky ovlivňuje kvalitu výsledné syntetické řeči. Celý proces vytváření pravidel je znázorněn na obr. č. 11. [5, 6]



Obr. 11 Schematické znázornění procesu manuálního hledání pravidel, zdroj: [6]

Dobře navržené formantové TTS systémy se vyznačují „konstantní" kvalitou vytvářené řeči, to v praxi znamená, že všechny věty zní přibližně stejně a nestane se, že by se na konci výstupu syntetizéru najednou objevila věta ve výrazně horší kvalitě. Díky své struktuře mohou formantové syntetizéry efektivně měnit identitu hlasu. „Pouhou" změnou několika parametrů tak lze změnit mužský hlas na ženský, šeptat apod. U formantových syntezátorů je možné generovat srozumitelnou řeč s malým počtem parametrů a lze tyto jednotlivé parametry nastavovat samostatně. Ovšem nastavení musí být prováděno velice citlivě, protože hodnoty jednoho parametru ve většině případů ovlivňují hodnoty jiných parametrů. Formantové syntetizéry se nesnaží modelovat do detailů lidské hlasové ústrojí, ale vytváří výsledný signál pomocí pracného manuálního hledání a nastavování pravidel. Základní nevýhodou je tedy pracné manuální hledání pravidel, která popisují nastavování jednotlivých parametrů syntetizéru v závislosti na fonetickém kontextu promluvy. Mezi další nevýhody řadíme časovou náročnost vývoje systému a vzájemnou interakci mezi hodnotami parametrů. Automatické techniky specifikace parametrů jsou sice intenzivně zkoumány, ale

zatím nedosahují uspokojivých výsledků. V porovnání s konkatenační syntézou působí formantová syntéza méně přirozeně, avšak při správně nastavených parametrech je údajně schopen formantový syntetizér vytvářet řeč, která je nerozlišitelná od přirozené promluvy. [5, 6, 11]

3.3.2 Konkatenační syntéza

Konkatenační syntéza je v současné době nejpoužívanější technikou syntézy řeči. Výsledná řeč se vytváří řetězením různých řečových jednotek, segmentů přirozené řeči. Před samotnou syntézou řeči je nutné vytvořit *inventář řečových jednotek*, z něhož se během syntézy konkrétní realizace řečových jednotek vybírá. Protože realizace řečových jednotek jsou v inventáři uloženy s jistými prozodickými a spektrálními vlastnostmi, které mohou být značně odlišné od vlastností požadovaných ve výsledné syntetické řeči, je vhodné jednotlivé realizace prozodicky modifikovat a vhodným způsobem řetězit tak, aby nedocházelo ke skokovým změnám prozodických ani spektrálních vlastností výsledné řeči. Na rozdíl od formantové syntézy nevyžaduje konkatenační syntéza znalost procesu vytváření řeči ani žádná pravidla, a tak odpadá jejich velmi pracné ruční nastavování. Základním stavebním kamenem a pracovním prvkem je tady segment přirozené řeči a předpoklad, že konkatenací takovýchto přirozených řečových segmentů je možné vytvářet vysoce kvalitní syntetickou řeč. Je důležité také zmínit, že v posledním desetiletí se zejména díky rychle rozvíjejícímu výkonu moderních počítačů začala hojně využívat *korpusově orientovaná syntéza*. Tato metoda využívá rozsáhlé řečové korpusy. Veškeré operace během vlastní syntézy řeči pak probíhají zcela automaticky právě na základě zvoleného řečového korpusu. Pro zajištění kvalitní syntetické řeči je tedy zapotřebí zajistit dostatečně kvalitní řečový korpus. [5, 6]

Vlastnosti, dle [5, 6]:

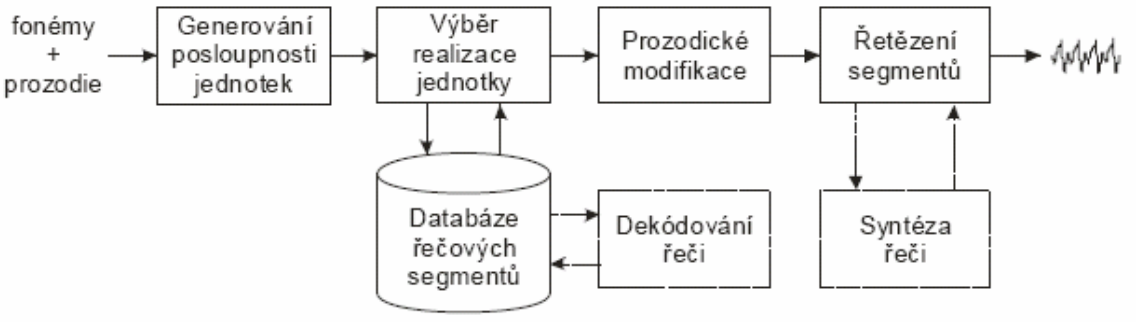
- značná kolísavost v kvalitě výsledné syntetické řeči
- kvalita značně závisí na inventáři řečových jednotek (pokud máme k dispozici kvalitní realizaci řečových jednotek, výsledná řeč může být nerozeznatelná od přirozené řeči člověka)
- při výskytu prozodické nespojitosti (v případě, že si na hranicích řečových segmentů neodpovídají prozodické charakteristiky) nebo spektrální nespojitosti (vzniká, pokud si na hranicích segmentů neodpovídají hodnoty formantů) kvalita generované řeči rapidně klesá

- ruční nebo automatické vytváření inventáře řečových jednotek
- neparametrické (přímo vzorky řeči) nebo parametrické (LCP, HNM) reprezentování řečových jednotek
- spektrální i prozodické jednotky mohou být buď bez modifikací (pouhé řetězení segmentů) anebo s modifikacemi (snaha o minimalizaci nespojitosti na hranici řetězených jednotek)
- syntetickou řeč je možno generovat buď s omezeným slovníkem (např. vlakové nádraží, telefonní automat, apod.) anebo s neomezeným slovníkem

Postup (velmi zjednodušený), dle [6]**:**

1. Nejprve je nutné rozhodnout, jaké řečové jednotky se budou používat. Důležité je si také uvědomit, zda inventář bude vytvářen ručně nebo automaticky.
2. Zde nastává fáze vytváření řečového korpusu. Slova či věty, tedy každá řečová jednotka, musí být v korpusu alespoň jednou zastoupena. Korpus je potom digitálně nahrán a uložen.
3. Dalším krokem je segmentace korpusu podle zvolených řečových jednotek. Jejím cílem je vymezit hranice každé jednotky v korpusu a vznikne tak *segmentový řečový korpus.* Takto vymezené segmenty se spolu s odpovídajícími informacemi o segmentech ukládají do inventáře řečových jednotek.
4. Tato fáze se nazývá *analýza řeči* má za úkol „předvybrat“ zástupce každé jednotky z inventáře řečových jednotek a uložit je do databáze řečových segmentů. Po provedení analýzy řeči je k dispozici *databázi řečových segmentů* a může se již přistoupit k samotnému generování syntetické řeči a skládá se z těchto kroků (viz obr. č. 12):
5. Přepokládá se, že je k dispozici podrobná informace (posloupnost hlásek či fonémů + prozodie) o promluvě, kterou je potřeba syntetizovat.
6. Z této informace si syntetizér vyrobí sekvenci řečových jednotek a z celkové prozodické informace o syntetizované promluvě odvodí prozodické vlastnosti jednotlivých jednotek. Dále si vybere jednoho zástupce řečové jednotky, jelikož databáze obsahuje jeden nebo více segmentů dané řečové jednotky.
7. Nyní je třeba vybrané segmenty dekódovat, v případě, že byly dříve zakódovány. Dekódovaní provádí dekodér řeči.

8. Dalším krokem je nastavení prozodie, jelikož segmenty uložené v databázi vykazují různé prozodické vlastnosti často odlišné od požadovaných.
9. Upravené segmenty se dále zřetězují, a tím vytvářejí výslednou promluvu. Při řetězení je snahou segmenty na sebe spojovat tak, aby se na hranicích neprojevily skokové změny spektrálních charakteristik.
10. Výsledná syntetická řeč se konečně vytváří na signální úrovni na základě řečového modelu použitého pro parametrizaci segmentů. Každý segment se převádí z parametrické do časové oblasti a řetězí se do výstupního proudu syntetické řeči.



Obr. 12 Blokové schéma konkatenační syntézy, zdroj: [5]

Shrnutí:

V současné době má konkatenační syntéza dominantní postavení mezi přístupy k syntéze řeči. Jde o zdaleka nejpoužívanější techniku vytváření řeči. Mezi výhody této syntézy patří *používání přirozených segmentů řeči*, to znamená, že konkatenační syntéza pracuje přímo se segmenty reálného řečového signálu. Z toho nám vyplývá další výhoda a tou je *redukovaná znalost procesu vytváření řeči* – na rozdíl od formantové syntézy není zapotřebí znát detailněji proces vytváření řeči člověkem. Dále *rychlý návrh syntetizéru* - na rozdíl od formantové syntézy zde není nutné hledat složitá pravidla, většina parametrů se totiž nastavuje automaticky z reálných dat. Nyní nám vyplývá další výhoda a tou je *vysoká kvalita syntetické řeči* – výsledná řeč působí daleko přirozeněji než u jiných syntetizérů z důvodu kopírovaní hlasu řečníka, který namluvil řečový korpus. Pro vytvoření požadovaného hlasu tedy není zapotřebí složitých pravidel, stačí „pouze“ mít k dispozici velký počet vět namluvených řečníkem s požadovaným hlasem. Mezi hlavní nevýhody konkatenační syntézy řadíme *výpočetní a paměťovou náročnost*, jelikož je zde zapotřebí uchovávat mnoho realizací každé řečové jednotky a v čase syntézy vybírat nejvhodnější realizace, mohou být nevýhodou poměrně značné výpočetní a paměťové nároky. Další nevýhodou je *těžkopádnost změny hlasu,* která vyžaduje nahrání nové databáze. V konkatenační syntéze je i *nebezpečí špatné*

kvality, což znamená, že jistá promluva (slovo, slabika) bude znít špatně. Toto nebezpečí lze redukovat pečlivým výběrem vět pro nahrávání řečového korpusu. Jako poslední nevýhodu je možno uvést *místo řetězení jednotek,* které bude vždy potenciálním zdrojem problémů. [5, 6]

3.3.3 Artikulační syntéza

Na rozdíl od dvou výše popsaných přístupů se artikulační syntéza snaží modelovat přímo systém vytváření řeči člověkem a jeho fyzikální chování. V principu tak vytváření řeči modeluje nejlepším možným způsobem. Artikulační model (model produkce řeči člověkem) zahrnuje modely hlasivek a mluvidel (artikulátorů) a jejich mechanické pohyby. Signál se vytváří pomocí matematické simulace výdechového proudu vzduchu a simulací činnosti jednotlivých artikulačních modelů. Artikulačními parametry artikulačního syntetizéru jsou relativní polohy jednotlivých artikulačních orgánů, jako např. velikost a tvar retní štěrbiny, výška a pozice špičky jazyka, výška a pozice jazyka, nebo pozice měkkého patra. Parametry buzení pak mohou být například velikost otvoru mezi hlasivkami, napnutí hlasivek nebo tlak plic. [6]

Artikulační syntéza řeči představuje atraktivní metodu syntézy řeči, která se problém syntézy řeči snaží řešit komplexně, přímo modelováním systému produkce řeči. Vyznačuje se však vysokou složitostí a poměrně výraznou výpočetní náročností, a tak syntéza není jednoduchá a je stále otázkou výzkumu. Stejně tak odhad artikulačních parametrů přímo z řečových dat se zdá být velkým problémem a nelze jej řešit standardními algoritmy zpracování signálu. I samotné měření artikulačních parametrů během řeči je poměrně obtížné – nutno použít např. rentgen nebo magnetickou rezonanci. V současné době kvalita vytvářené řeči ani zdaleka nedosahuje úrovně formantových nebo konkatenačních syntetizérů. Artikulační syntéza se zatím omezuje spíše na vytváření izolovaných zvuků, hlásek, slabik či jednoduchých slov. [6]

V současné době se zdá, že artikulační syntéza díky svému komplexnímu popisu a způsobu vytváření řeči má největší předpoklady stát se přístupem, který bude v budoucnu generovat řeč nejvyšší kvality.

3.4 Příklady řečových syntetizérů

V následujících dvou podkapitolách jsou uvedeny současné TTS systémy. Ovšem kvůli obrovskému vývoji technologií syntézy řeči a často nedostatečně zveřejňovaným

informacím o jednotlivých systémech se nejedná o kompletní výčet ani popis uvedených systémů. Vývoj a výzkum v dané problematice neustále pokračuje, stávající verze se zdokonalují a vyvíjejí se i systémy nové. Snahou této práce je postihnout nejznámější syntetizéry. Všechny níže uvedené syntetizéry mají společnou vlastnost a tou je konkatenační syntéza řeči (kromě systému DECTalk™), charakteristická použitím rozsáhlých řečových korpusů a inventářů řečových jednotek.

3.4.1 České systémy

- **ARTIC** (Artificial Talker in Czech) je TTS systém vyvíjený na Katedře kybernetiky Západočeské univerzity v Plzni. Podporuje syntézu dvou českých hlasů (mužský a ženský), slovenštiny a němčiny. V současné době se připravuje také francouzština, angličtina a ruština. Vyvíjen je i audiovizuální systém řeči (tzv. mluvící hlava), který vedle akustické syntézy provádí i syntézu vizuální (modeluje hlavu řečníka, zejména artikulaci rtů) a také syntézu znakové řeči, jejímž úkolem je vytvoření obrazu modelu člověka ukazujícího znakovou řeč například na obrazovce počítače. Spojením této animace se systémem překládajícím psaný text do znakové řeči dostáváme virtuálního tlumočníka překládajícího televizní zprávy. Komerční verze systému ARTIC – ERIS TTS – je nabízena firmou SpeechTech. Podporuje různá lokální a standardizovaná rozhraní na množství rozdílných platform. V příloze č. 1 nalezneme grafickou podobu tohoto programu. [6, 12, 13]
- **CS-VOICE®** je systém TTS firmy Frog Systems. Umožňuje syntézu jednoho českého hlasu. CS-VoiceWave je nadstavba CS-VOICE 97 pro převod textu na soubory WAV. DDETTS je nadstavba pro interpretaci DDE[3] příkazů a NetTTS nadstavba pro převod textu na soubory WAV pro síť. V příloze č. 2 nalezneme grafickou podobu tohoto programu. [6, 14]
- **EPOS** byl vyvinutý především pro výzkumné účely v Ústavu radioelektroniky České akademie věd v Praze. Systém je naprogramován v jazyce C++[4] a je volně dostupný. Nabízí syntézu několika českých a jednoho slovenského hlasu. Epos se snaží být

[3] DDE (Dynamic Data Exchange) umožňuje dvěma aplikacím zřídit spojení a používat jej pro přenos dat.

[4] C++ je objektově orientovaný programovací jazyk. Podporuje několik programovacích stylů (paradigmat) jako je procedurální programování, objektově orientované programování a generické programování, není tedy jazykem čistě objektovým. V současné době patří C++ mezi nejrozšířenější programovací jazyky.

nezávislý na zpracovávaném jazyce, lingvistických popisných metodách a na použitém rozhraní. Dnes aktuální stabilní verze projektu Epos je verze 2.4.85 vydaná v říjnu roku 2008. V příloze č. 3 nalezneme grafickou podobu tohoto systému. [6, 21]

- **WinTalker** je hlasový syntetizér od firmy Rosasoft určený pro zrakově postižené lidi. V současné době podporuje syntézu českého, slovenského a maďarského jazyka. [6]

3.4.2 Zahraniční systémy

Pro mezinárodní scénu jsou charakteristické vícejazyčné systémy, kterým jednoznačně dominuje angličtina. Součástí téměř každého známého systému TTS je nejméně jeden anglický syntetický hlas. Anglických hlasů bývá v jednom systému často i více, liší se podle pohlaví nebo akcentu (britská či americká angličtina apod.). Velmi často se můžeme setkat i s dalšími používanými světovými jazyky (španělštinou, francouzštinou, němčinou). Existuje i množství japonských systémů TTS. Zde je uveden přehled nejznámějších světových TTS systémů.

- **AT&T Natural Voices™** je systém vyvíjený americkou společností AT&T. V současné době umožňuje syntetický řečový výstup v americké a britské angličtině, latinskoamerické španělštině, němčině a francouzštině. Ke každému jazyku je k dispozici několik hlasů. Součástí toho systému jsou i specifické uživatelské slovníky, které umožňují uživatelům definovat výslovnost některých slov, dále upravit znění některých akronymů, zkratek, mluvenou rychlost, hlasitost a hlas. Tato technologie je velmi používaná v aplikacích jako jsou například e-knihy (elektronické knihy), e-learning (výuka po internetu), kde je výsledná řeč rozhodující pro přijetí uživatelem. V příloze č. 4 nalezneme grafickou podobu tohoto programu. [6, 15]
- **RealSpeak™** je systém americké společnosti Nuance (dříve ScanSoft®), který nabízí asi největší počet jazyků. RealSpeak podporuje na 30 jazyků a více než 50 hlasů. Ovšem pokud někdo v tomto systému nenalezne „svůj“ jazyk, je společnost Nuance schopna rozvíjet jazyky a hlasy na požádání. Vedle „tradičních“ jazyků umožňuje také syntézu japonštiny, kantonské a mandarínské čínštiny, korejštiny, dánštiny, holandštiny, italštiny, portugalštiny, mexické španělštiny, norštiny, polštiny, ruštiny, švédštiny, baskičtiny a také naší češtiny. Jedná se o software, který konvertuje text do pozoruhodně vysoké kvality projevu. V příloze č. 5 nalezneme grafickou podobu tohoto programu. [6, 16]

- **Loquendo.** Jde o systém stejnojmenné italské společnosti. Nyní nabízí syntézu devíti italských, sedmi španělských (včetně mexické, chilské a argentinské španělštiny), šesti anglických, tří francouzských, německých a portugalských, dvou řeckých, katalánských a holandských hlasů. Dále hlas dánský, švédský, finský, ruský, turecký, polský a nizozemský. Výslovnost Lexikon zajišťuje specializovanou slovní zásobu, zkratky, akronymy a dokonce i regionální rozdíly výslovnosti řeči. Charakteristika každého hlasu (tj. výšku, rychlost a hlasitost) může být vyladěna a plně kontrolována. Systém Loquendo podporuje také čtení speciálních formátů, jako jsou telefonní čísla, měny a e-mailové adresy. Loquendo TTS je k dispozici také pro telefonování, multimédia a vestavěné aplikace. V příloze č. 6 nalezneme grafickou podobu tohoto TTS systému. [6, 17]
- **Elan Sayso™** a **Elan Tempo™ .** Tyto systémy společnosti Elan speech z Francie (skupiny Acapela Group – kombinace tří velkých evropských společností) se liší kvalitou a náročností na hardware počítače. Elan Sayso™ je korpusově orientovaný systém podporující výběr jednotek a jako takový se vyznačuje vyšší kvalitou syntetické řeči. Podporuje syntézu 25 jazyků a více než 50 hlasů, několika hlasů v angličtině, němčině, holandštině, španělštině, francouzštině, italštině, polštině, švédštině, norštině, arabštině nebo jednoho hlasu v ruštině, finštině, norštině, turečtině, dánštině a také v češtině). Na druhou stranu Elan Tempo™ pracuje pouze s jedním reprezentantem každé řečové jednotky a je vhodný pro použití v zařízeních s omezenými paměťovými a výpočetními možnostmi. V příloze č. 7 nalezneme grafickou podobu programu Elan společnosti Acapela Group. [6, 18]
- **Cepstral Voices.** Cepstral® je americká firma nabízející syntézu 5 jazyků (angličtiny, španělštiny, francouzštiny, němčiny a arabštiny). Důraz je kladen na přípravu speciálních hlasů šitých na míru konkrétním aplikacím a požadavkům, tj. na syntézu z omezené oblasti, a to v telefonech, navigacích, mobilech, hračkách, hrách, počasí, lékařství a jiné. V příloze č. 8 nalezneme grafickou podobu tohoto syntetizéru. [6, 19]
- **DECTalk™.** Tento systém americké společnosti Fonix je zajímavý tím, že jde v současnosti v podstatě o jediný komerčně úspěšný systém TTS založený na formantové syntéze řeči. V porovnání s ostatními konkatenačními systémy syntézy řeči sice produkuje řeč nižší kvality ovšem s mnohem menšími paměťovými a výpočtovými nároky. Nabízí devět modifikovatelných hlasů (mužské, ženské i dětské) pro šest různých jazyků. [6]
- **SVOX** je systém TTS švýcarské společnosti SVOX Ltd. Specializuje se na syntézu němčiny, ale podporuje také syntézu jiných jazyků, jako francouzštinu, americkou a

britskou angličtinu, dánštinu, portugalštinu, ruštinu, polštinu, turečtinu, švédštinu, evropskou i mexickou španělštinu, italštinu nebo také čínštinu, japonštinu i korejštinu. SVOX se zabývá především syntézou v konkrétních aplikacích, například v telefonech, navigacích, ukazatelích, e-mailech a ostatních. V příloze č. 9 nalezneme grafickou podobu tohoto programu. [6, 20]

- **FESTIVAL** je kompletní systém syntézy řeči šířený pod licencí ve stylu BSD[5]. Nejde tedy o komerční, ale volně dostupný otevřený systém původem z Edinburghské univerzity. Festival je napsán v jazyce C++. Systém je volně dostupný pro vzdělávací, vědecké i individuální použití. Je vhodný jako jednoduchý TTS systém, pro experimenty se syntézou řeči (různé hlasy, specifické fráze, typické dialogy), případně pro vývoj a testování nových metod syntézy řeči. Díky své modulární architektuře by měl být použitelný pro všechny světové jazyky, ovšem například čeština není silnou stránkou Festivalu. V příloze č. 10 nalezneme grafickou podobu tohoto systému. [6, 21]
- **MBROLA** (Multi Band Resynthesis Overlap and Add) je významný akademický projekt, který vznikl v 90. letech na Faculté Polytechnique de Mons v Belgii. Cílem projektu byl vývoj kvalitní multilingvální řečové syntézy pro nekomerční účely. V současné době systém MBROLA zahrnuje 26 jazyků a 50 hlasů. MBROLA není skutečným TTS systémem, neboť jako vstup neakceptuje řádkový text. [21]
- **ReadPlease** od společnosti Readplease Corporation nabízí syntézu americké a britské angličtiny, němčiny, francouštiny a americké španělštiny. Nyní je k dispozici zdarma verze ReadPlease 2003, která nabízí plnou podporou společnosti Microsoft, umožňuje nastavit rychlost čtení, přidávat vlastní slova, přehrávat kdekoli v dokumentu a další. V příloze č. 11 nalezneme grafickou podobu tohoto programu. [22]

[5] BSD licence je licence pro svobodný software, mezi kterými je jednou z nejsvobodnějších. Umožňuje volné šíření licencovaného obsahu, přičemž vyžaduje pouze uvedení autora a informace o licenci, spolu s upozorněním na zřeknutí se odpovědnosti za dílo.

4 Testování vybraných softwarových nástrojů pro převod textu na řeč

Jak bylo již uvedeno v kapitole 3.2.3, systémy TTS jsou testovány pomocí *poslechových testů*, při nichž skupiny lidí subjektivně hodnotí syntetickou řeč. Do nich patří *testy srozumitelnosti*, které zjišťují, jak dobře posluchači rozumějí syntetické řeči (důraz je zde většinou kladen na vnímání přechodových zvuků, neboli koartikulaci – segmentální kvalita řeči) a *testy přirozenosti*, které hodnotí řeč z celkového hlediska (suprasegmentální kvalita řeči). [6]

4.1 Návrh vhodného způsobu testování TTS systémů

Vybrané TTS systémy budou testovány pomocí již zmiňovaných testů srozumitelnosti a testů přirozenosti, které jsou považovány za nejlepší „hodnotile“ kvality TTS systémů.

4.1.1 Testy srozumitelnosti

Testy srozumitelnosti rozdělujeme na *test diagnostikou rýmu* (Diagnostic Rhyme Test, DRT), *test modifikací rýmu* (Modified Rhyme Test, MRT) a *test identifikace skupin.*

Testy DRT a MRT jsou nejpoužívanější poslechové testy srozumitelnosti. U obou je posluchačům přehráno po řadě vždy jedno slovo a jejich úkolem je toto slovo identifikovat ve skupině podobných, předem daných slov. Testy DRT používají ve skupině dvě slova, zatímco testy MRT používají slov pět. Oba testy se soustředí na souhlásky, neboť právě správná syntéza souhlásek má větší vliv na srozumitelnost řeči. U testů DRT se slova liší pouze v počáteční souhlásce, kdežto u testů MRT se liší v počáteční nebo koncové souhlásce. V obou případech se používají slova jednoslabičná. Literatura uvádí, že pro správnou diagnostiku výsledků je nutno přehrát 192 slov v případě testu DRT a 50 slov v případě testu MRT. K výhodám těchto testů řadíme spolehlivost a k nevýhodám zase omezení testu na počáteční a koncové souhlásky a omezený počet alternativ v každé skupině. Pro testování vybraných TTS systémů byl vybrán test MRT. V tabulkách č. 1 a 2 jsou uvedeny skupiny slov, která poslouží pro testování vybraných TTS. [6]

Tabulka 1 – Skupina slov pro testování TTS systémů české řeči testem MRT, zdroj: vlastní

lok	tok	bok	mok	nok
řek	jek	bek	sek	šek
mor	mok	mop	mol	moc
byt	bych	byl	bys	bij
sál	vál	šál	žal	bál
pít	bít	jít	lít	vít
fuk	buk	puk	muk	suk
veš	ves	vem	web	ven
lis	vis	mis	rys	sis
růž	rub	ruch	rum	rus

Tabulka 2 – Skupina slov pro testování TTS systémů anglické řeči testem MRT, zdroj: [23]

tack	tab	tang	tan	tap
went	sent	bent	dent	tent
not	lot	got	pot	hot
pig	big	dig	rig	fig
kill	kit	kick	king	kid
bun	bus	but	bug	buck
heat	meat	feat	seat	meat
took	cook	look	hook	book
sag	sat	sack	sad	sap
fun	sun	bun	run	nun

Testy identifikace skupin představují obecnější metodu hodnocení řeči. Princip spočívá v tom, že se do krátkých frází vloží nesmyslné slovo. Posluchači mají k dispozici pouze seznam slov. Asi nejznámějším zástupcem této skupiny testů je test sémanticky nepredikovatelných[6] vět (Semantically Unpredictable Sentences, SUS). Jejich cílem je minimalizovat kontextovou a sémantickou informaci slov ve větách a tím ztížit proces rozpoznávání promluvy. Posluchač tak nemůže využít svých zkušeností a neznámé slovo z věty si „domyslet“ z kontextu okolních slov. Testované věty se vytvářejí na základě větných rámců, pevně daných struktur slov ve větě. Jednotlivé věty se pak sestavují tak, že do daných slotů v těchto větných rámcích jsou náhodně doplněna často používaná slova jazyka tak, aby

[6] Sémanticky nepredikovatelné věty jsou takové, které jsou gramaticky správně, ale nedávají smysl.

věty byly gramaticky správně. Úkolem posluchačů je přesně napsat chybějící slovo. Pomocí testu SUS budou testovány pouze TTS systémy, které umožňují syntézu řeči v českém jazyce. Větné rámce a věty pro testování vybraných TTS systémů jsou uvedeny v následujících tabulce č. 3. [6]

Tabulka 3 – Větné rámce a věty pro testování TTS systémů české řeči testem SUS, zdroj: upraveno na základě [6]

podst. jm. 1.p., č.j.	**sloveso se 7.p., 3.os., č.j., přít. čas**	**předložka se 7. p.**	**příd.jm.7.p., č.j., r.m./s.**	**podst.jm. 7.p., č.j., r.m./s.**
Sklenice	dělá	nad	děrovaným	rámem.
Padesátka	utírá	za	studeným	vládcem.
Kružnice	řve	pod	zmačkaným	úhlem.
příd.jm.1.p., č.j., r.m.	**podst. jm. 1.p., č.j., r.m.**	**sloveso „být“, č.j., r.m.**	**příd.jm.1.p., č.j., r.m.**	**podst. jm. 1.p., č.j., r.m.**
Vymetený	opičák	je	lesní	držák.
Visící	zvěrokruh	bude	zmrzlý	kůl.
Žlutý	osel	byl	vykutálený	pták.
sloveso rozkaz. způsobu	**příd.jm. 4.p., č.mn., r.m./ž.**	**podst.jm. 4.p., č.mn., r.m./ž.**	**spojka souř.**	**podst.jm. 4.p., č.mn.**
Podlej	černé	domy	a	nitě.
Nedělej	dlouhé	růže	i	drobky.
Popros	zlobivé	pastelky	nebo	nohy.
příslovce zp./m./příč.	**sloveso infinitiv**	**podst.jm. 4.p.**	**příd.jm. 2.p., č.mn.**	**podst.jm. 2.p., č.mn.**
Kdy	brát	pružinu	zelených	kostí?
Jak	naučit	sluchátka	tvrdých	papírů?
Proč	vidět	číslo	psích	náramků?
podst.jm. 1.p., č.j. r.m.	**sloveso 3.os., č.j., r.m. čas přít./min.**	**podst.jm. 4.p.**	**zvratné zájmeno**	**sloveso infinitiv**
Petr	uvolnil	štěstí	se	najíst.
Dělník	kouká	kytku	si	vzít.
Pes	trénuje	knížku	si	usušit.

Do testů srozumitelnosti bylo ještě zařazeno hodnocení kvality řeči u značek, symbolů, čísel, zkratek, vzorců aj. Vybrané TTS systémy budou podrobeny tomuto testu podle následující tabulky č. 4.

Tabulka 4 – Tabulka pro testování znaků u TTS systémů české i anglické řeči, zdroj: vlastní

Znak	**Přečteno (ano x ne)**
123	
1000000	
0,001	
9 – 5 = 4	
$V = a^3$	
O = 2 (a + b)	
1 < 2	
¾	
50 %	
§ 2	
α β δ λ π	
☺	
info@upce.cz	
www.upce.cz	
to & ono	
Ing.	
apod.	
ČSAD	

Znak	**Přečteno (ano x ne)**
123	
1000000	
0,001	
9 – 5 = 4	
$V = a^3$	
O = 2 (a + b)	
1 < 2	
¾	
50 %	
§ 2	
α β δ λ π	
☺	
king@yahoo.com	
www.google.com	
me & you	
Mrs.	
etc.	
OSN	

4.1.2 Testy přirozenosti

Testy přirozenosti hodnotí a porovnávají řeč podle celkové kvality. Neposuzují pouze to, jak bylo řeči rozumět, ale snaží se také hodnotit, jak byla dobře generována, tj. jak se dobře poslouchá. Vzhledem k tomu, že je velice těžké obecně definovat přirozenost řeči, používají se opět subjektivní metody. Nejrozšířenějším testem přirozenosti je *test MOS* (Mean Opinion Score) definovaný Mezinárodní telekomunikační unií. Hodnocení kvality řeči se provádí tak, že posluchači svůj názor vyjadřují na stupnici 1-5, viz tab. č. 5. Vyjadřují tak svůj názor na kvalitu řeči a označují, jakou námahu museli při poslechu syntetické řeči vynaložit.

Průměrná známka syntetizéru je získaná ze zapsaných výsledků. Tímto typem testu budou hodnoceny syntetizéry umožňující syntézu jak v českém, tak i anglickém jazyce. [6]

Tabulka 5 – Tabulka pro hodnocení TTS systémů testem MOS, zdroj: upraveno na základě [6]

Známka	**Kvalita řeči**	**Vynaložená námaha při poslechu**
1	špatná	zcela neodpovídající
2	horší	veliká
3	průměrná	průměrná
4	dobrá	bez větší námahy
5	výborná	bez jakékoli námahy

Druhou metodou testů přirozenosti je *porovnávání párů* (Category Comparison Rating, CCR), při nichž je hodnocena vždy jedna věta ve dvou vzorcích. Jednoduše řečeno, máme jednu větu a tu porovnáváme ve dvou různých TTS systémech (tts1 vs. tts2). Na posluchači je vybrat TTS systém, u kterého mu daná věta zněla lépe, což je jednodušší varianta: „upřednostňuji tts1 před tts2“. Složitější varianta je uvedena v následující tabulce č. 6. Také tímto typem testu budou hodnoceny syntetizéry umožňující syntézu řeči jak v českém, tak i anglickém jazyce. [6]

Tabulka 6 – Tabulka pro hodnocení TTS systémů testem CCR, zdroj: upraveno na základě [6]

Známka	**tts1 vs. tt2**
2	mnohem lepší
1	lepší
0	stejné
-1	horší
-2	mnohem horší

4.2 Realizace testování TTS systémů

Testování bylo provedeno pomocí tří posluchačů, kteří systematicky prováděli výše popsané testy. Testy následně vyhodnocovali na základě svého subjektivního dojmu. Jednotlivé testy probíhaly v časových prodlevách tak, aby nebyl ovlivněn výsledek testování.

4.2.1 Testování TTS systémů české řeči

Testování byly podrobeny **české systémy** - ARTIC, CS-VOICE, EPOS[7] a **zahraniční systémy** - RealSpeak™ a Elan společnosti Acapela.[8]

- **Test MRT**

  Testem srozumitelnosti MRT byly hodnoceny vybrané TTS systémy. Jak bylo uvedeno v kapitole 4.1.1, posluchači měli k dispozici tabulku č. 1, v níž je deset řádků a na každém z nich se nachází pět podobných jednoslabičných slov. Na daném TTS systému bylo vždy přehráno právě jedno slovo a úkolem posluchačů bylo ho správně označit. V následující tabulce č. 7 je uvedeno, kolik slov bylo v průměru z možných deseti správně rozpoznáno.

Tabulka 7 – Vyhodnocení testu MRT na TTS systémech české řeči, zdroj: vlastní

TTS systém	Počet bodů (10 max.)
ARTIC	9,33
CS-VOICE	6
EPOS	8,67
RealSpeak	9,67
Elan (Acapela)	10

- **Test SUS**

  U toho typu testu, jak bylo uvedeno také v kapitole 4.1.1, bylo úkolem posluchačů do tabulky č. 3, doplnit jedno chybějící slovo. Maximální počet získaných bodů byl 15 za 15 správně doplněných slov. V tabulce č. 8 je uveden průměrný počet získaných bodů u každého TTS systému české řeči.

[7] Syntézu české řeči provádí také český systém WinTalker, který se bohužel pro testování nepodařilo získat.

[8] Syntézu české řeči provádí také zahraniční systém SVOX a Festival, u kterých se bohužel čeština nepodařila zpřístupnit.

Tabulka 8 – Vyhodnocení testu SUS na TTS systémech české řeči, zdroj: vlastní

TTS systém	Počet bodů (15 max.)
ARTIC	15
CS-VOICE	13,67
EPOS	14,67
RealSpeak	15
Elan (Acapela)	15

- **Testování znaků**

  Při testování znaků byly v každém vybraném TTS systému přehrány znaky z tabulky č. 4. Zaznamenané výsledky jsou uvedeny v následující tabulce č. 9. Maximální počet bodů, které bylo možno získat, je 17.

Tabulka 9 – Vyhodnocení testování znaků na TTS systémech české řeči, zdroj: vlastní

Znak	ARTIC	CS-VOICE	EPOS	RealSpeak	Elan (Acapela)
123	✓	✓	✓	✓	✓
1000000	✓	✓	✓	✓	✓
0,001	✓	✓	×	✓	✓
9 – 5 = 4	×	×	✓ ×	✓ ×	✓ ×
$V = a^3$	×	×	×	×	×
O = 2 (a + b)	×	×	✓	✓	×
1 < 2	×	×	✓	✓	✓
¾	✓	×	✓	✓	✓
50 %	×	×	✓	✓	✓
§ 2	×	×	×	×	✓
α β δ λ π	×	×	×	×	×
☺	×	×	×	×	×
info@upce.cz	×	×	✓	✓	✓
www.upce.cz	×	×	✓	×	✓
to & ono	×	×	×	×	×
Ing.	×	×	✓	✓	×
apod.	×	×	✓	✓	✓
ČSAD	×	×	✓	×	×
BODY	**4**	**3**	**11,5**	**10,5**	**10,5**

Pozn.: u syntetizérů, kde je vyhodnocení kladné i záporné (✓ ×) byla přečtena vždy pouze část testovaného vzorce. Totéž platí i pro tabulku 13.

- **Test MOS**

  Test přirozenosti MOS byl proveden tak, jak je popsán v kapitole 4.1.2. Posluchači vyjadřovali svůj názor na kvalitu řeči a námahu, jakou museli při poslechu syntetické řeči vynaložit. V tabulce č. 10 jsou zaznamenány průměrné body, které získaly jednotlivé TTS systémy české řeči.

Tabulka 10 – Vyhodnocení testu MOS na TTS systémech české řeči, zdroj: vlastní

TTS systém	**Počet bodů (5 max.)**
ARTIC	5
CS-VOICE	2,33
EPOS	4
RealSpeak	4,67
Elan (Acapela)	5

- **Test CCR**

  V tomto testu byly porovnávány jednotlivé páry TTS systémů české řeči mezi sebou podle tabulky č. 6 v kapitole 4.1.2. Výsledky tohoto testu jdou uvedeny v následující tabulce č. 11.

Tabulka 11 – Vyhodnocení testu CCR na TTS systémech české řeči, zdroj: vlastní

ARTIC vs.	**Počet bodů (2 max, -2 min)**
CS-VOICE	2
EPOS	1
RealSpeak	1
Elan (Acapela)	0
Σ	**4**
CS-VOICE vs.	
ARTIC	-2
EPOS	-2
RealSpeak	-2
Elan (Acapela)	-2
Σ	**-8**
EPOS vs.	
ARTIC	-1
CS-VOICE	2
RealSpeak	0
Elan (Acapela)	-1
Σ	**0**

RealSpeak vs.	
ARTIC	-1
CS-VOICE	2
EPOS	0
Elan (Acapela)	-1
Σ	**0**
Elan (Acapela) vs.	
ARTIC	0
CS-VOICE	2
EPOS	1
RealSpeak	1
Σ	**4**

4.2.2 Testování TTS systémů anglické řeči

Následující TTS systémy umožňují syntézu řeči v mnoha světových jazycích. Testování probíhalo v jazyce anglickém a byly mu podrobeny tyto zahraniční TTS systémy[9]: AT&T Natural Voices™, RealSpeak™, Loquendo, Elan (Acapela), Cepstral Voices, ReadPlease a FESTIVAL.

- **Test MRT**

  Stejným způsobem jako TTS systémy české řeči byly hodnoceny TTS systémy anglické řeči. Vycházelo se zde také z kapitoly 4.1.1, konkrétně z tabulky č. 2. V následující tabulce č. 12 je uvedeno kolik slov bylo v průměru z možných deseti správně rozpoznáno.

Tabulka 12 – Vyhodnocení testu MRT na TTS systémech anglické řeči, zdroj: vlastní

TTS systém	**Počet bodů (10 max.)**
AT&T Natural Voices	8,67
RealSpeak	8
Loquendo	8,67
Elan (Acapela)	8,34
Cepstral Voices	8,34

[9] V současné době žádný český TTS systém angličtinu nepodporuje. Pouze u ARTICu se nyní připravuje francouzština, angličtina a ruština. Pro testování TTS systémů anglické řeči se bohužel nepodařily získat systémy MBROLA, DECTalk a systém SVOX byl nabídnut pouze v němčině.

FESTIVAL	6,34
ReadPlease	8,34

- **Test SUS**

  Tímto typem testu nebyly hodnoceny žádné TTS systémy anglické řeči, jelikož test SUS má význam pouze tehdy, rozumí-li posluchači bezchybně dané řeči.

- **Testování znaků**

  U tohoto typu testu byly jednotlivé znaky z tabulky č. 4. přehrány v uvedených TTS systémech anglické řeči. Vyhodnocení je uvedeno v následující tabulce č. 13.

Tabulka 13 – Vyhodnocení testování znaků na TTS systémech anglické řeči, zdroj: vlastní

Znak	**AT&T Natural Voices**	**Real-Speak**	**Loquendo**	**Elan (Acapela)**	**Cepstral**	**Festival**	**Read Please**
123	✓	✓	✓	✓	✓	✓	✓
1000000	✓	✓	×	✓	✓	✓	✓
0,001	×	×	×	×	✓	×	×
9 – 5 = 4	✓	✓ ×	✓ ×	✓ ×	✓ ×	✓ ×	✓
$V = a^3$	×	×	×	×	×	×	×
O = 2 (a+b)	×	×	×	×	×	×	×
1 < 2	✓	×	✓	✓	×	✓	×
¾	✓	✓	✓	✓	✓	✓	✓
50 %	✓	✓	✓	✓	✓	✓	✓
§ 2	×	×	✓	✓	×	×	✓
α β δ λ π	×	×	×	×	×	×	×
☺	✓	×	×	×	✓	×	✓
king@yahoo.com	✓	✓	✓	✓	✓	✓	✓
www.google.com	✓	✓	✓	✓	✓	✓	✓
Me & you	×	×	×	×	×	×	×
Mrs.	×	✓	✓	✓	✓	✓	✓
etc.	✓	✓	✓	✓	✓	✓	✓
OSN	✓	✓	✓	×	✓	✓	×
BODY	**11**	**9,5**	**10,5**	**10,5**	**11,5**	**10,5**	**11**

- **Test MOS**

  Test přirozenosti MOS u TTS systémů anglické řeči byl proveden stejně jako test MOS u TTS systémů české řeči. V tabulce č. 14 jsou znázorněny průměrné body, které získaly jednotlivé TTS systémy anglické řeči.

Tabulka 14 – Vyhodnocení testu MOS na TTS systémech anglické řeči, zdroj: vlastní

TTS systém	**Počet bodů (5 max.)**
AT&T Natural Voices	4,67
RealSpeak	5
Loquendo	5
Elan (Acapela)	4,67
Cepstral Voices	4
FESTIVAL	3
ReadPlease	4

- **Test CCR**

  V tomto testu byly také porovnávány jednotlivé páry TTS systémů anglické řeči mezi sebou podle tabulky č. 6 v kapitole 4.1.2. Výsledky testu jsou uvedeny v následující tabulce č. 15.

Tabulka 15 – Vyhodnocení testu CCR na TTS systémech anglické řeči, zdroj: vlastní

AT&T Natural Voices vs.	**Počet bodů (2 max, -2 min)**
RealSpeak	-1
Loquendo	-1
Elan (Acapela)	0
Cepstral Voices	1
FESTIVAL	2
ReadPlease	1
Σ	**2**
RealSpeak vs.	
AT&T Natural Voices	1
Loquedno	0
Elan (Acapela)	1
Cepstral Voices	1
FESTIVAL	2
ReadPlease	1
Σ	**6**

Loquendo vs.	
AT&T Natural Voices	1
RealSpeak	0
Elan (Acapela)	1
Cepstral Voices	1
FESTIVAL	2
ReadPlease	1
Σ	**6**
Elan (Acapela) vs.	
AT&T Natural Voices	0
RealSpeak	-1
Loquendo	-1
Cepstral Voices	1
FESTIVAL	2
ReadPlease	1
Σ	**2**
Cepstral Voices vs.	
AT&T Natural Voices	-1
RealSpeak	-1
Loquendo	-1
Elan (Acapela)	-1
FESTIVAL	1
ReadPlease	0
Σ	**-3**
FESTIVAL vs.	
AT&T Natural Voices	-2
RealSpeak	-2
Loquendo	-2
Elan (Acapela)	-2
Cepstral Voices	-1
ReadPlease	-1
Σ	**-10**
ReadPlease vs.	
AT& Natural Voices	-1
RealSpeak	-1
Loquendo	-1
Elan (Acapela)	-1
Cepstral Voices	0
FESTIVAL	1
Σ	**-3**

4.3 Vyhodnocení výsledků testování TTS systémů

V následujících dvou odstavcích jsou zhodnoceny testované syntetizéry řeči:

- **TTS systémy české řeči**

  Z pěti testovaných syntetizérů české řeči byly na základě provedených testů nejlépe hodnoceny dva systémy, a to český systém ARTIC (společnosti SpeechTech) a zahraniční systém Elan (společnosti Acapela). Jako první byly u těchto systémů prováděny testy srozumitelnosti (MRT, SUS a testování znaků). V této skupině testů získal systém Elan oproti systému ARTICu vždy maximální počet bodů. Ovšem tyto testy srozumitelnosti nemají takovou vypovídací hodnotu jako testy přirozenosti (MOS a CCR), které hodnotí řeč podle celkové kvality. Dle těchto testů bezkonkurenčně „zvítězily“ zmiňované syntetizéry řeči. Vezmeme-li ale v úvahu všechny testy, je nutno konstatovat, že syntetizér Elan je ze všech testovaných systémů nejlepší, avšak z celkového pohledu by mohl být zařazen na stejnou úroveň jako syntetizér ARTIC. Jako velmi kvalitní TTS systém je nutno uvést zahraniční systém RealSpeak (společnosti ScanSoft), který také dosáhl vynikajících výsledků. O něco hůře se umístil český syntetizér EPOS (vyvíjený Ústavem radiotechniky a elektroniky Akademie věd ČR), který získal ve většině testů o stupeň horší hodnocení než předchozí tři systémy, na druhé straně při testování znaků dostál nejlepšího výsledku. Poslední příčku při hodnocení TTS systémů české řeči český získal syntetizér CS-VOICE (společnosti Frog Systems). Tento systém příliš neuspěl při testech srozumitelnosti ani při testech přirozenosti.

  Shrnutí: systémy ARTIC, Elan, RealSpeak a EPOS jsou vysoce kvalitní syntetizéry řeči. Jimi produkovaná řeč je přirozená, je jí dobře rozumět a při poslechu není třeba vynaložit větší námahy. Jsou schopné přečíst nejrůznější zkratky, čísla, e-mailové adresy apod. Na základě provedených testů se nejhůře umístil CS-VOICE, který musí být označen za nepříliš kvalitní. Produkovaná řeč má špatnou artikulaci, při jejím poslechu je nutno se velmi soustředit. Je schopný ze všech znaků přečíst pouze čísla, ostatní znaky nebo adresy bohužel ne. Závěrečné bodové hodnocení TTS systémů české řeči je přehledně uvedeno v následující tabulce 16.

Tabulka 16 – Závěrečné bodové hodnocení TTS systémů české řeči, zdroj: vlastní

Test	ARTIC	CS-VOICE	EPOS	RealSpeak	Elan
MRT	9,33	6	8,67	9,67	10
SUS	15	13,67	14,67	15	15
Testování znaků	4	3	11,5	10,5	10,5
∑ testy srozumitelnosti	**28,33**	**22,67**	**34,84**	**35,17**	**35,5**
MOS	5	2,33	4	4,67	5
CCR	4	-8	0	0	4
∑ testy přirozenosti	**9**	**-5,67**	**4**	**4,67**	**9**
∑∑	**37,33**	**17**	**38,84**	**39,84**	**44,5**

Pozn.: Bodové hodnocení je řazeno vzestupně, tzn. čím více bodů, tím lepší TTS systém. Totéž platí i pro tabulku 17.

- **TTS systémy anglické řeči**

  Mezi sedmi testovanými TTS systémy anglické řeči nejsou žádné systémy z české produkce softwarových firem. Stejně jako u systémů české řeči dosahují i tyto velmi dobrých výsledků, proto je dobré připomenout, že hodnocení je subjektivní. Jako nejlepší byl podle hodnocení celkové kvality řeči (testů přirozenosti) vyhodnocen systém RealSpeak, který byl testován i v systémech české řeči a systém Loquendo (italské společnosti). S přihlédnutím na testy srozumitelnosti se na prvním místě umístil systém Loquendo. Pouze o něco hůře dopadl systém AT&T Natural Voices (americké společnosti) a systém Elan, který byl již také testován v systémech české řeči a rozdílem jediného bodu se umístil za zmiňovaný AT&T Natural Voices. Jako další velmi kvalitní syntetizéry se ukázaly Cepstral Voices a ReadPlease, které získaly stejné hodnocení, ale oproti již zmiňovaným syntetizérům byly o stupeň horší ve všech prováděných testech. Na poslední příčce se na základě realizace testů umístil syntetizér Festival (vyvíjený Edinburghskou univerzitou).

  Shrnutí: u syntetizérů anglické řeči vyšly podobné výsledky jako u těch s českou řečí. Všechny zmiňované systémy se s výjimkou jediného vyznačují velmi vysokou kvalitou, tzn. produkovaná řeč je přirozená, je jí dobře rozumět a při poslechu není třeba vynaložit větší námahy. Jsou schopné přečíst nejrůznější zkratky, čísla, e-mailové adresy apod. Pouze systém Festival těchto výsledků nedosahoval. Zkratky, čísla, e-mailové adresy přečetl a produkovaná řeč nezněla příliš přirozeně (mezi slovy byly divné pauzy). Je dobré zmínit, že systémy, které

byly porovnávány v obou jazycích, se při testování znaků občas lišily, např. RealSpeak přečetl v češtině číslo 0,001, avšak v angličtině to nedokázal. Závěrečné bodové hodnocení TTS systémů anglické řeči je přehledně uvedeno v následující tabulce 17.

Tabulka 17 – Závěrečné bodové hodnocení TTS systémů anglické řeči, zdroj: vlastní

Test	**AT&T**	**Real-Speak**	**Loquendo**	**Elan**	**Cepstral**	**Festival**	**Read-Please**
MRT	8,67	8	8,67	8,37	8,34	6,34	8,34
Testování znaků	11	9,5	10,5	10,5	11,5	10,5	11
∑ testy srozumitelnosti	**19,67**	**17,5**	**19,17**	**18,87**	**19,84**	**16,84**	**19,34**
MOS	4,67	5	5	4,67	4	3	4
CCR	2	6	6	2	-3	-10	-3
∑ testy přirozenosti	**6,67**	**11**	**11**	**6,67**	**1**	**-7**	**1**
∑∑	**26,34**	**28,5**	**30,17**	**25,54**	**20,84**	**9,84**	**20,34**

5 Závěr

S vývojem výpočetní techniky nastává zlom ve vývoji syntézy řeči a díky intenzivnímu výzkumu a zdokonalujícím se softwarovým nástrojům se syntéza řeči stává daleko snazší než tomu bylo dříve. Ovšem ani v dnešním světě nejmodernějších technologií nejsou všechny softwary pro převod textu na řeč příliš dokonalé. Proto hlavním tématem této práce bylo testování vybraných syntetizérů řeči.

Syntéza řeči má neocenitelnou úlohu nejen pro lidi s poruchami sluchu, ale i pro ty, kteří naopak svou řeč ztratili. V současné době je čím dál populárnější v zábavném průmyslu, dále pak nabízí širokou škálu služeb v oblastech, kde není možné využít jiný způsob komunikace. Dokáže dokonale nahradit i skutečného lidského řečníka. Uplatňuje se také v informačních systémech a systémech pro čtení textů (SMS, knih, e-mailů apod.)

První část práce se teoreticky zabývala popisem technologie pro převod textu na mluvenou řeč. Byl zde zdůrazněn především samotný „výrobní“ postup řeči, základní přístupy při modelování řeči a představení současných českých i zahraničních TTS systémů.

Druhá část práce se věnovala samotnému testování popsaných TTS systémů. Nejprve byl popsán nejvhodnější postup, jak dané syntetizéry otestovat, a následně byl test zrealizován. Pro možnost porovnání výsledků bylo testování prováděno ve dvou krocích. V tom prvním se testovaly čtyři české a dva zahraniční systémy, které umožňují syntézu řeči v českém jazyce. Původně měl být do těchto testů zahrnut i český systém WinTalker, ale bohužel nebyl zpřístupněn. Ve druhém kroku se testovalo sedm zahraničních syntetizérů, které umožňují syntézu řeči v jazyce anglickém. I zde bohužel nebyly zpřístupněny některé syntetizéry jako DECTalk, MBROLA a SVOX. Syntetizéry Elan a RealSpeak umožňují syntézu v českém i v anglickém jazyce, a proto byly otestovány pro oba případy. Ne vždy však dosahovaly stejných výsledků pro oba jazyky. Popsané systémy byly vyzkoušeny pomocí testů srozumitelnosti - test MRT, SUS (pouze systémy české řeči), testování znaků a testů přirozenosti - test MOS a CCR. Větší důraz je však kladen na testy přirozenosti, které hodnotí systémy podle celkové kvality produkované řeči.

Na základě realizace zmiňovaných testů mohou být za nejkvalitnější syntetizéry české řeči považovány zahraniční systém Elan a český systém ARTIC. Syntetizéry RealSpeak a EPOS jsou také velice kvalitní. Pouze o systému CS-VOICE kladné hodnocení napsat nelze.

Za vysoce kvalitní syntetizéry anglické řeči je možné na základě provedených testů považovat všechny, tj. AT&T Natural Voices, RealSpeak, Loquendo, Elan, Cepstral a ReadPlease, pouze s výjimkou systému Festival, který není příliš srozumitelný.

Celkově se dá říci, že současné české i zahraniční TTS systémy mají, až na pár výjimek, velmi dobrou kvalitu produkované řeči.

Na začátku práce byly stanoveny čtyři cíle. Popsat technologii převodu textu na řeč, představit současné softwarové nástroje pro převod textu na řeč, otestovat tyto softwarové nástroje a vyhodnotit výsledky.

Závěrem lze říci, že tyto cíle práce byly splněny.

6 Použitá literatura

[1] JAN, Uhlíř, et al. *Technologie hlasových komunikací*. [s.l.] : [s.n.], 2007. 276 s.

[2] *Fonetika a fonologie : Aktivní artikulační orgány* [online]. 2008 [cit. 2009-03-20]. Dostupný z WWW: <http://is.muni.cz/elportal/estud/ff/js08/fonetika/ucebnice/ch05s02s04.html>.

[3] ČERMÁKOVÁ, Kristýna. *Techniky zpěvu* [online]. 2008 [cit. 2009-03-20]. Dostupný z WWW: <www.gymnzidlo.cz/supl/123.doc>.

[4] VOKÁČOVÁ, Jarmila. *Řečový signál a jeho zpracování pomocí kepstrální analýzy* [online]. [2007-2008] [cit. 2009-03-12]. Dostupný z WWW: <http://homel.vsb.cz/~mor196/vok028.pdf>.

[5] MATOUŠEK, Jindřich. *Syntéza řeči : Přednáška z předmětu Neuronové sítě pro humanitní studia (NEUH)* [online]. 2004 [cit. 2009-01-15]. Dostupný z WWW: <http://control.zcu.cz/~radova/teaching/neuh/synteza.pdf>.

[6] PSUTKA, Josef, et al. *Mluvíme s počítačem česky*. [s.l.] : [s.n.], 2006. 752 s.

[7] *LINUXZONE* [online]. 2002 [cit. 2008-10-15]. Dostupný z WWW: <http://www.linuxzone.cz>.

[8] GRANDISCH, Michal. *Syntéza řeči* [online]. 2003 [cit. 2008-10-15]. Dostupný z WWW: <http://www.fi.muni.cz/usr/jkucera/pv109/2003/xgrandis.htm>.

[9] *Encyklopedie Wikipedia* [online]. [2007] [cit. 2008-11-20]. Dostupný z WWW: <http://wikipedia..cz>.

[10] KINCL, Jiří. *Rozpoznávání a syntéza řeči*. [s.l.] : [s.n.], 1985. 32 s.

[11] *DZR : Formantová syntéza* [online]. [2001] [cit. 2008-11-01]. Dostupný z WWW: <http://noel.feld.cvut.cz/vyu/dzr/dzr12/>.

[12] *SpeechTech* [online]. 2006 [cit. 2009-02-05]. Dostupný z WWW: <http://www.speechtech.cz>.

[13] *Katedra kybernetiky ZČU : Umělá inteligence* [online]. 2009 [cit. 2009-02-05]. Dostupný z WWW: <http://ui.kky.zcu.cz>.

[14] *FROG Systems* [online]. 1995-2003 [cit. 2009-02-05]. Dostupný z WWW: <http://www.frog.cz>.

[15] *Wizzard Software Corporation* [online]. 1995-2008 [cit. 2009-02-12]. Dostupný z WWW: <http://www.wizzardsoftware.com/searchresult.php?sw=voices>.

[16] *Nuance : RealSpeak* [online]. 2002-2009 [cit. 2009-02-12]. Dostupný z WWW: <http://www.nuance.com/realspeak/>.

[17] *Loquendo : global supplier of speech recognition and speech synthesis technology and solutions* [online]. 2001-2009 [cit. 2009-02-12]. Dostupný z WWW: <http://www.loquendo.com/>.

[18] *Text to Speech & Voice Solutions - Acapela Group* [online]. [2003] [cit. 2009-02-12]. Dostupný z WWW: <http://www.acapela-group.com>.

[19] *Cepstral Text-to-Speech* [online]. 2009 [cit. 2009-02-12]. Dostupný z WWW: <http://www.cepstral.com>.

[20] *SVOX - Embedded Text-to-Speech* [online]. 2000-2008 [cit. 2009-02-13]. Dostupný z WWW: <http://www.svox.com/>.

[21] *Formantová analýza a syntéza* [online]. 2007 [cit. 2009-03-20]. Dostupný z WWW: <http://amber.feld.cvut.cz/user/cmejla/dzr12/formant.htm>.

[22] *ReadPlease* [online]. 1999-2005 [cit. 2009-03-20]. Dostupný z WWW: <http://www.readplease.com/english/order/>.

[23] *Meyer Sound Laboratories Inc. 2009* [online]. 2009 [cit. 2009-02-19]. Dostupný z WWW: <http://www.meyersound.com/support/papers/speech/mrtlist.htm>.

[24] *Epos : A Free Text Speech Synthesis System* [online]. 2005 [cit. 2009-02-19]. Dostupný z WWW: <http://epos.ure.cas.cz/>.

[25] *AT&T Labs Research* [online]. 2008 [cit. 2009-02-12]. Dostupný z WWW: <http://www.research.att.com>.

[26] *Interactive Demo of SVOX* [online]. 2007 [cit. 2009-02-13]. Dostupný z WWW: <http://www.tik.ee.ethz.ch/cgi-bin/w3svox>.

[27] *The Centre for Speech Technology Research : Festival* [online]. [2004] [cit. 2009-02-19]. Dostupný z WWW: <http://www.cstr.ed.ac.uk/projects/festival>.

7 Seznam použitých zkratek

TTS	syntéza řeči z textu neboli převod textu na řeč (Text-to-speech)
NLP	zpracování přirozeného jazyka (Natural Language Processing)
LTS	modul fonetické transkripce (Letter-to-sound)
GP	generátor prozodie (Prosody generator)
MSA	morfologicko-syntaktický analyzátor (Morpho-syntactit analyzer)
F_0	základní hlasivkový tón
ToBI	symbolický transkripční systém, který vyznačuje a rozlišuje hranice mezi promluvovými úseky (Tones and Break Indices)
INTSINT	symbolický transkripční systém popisující význačné události intonační křivky pomocí symbolů (INTernational Transciription System for INTonation)
ARTIC	český softwarový nástroj pro převod textu na řeč (Artifical Talker in Czech)
WAV	zvukový formát (WAVeform audio format)
DDETTS	nadstavba pro interpretaci DDE u programu CS-VOICE
DDE	technologie pro komunikaci mezi více aplikacemi (Dynamic Data Exchange)
NetTTS	nadstavba pro převod textu na soubory WAV pro síť u programu CS-VOICE
C++	objektově orientovaný programovací jazyk
e-kniha	elektronická kniha
e-learning	výuka po internetu
BSD	licence pro svobodný software (Berkeley Software Distribution)
MBROLA	zahraniční softwarový nástroj pro převod textu na řeč (Multi Band Resynthesis Overlap and Add)
DRT	test diagnostikou rýmu patřící do skupiny testů srozumitelnosti (Diagnostic Rhyme Test)
MRT	test modifikací rýmu patřící do skupiny testů srozumitelnosti (Modified Rhyme Test)

SUS	test sémanticky nepredikovatelných vět patřící do testů srozumitelnosti, konkrétně do testů identifikace skupin (Semantically unpredictable Sentence)
MOS	test přirozenosti posuzující řeč podle celkové kvality (Mean Opinion Score)
CCR	test porovnávání párů patřící do testů přirozenosti (Category Comparison Rating)

8 Seznam příloh

PŘÍLOHA 1: DEMO verze programu ARTIC, zdroj [12]

PŘÍLOHA 2: program CS-VOICE, zdroj [14]

PŘÍLOHA 3: DEMO verze programu EPOS, zdroj [24]

PŘÍLOHA 4: DEMO verze programu AT&T Natural VoicesTM, zdroj [25]

PŘÍLOHA 5: DEMO verze programu RealSpeak ™, zdroj [16]

PŘÍLOHA 6: DEMO verze programu Loquendo, zdroj [17]

PŘÍLOHA 7: DEMO verze programu Elan, zdroj [18]

PŘÍLOHA 8: DEMO verze programu Cepstral, zdroj [19]

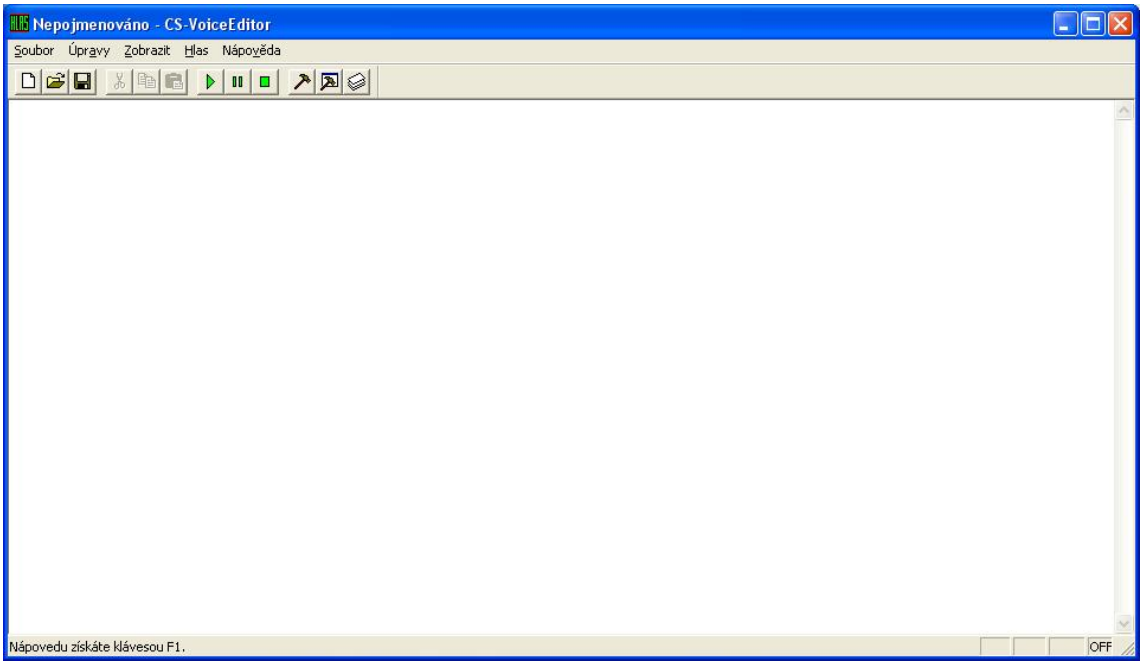
PŘÍLOHA 9: DEMO verze programu SVOX, zdroj [26]

PŘÍLOHA 10: DEMO verze programu Festival, zdroj [27]

PŘÍLOHA 11: program ReadPlease, zdroj [22]

PŘÍLOHA 1: DEMO verze programu ARTIC, zdroj [12]

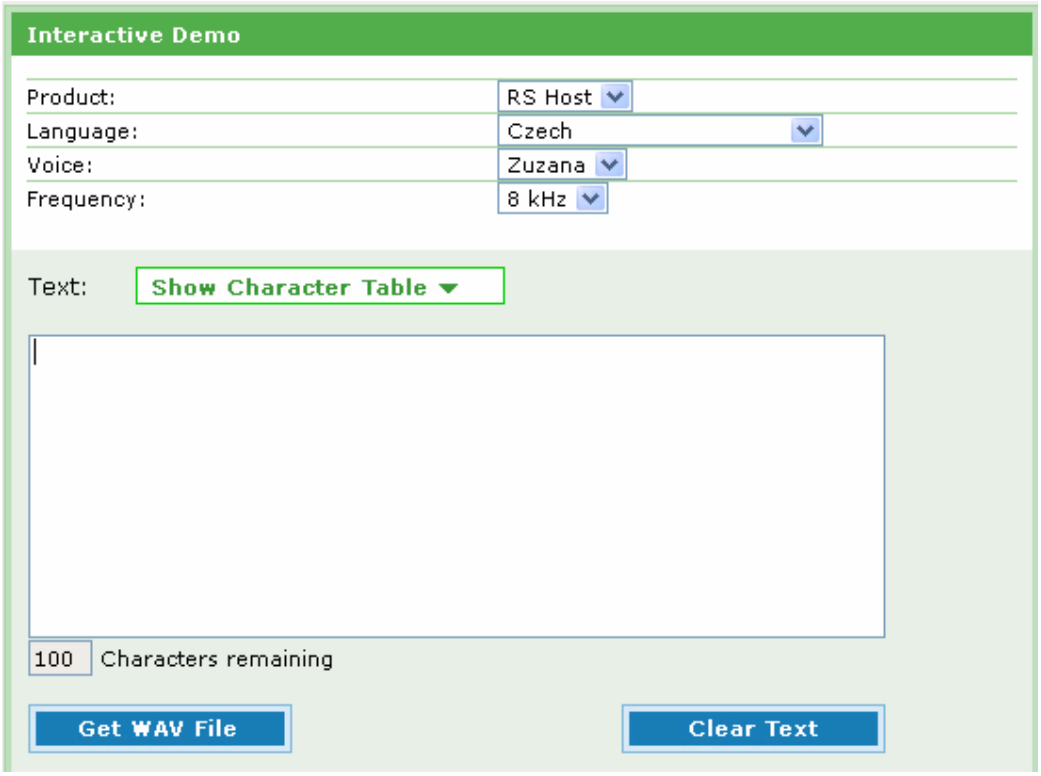
PŘÍLOHA 2: program CS-VOICE, zdroj [14]



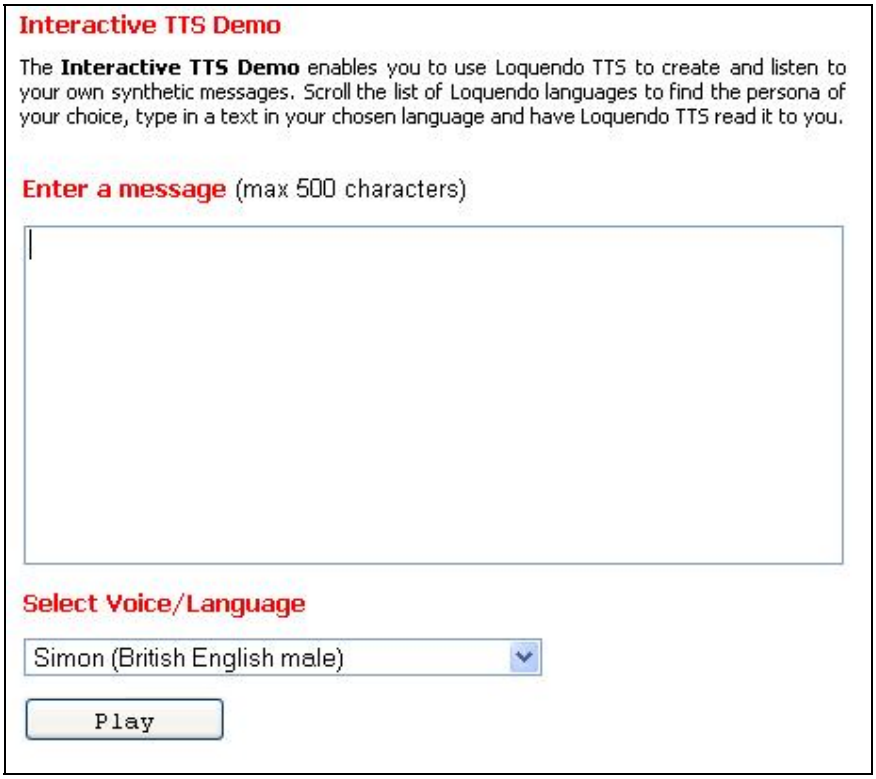
PŘÍLOHA 3: DEMO verze programu EPOS, zdroj [24]

PŘÍLOHA 4: DEMO verze programu AT&T Natural VoicesTM, zdroj [25]

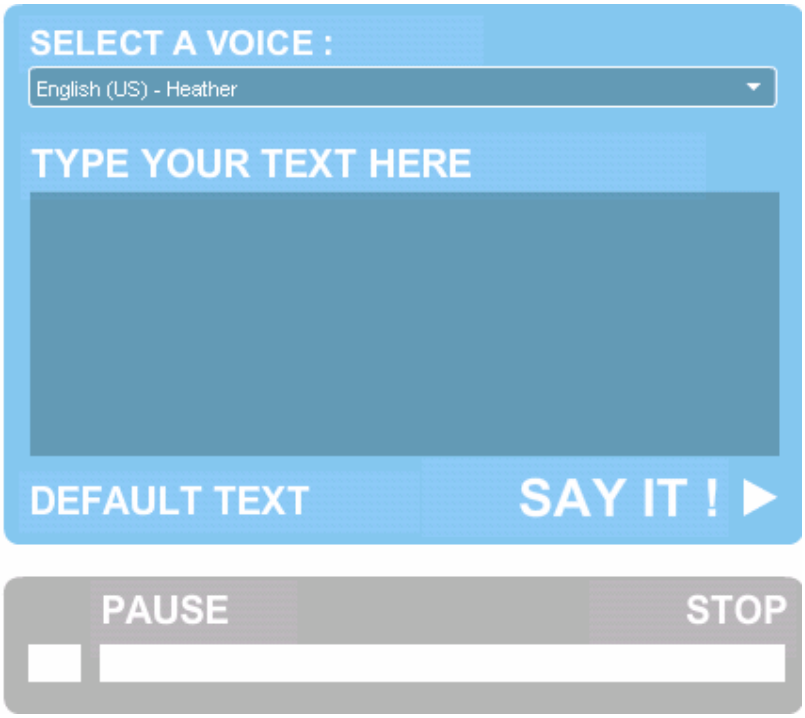
PŘÍLOHA 5: DEMO verze programu RealSpeak ™, zdroj [16]



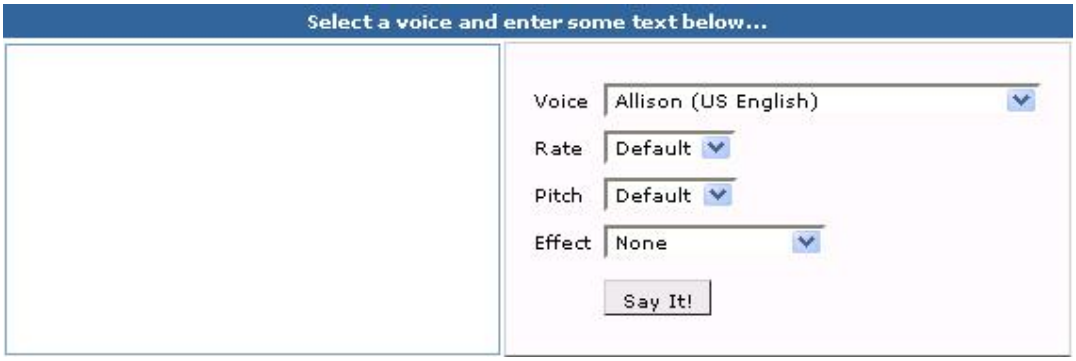
PŘÍLOHA 6: DEMO verze programu Loquendo, zdroj [17]



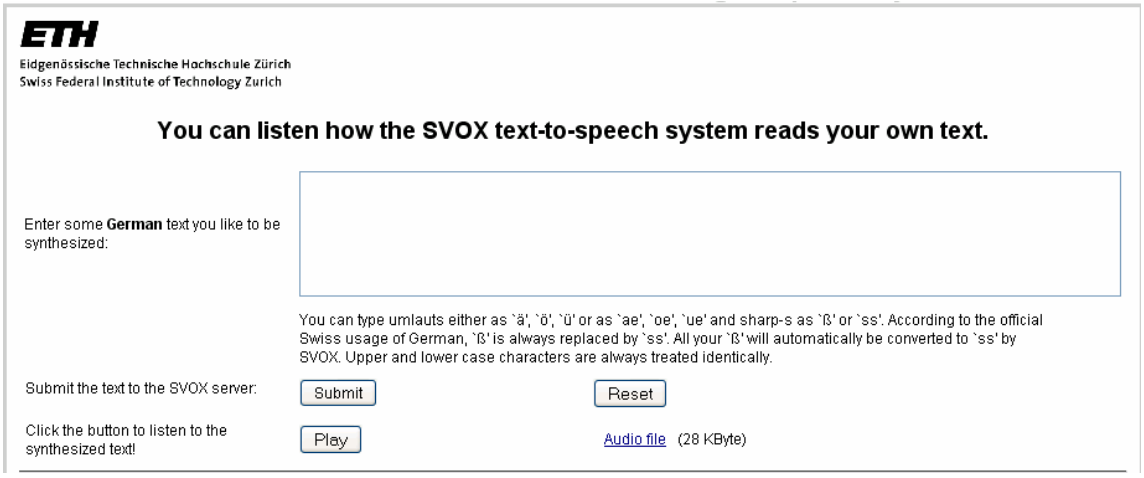
PŘÍLOHA 7: DEMO verze programu Elan, zdroj [18]



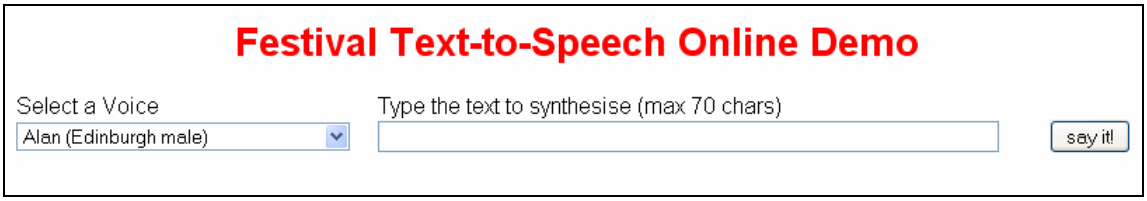
PŘÍLOHA 8: DEMO verze programu Cepstral, zdroj [19]



PŘÍLOHA 9: DEMO verze programu SVOX, zdroj [26]



PŘÍLOHA 10: DEMO verze programu Festival, zdroj [27]



PŘÍLOHA 11: program ReadPlease, zdroj [22]