# METÓDY UČENIA PRE SYNTÉZU REČI

Renata Rybárová, Gregor Rozinaj


STU
SLOVENSKÁ TECHNICKÁ
UNIVERZITA V BRATISLAVE

# METÓDY UČENIA PRE SYNTÉZU REČI

Renata Rybárová, Gregor Rozinaj

Slovenská technická univerzita v Bratislave
2013

Ing. Renata Rybárová, PhD., doc. Ing. Gregor Rozinaj, PhD.

**METÓDY UČENIA PRE SYNTÉZU REČI**



Oponenti: prof. Ing. Jozef Juhár, CSc.
prof. Ing. Dušan Levický, CSc.

Edícia a návrh obálky: Ing. Ivan Minárik

Vydalo: Nakladateľstvo STU v Bratislave
Bratislava 2013

ISBN 978-80-227-4033-3 (tlačená verzia)
ISBN 978-80-227-4034-0 (CD verzia)

# Predslov

Predkladaná monografia so zameraním na inteligentnú syntézu reči je výsledkom dlhej činnosti orientovanej na túto problematiku, realizovanej na báze viacerých prác riešených na Ústave telekomunikácií FEI STU v Bratislave.

Hlavnou motiváciou na vydanie tejto publikácie je aj absencia literatúry zameranej na inteligentnú syntézu. Monografia sa zaoberá návrhom inteligentného rečového syntetizátora a modifikáciou prozódie slovenskej reči. Jedným z hlavných pilierov je analýza a návrh modulového syntetizátora s implementovaným učiacim sa systémom pre každý modul (modul fonetickej transkripcie, fonetickej transkripcie skratiek, určovania slovných druhov a modifikácie prozódie). Druhým nosným pilierom je analýza zmeny prozódie pomocou sínusoidálnych modelov a ich využitie na komprimovanie databázy pre rečový syntetizátor.

K vytvoreniu predloženej monografie významnou mierou prispela dlhoročná vedecko-výskumná činnosť autorov, ich kolegov a študentov, ktorí výsledkami experimentov a radami prispeli k tvorbe tohto diela.

Poďakovanie patrí oponentom, prof. Dušanovi Levickému a prof. Jozefovi Juhárovi za pozorné prečítanie a cenné rady a pripomienky, ktoré prispeli k zvýšeniu kvality publikácie.

Naše poďakovanie patrí Ing. Martinovi Turi Nagyovi za podporu, cenné rady a program na sinuoidálnu analýzu a tiež teoretické podklady ku kompresii reči. Ďalšie poďakovanie patrí prof. Pavlovi Podhradskému za pomoc a ochotu pri dokončovaní našej knižky. Tiež sa chceme poďakovať Ing. Ivanovi Minárikovi, editorovi, ktorý sa zanietene venoval finálnej grafickej a vizuálnej úprave monografie.

V neposlednom rade patrí vďaka našim rodinám za ich podporu pri tvorbe tejto publikácie.

Vydanie tejto publikácie bolo realizované s podporou projektu HBB-Next, FP7-ICT-2011-7, 287848 a projektu IMUROSA, VEGA 1/0708/13.

Bratislava, august 2013 Autori

# Obsah

Predslov ..... 7
Zoznam použitých skratiek ..... 13
1 Úvod ..... 17
2 Syntéza reči ..... 19
2.1 Reč a jej vlastnosti ..... 19
2.2 História syntézy reči ..... 26
2.3 Syntéza reči ..... 28
*2.3.1 Syntéza spájaním jednotiek ..... 28*
*2.3.2 Formantová syntéza ..... 29*
*2.3.3 Artikulačná syntéza ..... 29*
*2.3.4 HMM syntéza ..... 30*
2.4 Základná schéma syntetizátora ..... 30
*2.4.1 SAMPA abeceda ..... 31*
2.5 Najnovšie metódy v oblasti syntézy reči ..... 34
2.6 Skvalitnenie syntézy a princíp učenia v procese syntézy reči ..... 35
2.7 Prozódia reči ..... 38
2.8 Úprava hlasu a emócií ..... 41
2.9 Možnosti modifikácie ľudskej reči ..... 42
2.10 Kódovanie reči pomocou TTS ..... 43
3 Využitie sínusoidálnych modelov na modifikáciu reči ..... 47
3.1 Štandardný model sínusoíd plus šum ..... 48
3.2 Analýza a odhad parametrov signálu pomocou sínusoidálneho modelu ..... 49
3.3 Harmonický plus šumový model ..... 55
3.4 Analýza reči pomocou HNM modelu ..... 56
3.5 Syntéza pomocou HNM modelu ..... 58
*3.5.1 Modifikácia prozódie akustických jednotiek ..... 58*
*3.5.2 Spájanie akustických jednotiek ..... 59*
*3.5.3 Generovanie tvaru vlny v časovej oblasti ..... 61*
4 Kompresia a modelovanie reči ..... 63
4.1 História kódovania reči ..... 64
4.2 Vlnové kodéry ..... 66
*4.2.1 Modelovanie v časovej oblasti ..... 66*
*4.2.2 Modelovanie vo frekvenčnej oblasti ..... 67*
*4.2.3 Modelovanie založené na vlastnostiach reči ..... 68*
*4.2.4 Hybridné modelovanie ..... 70*
4.3 Kompresné algoritmy ..... 73
*4.3.1 Wave ..... 75*

*4.3.2 MPEG-1* ..... *76*
*4.3.3 MP3* ..... *77*
*4.3.4 MPEG-2* ..... *79*
*4.3.5 MPEG-4 ACC* ..... *79*
*4.3.6 Ogg* ..... *81*
*4.3.7 Windows Media Audio* ..... *82*
*4.3.8 Porovnanie kompresných algoritmov* ..... *82*
4.4 Vyhodnotenie kódovacích algoritmov ..... 83
5 Skvalitnenie syntézy a princíp učenia v procese syntézy reči ..... 87
5.1 Inteligentný systém pre syntézu reči ..... 87
5.2 Spracovanie prirodzeného textu ..... 90
*5.2.1 Morfologicko-syntaktická analýza* ..... *91*
*5.2.2 Fonetická transkripcia* ..... *93*
*5.2.3 Generovanie prozódie* ..... *93*
5.3 Modifikácia pravidiel výslovnosti pri fonetickej transkripcii ..... 96
5.3.1 Spodobovanie ..... 97
5.3.2 Návrh modulu transkripcie ..... 99
5.3.3 Implementácia modulu transkripcie s možnosťou učenia ..... 100
5.3.4 Databáza opráv ..... 102
5.3.5 Inteligentné učenie ..... 106
5.4 Modul určovania slovných druhov ..... 107
*5.4.1 Návrh modulu morfologickej analýzy* ..... *110*
*5.4.2 Vytváranie trojfonémového slovníka* ..... *112*
*5.4.3 Vytváranie štvorfonémového a päťfonémového slovníka* ..... *113*
*5.4.4 Prehľadávanie slovníkov* ..... *114*
*5.4.5 Implementácia modulu morfológie* ..... *115*
*5.4.6 Frázovanie* ..... *117*
*5.4.7 Inteligentné učenie* ..... *120*
5.5 Voľba fonetickej transkripcie skratiek podľa kontextu ..... 121
*5.5.1 Návrh modulu transkripcie skratiek* ..... *122*
*5.5.2 Lexikón skratiek a akronymov* ..... *123*
*5.5.3 Hláskovanie skratiek a akronymov* ..... *124*
*5.5.4 Vysloviteľnosť skratiek a akronymov* ..... *126*
*5.5.5 Expanzia skratiek a akronymov* ..... *126*
*5.5.6 Skloňovanie skratiek a akronymov* ..... *128*
*5.5.7 Inteligentné učenie* ..... *131*
5.6 Modifikácia prozódie syntetizovaného textu ..... 132
*5.6.1 Typy viet v slovenskom jazyku* ..... *133*
*5.6.2 Návrh zmeny prozódie* ..... *140*
*5.6.3 Implementácia modulu na zmenu prozódie* ..... *143*
*5.6.4 Inteligentné učenie* ..... *145*

*5.7 Vzťah medzi parametrami sínusoíd pre rôzne tóny ..... 146*
*5.8 Využitie sínusoíd ako korpus pre rečový syntetizátor ..... 156*
6 Zhrnutie ..... 159
Zoznam použitej literatúry ..... 161
Register ..... 171

# Zoznam použitých skratiek

| Skratka | Anglický názov | Slovenský názov |
|---|---|---|
| AAC | Advanced Audio Coding | Štandard pre stratovú kompresiu |
| ADPCM | Adaptive Differential PCM | Adaptívna diferenčná PCM |
| ANA | Adaptive Noise Addition | Adaptívne pridávanie šumu |
| AR | Autoregressive Model | Autoregresívny model |
| ASCII | American Standard Code for Information Interchange | Americký štandard kódovanie pre výmenu informácií |
| ATC | Adaptive Transform Coding | Adaptívne transformačné kódovanie |
| CART | Classification And Regression Trees | Klasifikačné a regresné stromy |
| CCITT | International Telegraph and Telephone Consultative Committee | Medzinárodná telegrafná a telefónna organizácia zaoberajúca sa štandardami v tejto oblasti |
| CELP | Code Excited Linear Prediction | Kódovo-budená lineárna predikcia |
| DAM | Diagnostic Acceptability Measure | Diagnostické meranie akceptovateľnosti |
| DCT | Discrete Cosine Transform | Diskrétna kosínusová transformácia |
| DFT | Discrete Fourier transform | Diskrétna Fourierova transformácia |
| DI | Distortion Index | Index skreslenia |
| DM | Delta Modulation | Delta modulácia |
| DPCM | Differential PCM | Diferenčná PCM |
| DRT | Diagnostic Rhyme Test | Diagnostický rýmovací test |
| DSP | Digital Signal Processing | Digitálne spracovanie signálov |
| DTW | Dynamic Time Warping | Algoritmus nelineárnej časovej normalizácie používaný k porovnávaniu rečových segmentov s nerovnakým počtom rámcov |
| $F_0$ | | Základná hlasivková frekvencia |
| FEI STU | | Fakulta elektrotechniky a informatiky Slovenskej technickej univerzity |
| FFT | Fast Fourier Transform | Rýchla Fourierova transformácia |
| FIFA | Fédération Internationale de Football Association (English: International Federation of Association Football) | Medzinárodná futbalová federácia |
| GMM | Gaussian Mixture Model | Funkcia hustoty pravdepodobnosti s normálnym rozdelením s viacerými mixami |

| Skratka | Anglický názov | Slovenský názov |
|---|---|---|
| GSM | Global System for Mobile Communications | Systémy pre globálnu mobilnú komunikáciu |
| HMM | Hidden Markov Models | Skrytý Markovov model |
| HNM | Harmonic plus Noise Model | Harmonický plus šumový model |
| CHATR | | Typ syntetizátora vykonávajúceho syntézu výberom segmentov z korpusu vyvinutý v laboratóriách ATR (Advanced Telecommunications Research Institute) |
| INSINT | International Transcription System for Intonation | Medzinárodný systém pre transkripciu intonácie |
| IPA | International Phonetic Alphabet | Medzinárodná fonetická abeceda |
| KLT | Karhunen-Loeve Transformation | Karhunen-Loeveho transformácia |
| LNRE | Large Number of Rare Events | Veľký počet zriedkavých udalostí |
| LPC | Linear Predictive Coding | Lineárne predikčné kódovanie |
| LTP | Long Term Prediction | Dlhodobá predikcia |
| LTS | Letter to Sound | Pravidlá na prepis písmen na znaky |
| MBE | Multiband Excitation | Multipásmové budenie |
| MBR TD PSOLA (=MBROLA) | Multi-band Resynthesis Time – Domain Pitch Synchronous Overlap Add | Metóda resyntézy rečových segmentov za účelom dosiahnutia lepšej kvality syntézy TD-PSOLA algoritmom |
| MBROLA | MultiBand Resynthesis Overlap Add Method | Metóda resyntézy rečových segmentov za účelom dosiahnutia lepšej kvality syntézy TD-PSOLA algoritmom |
| MDCT | Modified Discrete Cosine Transformation | Modifikovaná diskrétna kosínusová transformácia |
| MIPS | Microprocessor Instructions per Second | Počet mikroprocesorových inštrukcií za sekundu |
| MIS | Multiple Instance System | Systém výberu segmentov, v ktorom je pre jeden segment z cieľa syntézy viacero kandidátov |
| MOS | Mean Opinion Score | Skóre priemernej mienky |
| MOVs | Model Output Variables | Výstupné premenné modelu |
| MPC | Multi-pulse Coding | Multipulzné kódovanie |
| MPEG | Motion Picture Experts Group | Skupina expertov pre pohyblivý obraz |
| MSS | Miles Sound System | Zvukový systém Miles |
| NLP | Natural Language Processing | Spracovanie prirodzeného jazyka |
| ODG | Objective Difference Grade | Objektívny stupeň rozdielu |

| Skratka | Anglický názov | Slovenský názov |
|---|---|---|
| OLA | Overlap Add Method | Súčet s prekryvom |
| OSN | United Nations Organization | Organizácia Spojených národov |
| PCM | Pulse Code Modulation | Pulzná kódová modulácia |
| PEAQ | Perceptual Evaluation of Audio Quality | Percepčné vyhodnotenie kvality zvuku |
| PESQ | Perceptual Evaluation of Speech Quality | Percepčné vyhodnotenie kvality reči |
| PK | Parametric Coding | Parametrické kódovanie |
| PNS | Perceptual Noise Substitution | Vnemové nahradenie šumu |
| PPP | Point to Point Protocol | Komunikačný protokol linkovej vrstvy |
| PPP | Purchasing Power Parity | Parita kúpnej sily |
| PPP | Public Private Partnership | Projekty verejno – súkromného partnerstva |
| PS | Parametric Stereo | Parametrické stereo |
| PSOLA | Pitch Synchronous Overlap Add Method | metóda modifikácie prozodických parametrov prekladaním a spočítavaním rečových rámcov synchronizovanými s $F_0$ |
| PSQM | Perceptual Speech Quality Measurement | Percepčné meranie kvality reči |
| RELP | Residual Excited Linear Prediction | Reziduálne budená lineárna predikcia |
| RIFF | Resource Interchange File Format | Súborový formát pre ukladanie multimediálnych zvukových a obrazových predlôh |
| SAMPA | Speech Assessment Methods Phonetic Alphabet | fonetická abeceda vychádzajúca z medzinárodnej IPA abecedy používajúca 7 bitové tlačiteľné ASCII znaky |
| SAV | | Slovenská akadémia vied |
| SBR | Spectral Band Replication | Spektrálna odpoveď pásma, metóda používaná na zlepšenie kódovania reči |
| SDKU | | Slovenská demokratická a kresťanská únia |
| SEGSNR | Segmental SNR | Segmentálne SNR |
| SFT(SFT) | Short-time Fourier Transform | Krátkodobá Fourierova transformácia |
| SII | Speech Intelligibility Index | Index zrozumiteľnosti reči |
| SIS | Single Instance System | Systém výberu segmentov v ktorom je pre jeden segment z cieľa syntézy jeden kandidát |
| SN(SNM) | Sinusoidal Plus Noise Model | Sínusoidálny plus šumový model |
| SNK | | Slovenský Národný Korpus |
| SNR | Signal to Noise Ratio | Odstup signál/šum |

| Skratka | Anglický názov | Slovenský názov |
|---|---|---|
| SR | Sinusoidal Regeneration | Sínusoidálna regenerácia |
| STC | Sinusoidal Transform Coding | Sínusoidálne transformačné kódovanie |
| STI | Speech Transmission Index | Index prenosu reči |
| STU | | Slovenská technická univerzita |
| TD PSOLA | Time-Domain Pitch Synchronous Overlap Add Method | Metóda modifikácie prozodických parametrov prekladaním a spočítavaním rečových rámcov synchronizovanými s F0, pracujúca v časovej oblasti |
| Tilt | Technology in Learning and Teaching | Technika na učenie |
| TNS | Temporal Noise Shaping | Audio technika používaná zvyčajne na digitálne spracovanie |
| ToBI | Tones and Break Indices | Indexovanie tónov a páuz |
| TTS | Text to Speech | Syntéza reči z textu |
| Twin VQ | Transform-domain Weighted Interleaved Vector Quantization | Audio štandard založený na váhovanom prekryve a vektorovej kvantizácií |
| UNESCO | United Nations Educational, Scientific and Cultural Organization | Organizácia Spojených národov pre výchovu, vedu a kultúru |
| USA | United States of America | Spojené štáty americké |
| UTF | 8-bit Unicode Transformation Format | 8-bitový transformačný formát typu Unicode |
| XML | eXtensible Markup Language | Rozšíriteľný značkovací jazyk |

# 1 Úvod

Komunikácia prostredníctvom hovorovej reči je základný a najdôležitejší prostriedok prenosu informácií medzi inteligentnými bytosťami. Tento proces je postup akcií, ktoré začínajú prípravou správy v mozgu hovoriaceho a končia rozpoznaním u poslucháča, zahŕňajúce aj rozpoznanie významu správy. Reč je základná súčasť nášho každodenného života. Či už ňou vyjadrujeme naše myšlienky, ciele, reakcie alebo pocity, sotva si uvedomujeme, že nielen hovoríme, ale prostredníctvom konverzácie priamo ovplyvňujeme myšlienky nášho komunikačného partnera.

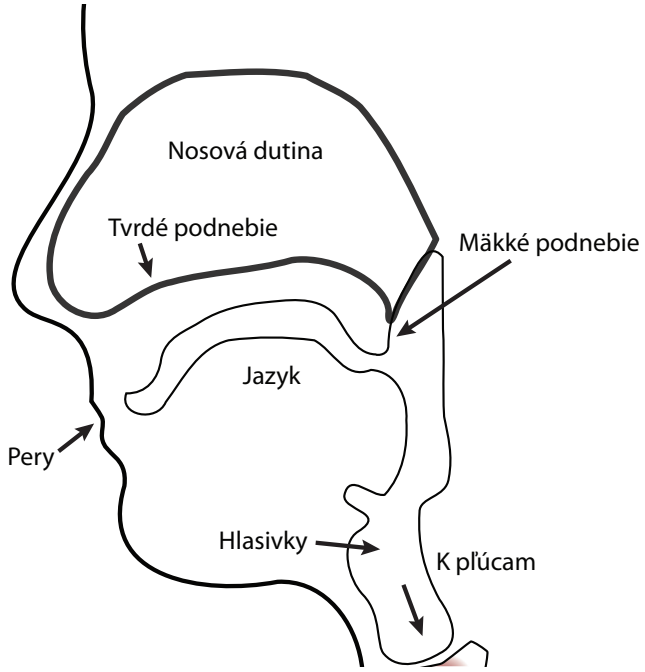
V niekoľkých posledných rokoch výskum v oblasti syntézy reči ako aj výpočtovej techniky zaznamenali výrazný pokrok. Môžeme zostrojiť rozpoznávač reči takmer pre každý jazyk. Avšak ak očakávame, že získame informácie z počítača pomocou hovorenej reči rovnako ľahko, ako sme zvyknutí v každodennej konverzácii, bude nás to stáť veľa práce. Syntetizovaná reč musí byť prirodzená, kontrolovateľná a efektívna ako na strane rozpoznávača, tak aj na strane syntetizátora.

Syntéza reči je vedná disciplína, ktorá sa zaoberá tvorbou umelého signálu ľudskej reči. Ľudia sa myšlienkou syntézy reči zaoberajú už dlhšiu dobu. Prvý opis hovoriaceho stroja publikoval už v roku 1791 Jan Wolfgang Kempelen. Postupne sa techniky syntézy zdokonaľovali. Pomocou softvérových syntetizátorov je možné robiť stále dokonalejšiu syntézu. Rovnako sa vylepšili metódy post spracovania reči, ktorým sa snažíme čo najviac priblížiť originálu a umožniť tým dokonalú komunikáciu stroja s človekom. Pri prenose reči nie je základnou podmienkou verný prenos pôvodného signálu, ale v prvom rade dobrá zrozumiteľnosť reči. Ak by sa podarilo vytvoriť dokonalú syntézu, zohľadňujúcu intonáciu, rytmus a farbu reči rečníka, umožnilo by to minimalizovať veľkosť dát prenášaných komunikačným kanálom pri prenose reči. Nemusela by sa prenášať celá reč, ale postačili by dáta charakterizujúce rozdiel medzi rečou rečníka a rečou syntetizátora. V súčasnosti existujú metódy, ktoré sa nesnažia zachovať pôvodný priebeh rečového signálu, ale signál modelujú s ohľadom na vlastnosti ľudského vnímania reči. Existuje veľa prístupov, ako syntetizovanú reč zdokonaľovať a kódovať. Tie najznámejšie sú opísané v tejto publikácii.

# 2 Syntéza reči

## 2.1 Reč a jej vlastnosti

Zdrojom rečových kmitov, ktoré sú fyzikálnou reprezentáciou reči, sú ľudské rečové orgány. Tie sa skladajú z hlasiviek, nosovej dutiny, hrdelnej dutiny, ústnej dutiny, mäkkého a tvrdého podnebia, zubov a jazyka (Obr. 1). K týmto orgánom je však nutné pripočítať ešte fundamentálny zdroj hlasovej energie – pľúca a s nimi funkčne spojené orgány.




***Obr. 1.*** *Rečové orgány človeka*

Rečové signály sú nestacionárne, iba na krátkych úsekoch (časové okno veľkosti 5-20 ms) ich môžeme považovať za kvázi-stacionárne. Z tohto dôvodu definujeme štatistické a spektrálne vlastnosti reči na krátkych blokoch. Z hľadiska kódovania môže byť reč klasifikovaná ako znelá a neznelá. Znelá reč je kvázi-periodická v časovej oblasti a harmonická vo frekvenčnej oblasti, zatiaľ čo neznelá reč je podobná náhodným signálom a je širokopásmová. Energia znelých hlások je väčšia ako energia neznelých hlások.

Základnou stavebnou jednotkou reči je fonéma. Je to najmenšia zmysluplná časť jazyka. Hláska je najmenšia samostatná, sluchom rozlíšiteľná udalosť hovorového jazyka. Difón je postupnosť dvoch nasledujúcich hlások. Na rozdiel od samostatne stojacich hlások je difón silne ovplyvnený prítomnosťou susedných foném. Ďalším špecifikom difónového prístupu je odlišnosť rovnakých difónov na začiatku, v strede a na konci slova, ako aj ich odlišnosť pri zmene prozódie vety. Spájaním foném a difónov do väčších celkov vznikajú slová a vety. Z viet vznikajú ucelené texty.

V slovenčine môžeme hlásky rozdeliť do dvoch základných skupín a ďalej podľa rôznych charakteristík v rámci skupiny [87]:

- samohlásky
  - krátke
  - dlhé
  - dvojhlásky (podľa trvania sú tieto tiež dlhé)
- spoluhlásky
  - podľa účasti hlasu
    - znelé
    - neznelé
  - podľa sluchového dojmu
    - plozíva
    - frikatívy
    - afrikatívy
  - podľa účasti nosnej dutiny
    - sonórne
    - orálne

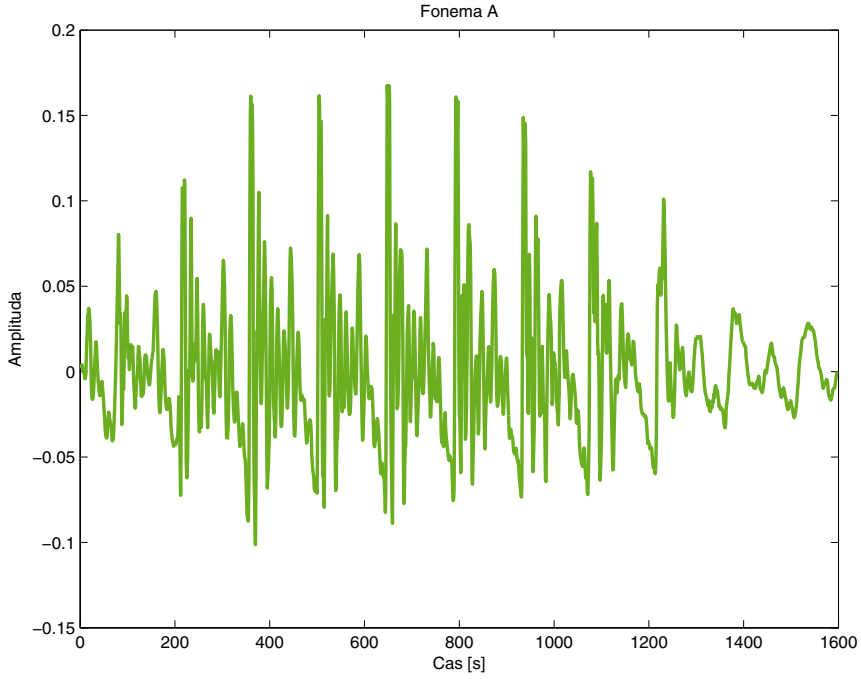
Základné delenie foném je na samohlásky a spoluhlásky, tie ďalej delíme na znelé a neznelé. Na tomto delení sa veľmi dobre opisuje vznik hlások.

Znelé hlásky vznikajú tým, že prúd vzduchu, ktorý vydychujeme pri rozprávaní prechádza okolo hlasiviek, kmitajúcich s určitou frekvenciou - fundamentálnou frekvenciou $F_0$. Následne tento periodický signál môže byť obohatený o ďalšiu zložku, ktorá vzniká rezonovaním v následných častiach hovorového ústrojenstva, ako sú hrtan, nosová dutina alebo ústna dutina. Frekvencia kmitov závisí od tlaku vzduchu a od svalového napätia hlasiviek. Fundamentálna frekvencia charakterizuje základný tón ľudského hlasu. Táto je iná u detí ako u dospelých, a iná u mužov a žien. U väčšiny ľudí sa pohybuje v rozmedzí 80 až 400 Hz (muži v rozmedzí 80 až 180 Hz, ženy v rozmedzí 165 až 255 Hz, deti v rozmedzí 260 až 400 Hz). Základná hlasivková frekvencia je prítomná pri tvorení všetkých znelých zvukov, čiže samohlások a znelých spoluhlások. Ďalšie frekvencie, ktoré sa vyskytujú v akustickom spektre samohlások, sa nazývajú formanty. Označujú sa číslami, začínajúc formantom s najnižšou frekvenciou $F_1$, $F_2$,.., $F_N$. Formanty obsahujú skoro celú informáciu zo signálu. Hlavné anatomické komponenty, ktoré zapríčiňujú zmenu výšky a intenzity samohlások, sú pery, čeľuste, jazyk a mäkké podnebie. Rečový trakt môžeme efektívne modelovať pomocou celopólového lineárneho systému. Znelá reč preukazuje

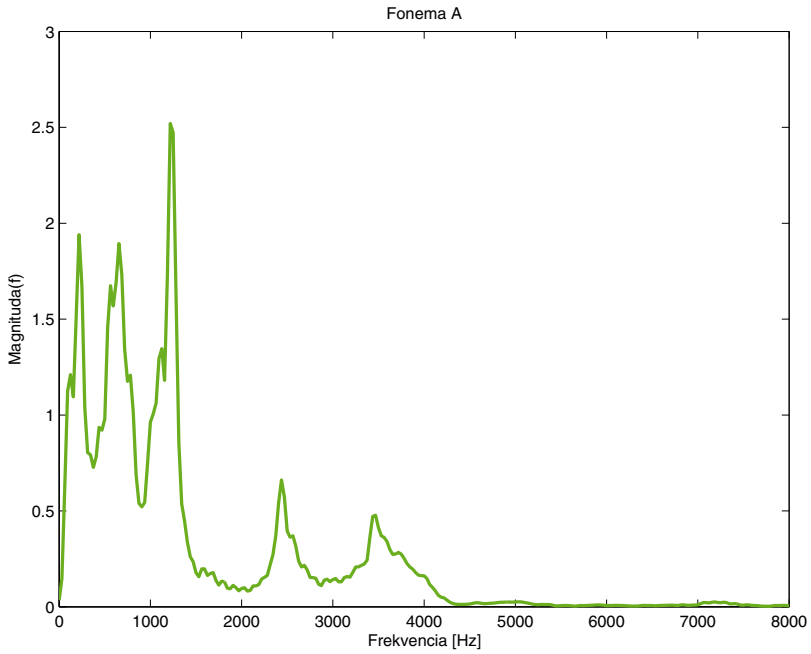
efekty vibrácie hlasiviek a tento efekt nazývame kvázi-periodicita. Na obrázku spektra samohlásky (Obr. 3, Obr. 4) je vidieť prvé formanty. Je dôležité podotknúť, že formantová štruktúra sa nad 4 kHz rozpadá a dominuje šum, ktorý vzniká turbulentným tokom vzduchu cez hlasový trakt. Z obrázkov je zrejmá aj dolnopriepustná povaha znelej reči.

Podstatná vlastnosť spoluhlások je prítomnosť šumu v ich akustickom spektre. Spoluhlásky sa vytvárajú vzduchovou turbulenciou, ktorá vzniká trením vydychovaného vzduchu o prekážku vytvorenú artikulačnými orgánmi (napr. špičkou jazyka, zubami a perami). Šumovú povahu spoluhlások je vidieť z časovej aj frekvenčnej oblasti. Spektrum neznelej reči má hornopriepustný charakter a jej energia je podstatne nižšia ako pri znelej reči.

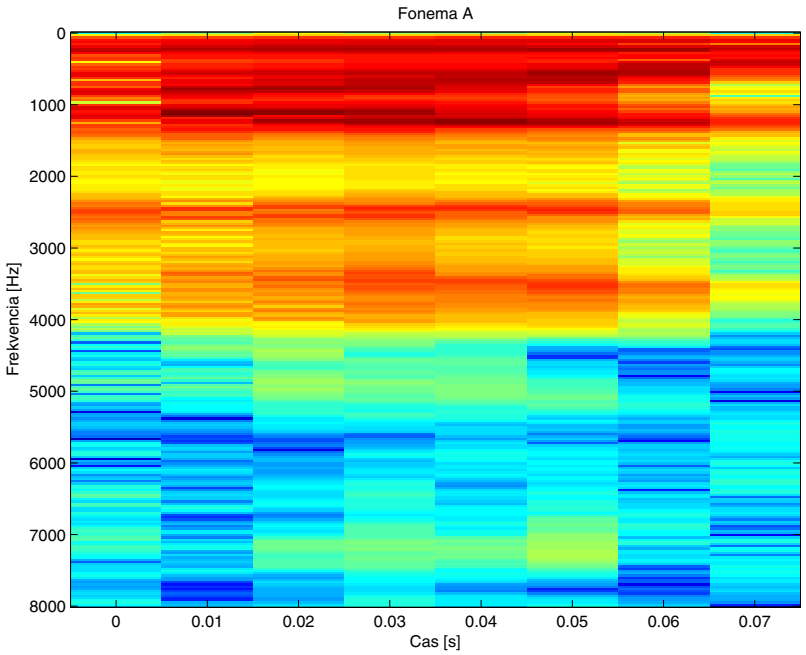
Zvuky delíme podľa priebehu na periodické a neperiodické. Periodické označujeme ako tóny (patria tam všetky samohlásky a znelé spoluhlásky), neperiodický priebeh majú neznelé spoluhlásky. Príklady časového priebehu samohlásky a znelej a neznelej spoluhlásky sú na obrázkoch (Obr. 2, Obr. 5, Obr. 8). Veľa zvukov v prírode je kombináciou tónových a šumových zložiek. Aj v ľudskej reči sa uplatňujú zložky tónové (samohlásky), ďalej zvuky kombinujúce tónovú a šumovú zložku (znelé spoluhlásky) a čisté šumy (neznelé spoluhlásky).



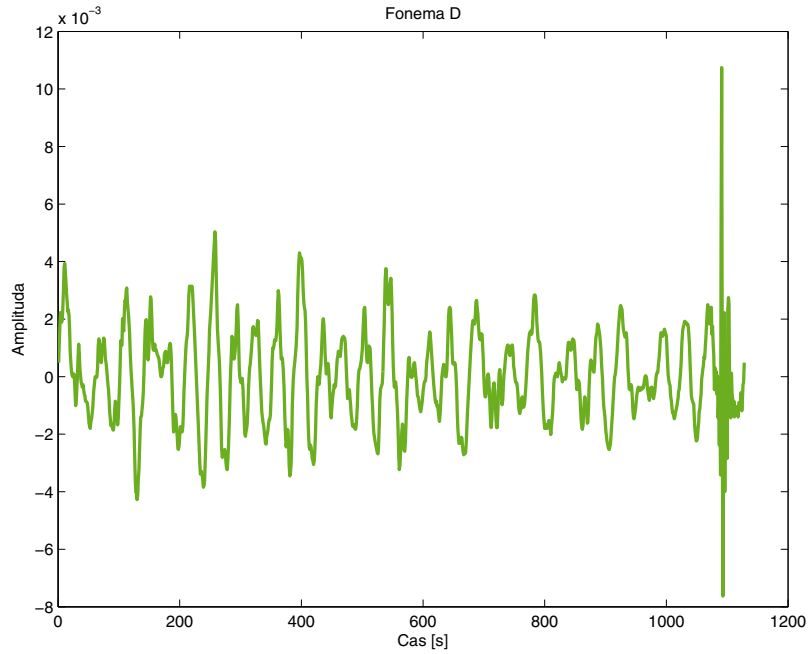

***Obr. 2.*** *Časový priebeh samohlásky „a“*



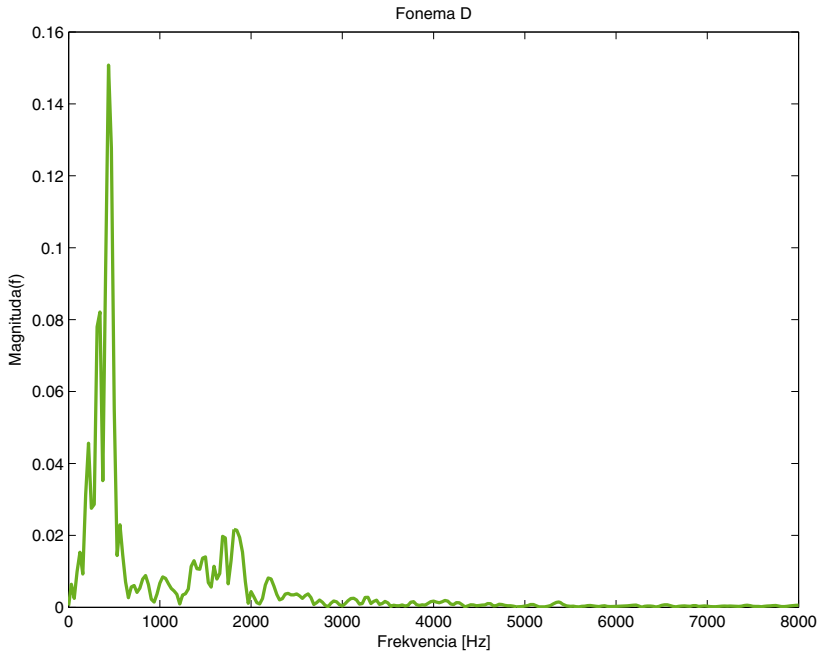

***Obr. 3.*** *Frekvenčný priebeh samohlásky „a“*



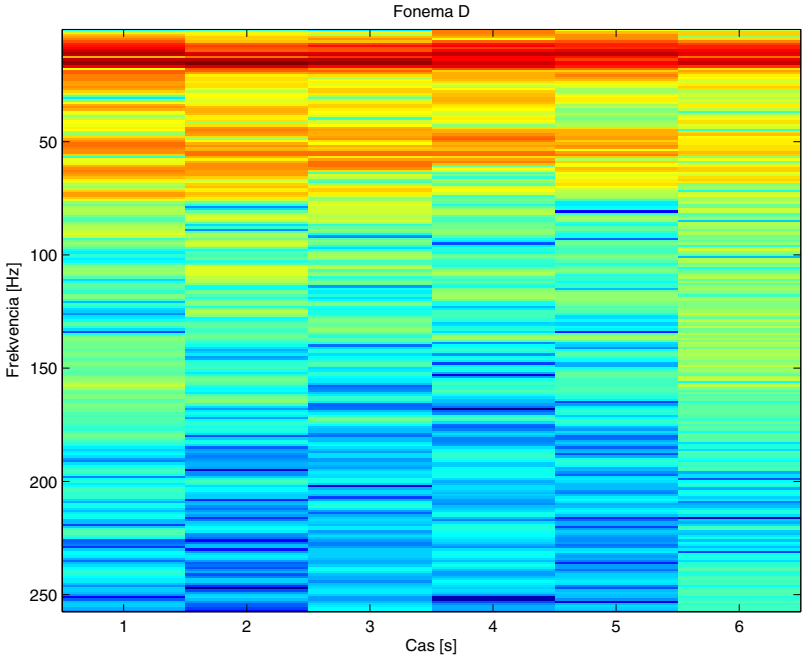

***Obr. 4.*** *Spektrogram samohlásky „a“*



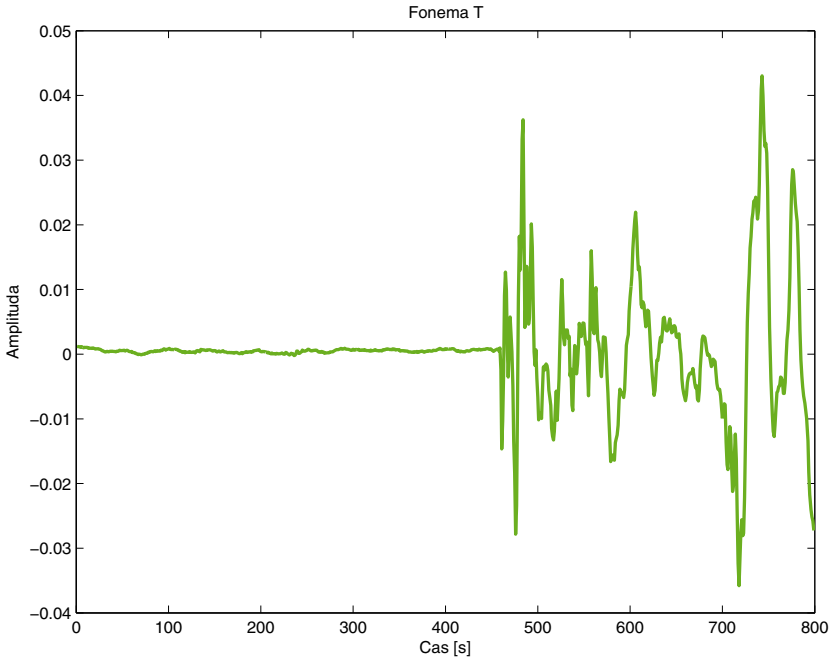

***Obr. 5.*** *Časový priebeh spoluhlásky „d“*



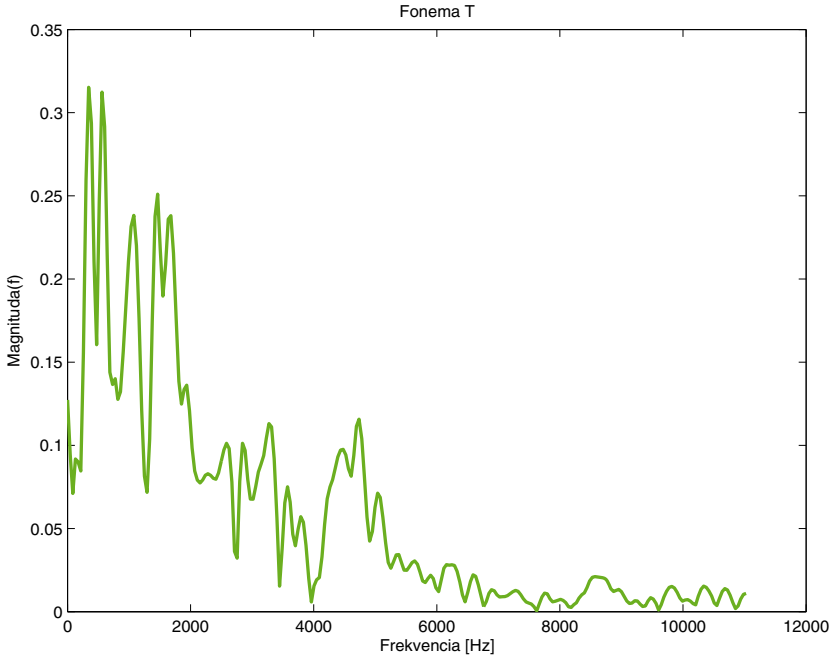

***Obr. 6.*** *Frekvenčný priebeh spoluhlásky „d“*




***Obr. 7.*** *Spektrogram spoluhlásky „d“*




***Obr. 8.*** *Časový priebeh spoluhlásky „t“*




***Obr. 9.*** *Frekvenčný priebeh spoluhlásky „t“*

Šum tvoriaci spoluhlásky môžeme rozdeliť na tri typy:

- jednorazový, veľmi krátky šum pulzového typu, pred vznikom šumu dochádza k prerušeniu zvukového prúdu. Z auditívneho hľadiska sa označujú ako plozíva. Patria sem: *p, b, t, ť, d, ď, k, g*.
- trvajúci šum nadväzuje na predošlé zvuky súvisle. Z auditívneho hľadiska sa označujú ako frikatívy. Patria sem: *f, v, s, š, z, ž, h, ch*.
- ako kombináciu oboch spôsobov, to znamená krátka pauza a potom explózia. Z auditívneho hľadiska sa ich zvuk podobá frikatívam, preto sa nazývajú afrikatívy. Patria sem: *c, č, dz, dž*.

Pre samohlásky je typická existencia väčšieho počtu formantov – miest s vysokou koncentráciou akustickej energie. Poznáme však aj hlásky, ktoré majú slabú šumovú zložku, a ani tónová zložka u nich nie je rozvinutá. Tieto hlásky sú z akustického hľadiska na rozmedzí medzi samohláskami a spoluhláskami, ale radíme ich medzi spoluhlásky.

Sonóry sú hlásky, ktoré majú šumovú zložku podobnú spoluhláskam, je však oslabená (napr. šumová zložka *m* je slabšia ako šumová zložka znelého *b*). Okrem základného formantu majú iba jeden ďalší formant, ktorý vznikol rezonanciou laryngálneho hlasu. Tento charakter majú hlásky *l* – ové, *r* – ové a väčšina nosových hlások ako *m, n, j*. Delíme ich na nazály (*m, n, ň*) a likvidity (*l, ľ, ĺ, r, ŕ*) [87].

Pri vyslovovaní zvukov musia časti nášho ústrojenstva zaujať počiatočnú polohu zodpovedajúcu danému zvuku. Počas doby vyslovovania zvuku orgány menia svoju polohu, pričom k tejto zmene nemôže dochádzať okamžite. Nadobudnutie finálnej polohy

pre každú hlásku potrebuje istý čas. Zásadný jav, ktorý možno pri tomto procese pozorovať, a ktorý komplikuje ďalšie spracovanie, je, že fonéma sa môže značne meniť v závislosti od vysloveného kontextu. Jeho akustická realizácia závisí od predchádzajúceho a nasledujúceho zvuku, tempa a intonácie. Táto závislosť je známa ako koartikulácia[2].

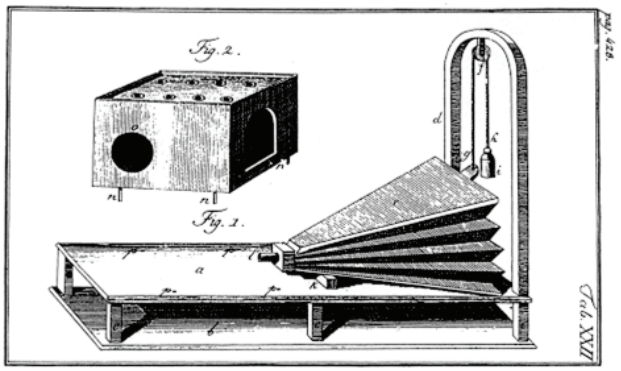
## 2.2 História syntézy reči

Pokusy napodobniť ľudskú reč sú známe dávno, dokonca i zo staroveku. Na ostrove Lesbos zvuky ľudského hlasu vychádzali z úst Orfeovho orákula. V roku 999 francúzsky mních Gelbert - neskôr pápež známy pod menom Silvester II. - vyrobil bronzovú hlavu, ktorá „hovorila“, a „hovoriacu hlavu“ zostrojil údajne aj scholastický filozof a prírodovedec Albertus Magnus. Išlo tu zrejme vždy len o mystifikácie. Zaujímave bolo 18. storočia, známe najmä osobitnou záľubou v mechanických hračkách a automatoch, niekedy až prekvapujúcej dokonalosti.

V 18. storočí sa však objavili aj seriózne vedecké pokusy napodobniť ľudskú reč pomocou prístrojov, spojené so štúdiom reálnych rečových procesov, a to v osobe W. von Kempelena (Obr. 10), ale aj Ch. G. Kratzensteina, profesora fyziológie na univerzite v Halle a v Kodani. Išlo tu už skutočne o vážne vedecké pokusy, a nie o dokonalé mystifikácie ako v ostatných podobných prípadoch, čoho dokladom sú spisy, ktoré títo autori zanechali. Ch. G. Kratzenstein zostrojil už v r. 1773 prístroj, ktorý mechanicky tvoril vokály. Jeho spis vyšiel v r. 1780 pod názvom *Christiani Theophili Kratzensteinii. Tentamen resolvendi problema ab Academia scientiarum imperiali petropolitana ad annum 1780 publice propositum: 1. Qualis sit natura et character sonorum vocalium a, e, i, o, u, tam insigniter inter se diversorum; 2. Annon construi queant instrumenta ordine tuborum organicorum, sub termine vocis humanae noto, similia, quae litterarum vocalium sonos exprimant, in publico Academiae conventu, die XIX septebris 1780, praemio coronotum, Petropoli,* (fr. preklad *Sur la naissance de la formation des voyelles.* Journal de Physique, 21, 1782, s. 358–380). Najmä však rozsiahly spis Mechanizmus ľudskej reči W. von Kempelena predstavuje v dejinách fonetiky ojedinelé dielo, ktoré malo významný vplyv na ďalší vývoj vedy o zvukovej stránke jazyka, a ktoré si preto zaslúži preto čo najväčšiu publicitu. Slovenský preklad tohto spisu, vybavený pomerne bohatým poznámkovým aparátom, vyšiel v r. 1990 vo vydavateľstve Tatran (preložili Slávo Ondrejovič a Peter Ďurčo).

S prácou Wolfganga von Kempelena na hovoriacom stroji, umožňujúcom produkciu hlások a slabík, a teda i slov a krátkych viet azda vo všetkých európskych jazykoch, súvisí spis, ktorý vyšiel vo Viedni v r. 1791 v nemeckej a francúzskej mutácii u vydavateľa J. V. Degena pod názvom *Mechanismus der menschlichen Sprache nebst Beschreibung liner sprechende Maschine - Le Mechanizme de la parole, suivi de la description d'une machine parlante.* Táto práca dokazuje, že v prípade Kempelenovho hovoriaceho stroja nešlo o ni-

jakú mystifikáciu, ani šikovnú mechanickú hračku. Kniha Mechanizmus ľudskej reči v každom prípade potvrdila mimoriadne schopnosti Kempelena prenikavo uvažovať i v abstraktných filozofických polohách a spájať ich s presným technicko-konštruktérskym myslením [4].




***Obr. 10.*** *Kempelenov hovoriaci stroj*

Pokrok v oblasti výkonu výpočtovej techniky a ďalších technológií, ktoré boli vyvinuté v posledných rokoch, umožnili strojom rozprávať, čítať alebo viesť dialóg. To všetko je možné vďaka digitálnemu spracovaniu signálov, teda aj reči ako akustického signálu.

Jednou z prvých praktických aplikácií syntézy boli hovoriace hodiny v roku 1936, ktoré uviedla U.K. Telephone Company. V tomto systéme boli uložené frázy, slová a časti slov, ktoré boli vhodne pospájané, aby vytvorili celé vety. V tom istom období Homer Dudley z Bell Laboratories vytvoril mechanický stroj, ktorý sa ovládal pohybom pedálov a kláves. Ak človek vedel tento prístroj ovládať, dokázal vytvoriť zvuky, ktoré zneli skoro ako reč a boli zrozumiteľné.

Ľuďom sa podarilo rozložiť reč na viacero zložiek a vznikli formantové syntetizátory a ďalšie druhy syntetizátorov. Veľkým problémom v 80. rokoch bola vysoká cena hardvéru, takže venovať sa tejto oblasti mohli iba väčšie laboratória a spoločnosti. V polovici osemdesiatych rokov však cena začala klesať a stále viac univerzít a laboratórií sa začalo oblasti syntézi reči venovať a tento stav trvá až dodnes.

V dnešnej dobe viacero spoločností ponúka ako syntézu, tak aj rozpoznávacie technológie. Najčastejšie sa so syntézou reči môžeme stretnúť počas telefonických hovorov. Každý z nás sa určite stretol s automatom, ktorý mu povedal napr. aktuálny čas, počasie alebo telefónne číslo iného účastníka. Tieto formy sú základné spôsoby syntézy, ktoré sú veľmi ľahko aplikovateľné, keďže používajú obmedzenú množinu slov.

Podobnou aplikáciou sú aj oznamy v dopravných prostriedkoch, kde sú vopred nahovorené nie jednotlivé hlásky alebo slová, ale dokonca celé slovné spojenia alebo vety.

## 2.3 Syntéza reči

Pod syntézou reči rozumieme vytváranie umelého hlasového prejavu pomocou prístroja, ktorý sa nazýva rečový syntetizátor. Tento môže byť v dnešnej dobe vo forme hardvéru, alebo softvéru. „Text-To-Speech“ (TTS) systém konvertuje text na reč, iné systémy prevádzajú symbolickú reprezentáciu jazyka do reči.

Rôzne typy syntézy reči sú analyzované v ďalšej časti publikácie. V súčasnosti najpoužívanejšia forma syntézy je taká, kde syntetizovaná reč môže byť vytvorená zreťazením krátkych rečových záznamov, ktoré sú uložené v databáze. Systémy sa odlišujú vo veľkosti uložených rečových jednotiek. Systémy, ktoré obsahujú fonémy a difóny, poskytujú veľké možnosti na výstupe, ale môže chýbať jednoznačnosť pri výbere vhodného kandidáta z databázy. Niekedy môžu syntetizátory obsahovať aj model vokálneho traktu a iné charakteristiky ľudského hlasu, aby čo najvernejšie vytvorili syntetický hlasový výstup.

Kvalita rečového syntetizátora je posudzovaná na základe podobnosti s ľudským hlasom a úrovňou zrozumiteľnosti. Pre kvalitné syntetizátory je potrebné sledovať prirodzenosť a prozódiu syntetizovanej reči.

Poznáme viacero druhov syntetizátorov, ale pri každom ide o rovnaký cieľ: čo najvernejšie a najzrozumiteľnejšie reprodukovali požadovaný text. Medzi základne prístupy syntézy patrí:

- syntéza spájaním jednotiek
- formantová syntéza
- artikulačná syntéza
- HMM syntéza

### 2.3.1 Syntéza spájaním jednotiek

Syntéza spájaním jednotiek je založená na spájaní segmentov zaznamenanej reči. Vo všeobecnosti výsledok tejto metódy je najviac prirodzene znejúca syntetizovaná reč. Avšak rozdiel medzi prirodzenými obmenami v reči a prirodzenosťou automatizovanej techniky na segmentáciu zvukových segmentov niekedy spôsobí počuteľný rozdiel vo výstupe.

Poznáme tri typy tejto metódy:

- **syntéza založená na jednotkovom výbere** (unit selection) – požíva veľkú databázu nahranej reči. Počas vytvárania databázy je každý záznam segmentovaný do jednej alebo do všetkých z nasledujúcich skupín: fonémy, slabiky, morfémy, slová, alebo vety. Delenie do segmentov sa robí použitím špeciálneho rečového rozpoznávača, ktorý nahrávky zarovnáva. Potom sa môžu ešte ručne upraviť akustickou kontrolou alebo použitím vizuálnej reprezentácie nahrávok, ako sú priebeh alebo spektogramy. Index jednotky v rečovej databáze je vytvorený na zá-

klade segmentácie a akustických parametrov ako základná hlasivková frekvencia, dĺžka trvania, pozícia v slabike, susedné fonémy. Táto metóda poskytuje vysokú prirodzenosť, pretože digitálne spracovanie signálov (DSP) iba málo ovplyvňuje rečové nahrávky. DSP často spôsobí, že rečové nahrávky znejú menej prirodzene. Napriek tomu niektoré systémy používajú DSP na vyhladenie spojových bodov. Na dosiahnutie maximálne prirodzeného výstupu sa vyžaduje veľmi veľká rečová databáza (môžeme hovoriť až o gigabytoch zaznamenaných dát) [87].

- **difónová syntéza** – používa minimálnu databázu obsahujúcu všetky difóny vyskytujúce sa v jazyku. Počet difónov závisí od fonetickosti daného jazyka, napríklad španielčina má asi 800 difónov, nemčina asi 2500. V databáze je vždy len jedna vzorka daného difónu, aj keď možností jeho nahovorení je viac. Prozódia výslednej reči je daná superpozíciou minimálnych jednotiek pomocou digitálneho spracovania signálov ako je lineárna predikcia, PSOLA a MBROLA. Kvalita výslednej reči je vo všeobecnosti horšia ako pri prvej spomenutej metóde, avšak lepšia ako pri formantovej metóde [87], [99].
- **syntéza na špecifickej oblasti** (domain - specific) – zlučuje nahraté slová a frázy a vytvára kompletnú reč. Používa sa v aplikáciách, kde rečový výstup je limitovaný na určitú oblasť, ako dopravné hlásenia alebo predpoveď počasia. Úroveň prirodzenosti týchto systémov je vysoká, pretože rozmanitosť viet je limitovaná a zhoda s originálnymi nahrávkami je veľká [87].

### 2.3.2 Formantová syntéza

Formantová syntéza nepoužíva vzorky ľudskej reči. Namiesto toho je rečový výstup vytvorený zo zvukových modelov. Zvukové modely reči obsahujú informácie o formantoch, znelosti a šume v jednotlivých fonémach. Syntetizátor generuje/upravuje formanty, ich pomery, množstvo šumu a pod. v čase, čím vytvára syntetický rečový výstup. Systémy založené na tejto metóde majú kompletnú kontrolu nad všetkými aspektmi výstupnej reči, širokú obmenu prozódie a intonácie, nemajú však obmenu emócií a tónov [87].

### 2.3.3 Artikulačná syntéza

Artikulačná syntéza popisuje techniku pre syntézu reči založenú na modeli ľudského hlasového traktu a artikulácie. Tento typ syntetizátora sa donedávna používal hlavne na akademické účely [87].

### 2.3.4 HMM syntéza

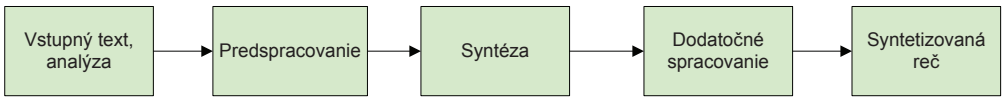
HMM syntéza je založená na metóde skrytých Markovových modelov (hidden Markov models). V tomto systéme sú frekvenčné spektrum, základná hlasivková frekvencia a prozódia reči modelované súbežne pomocou HMM. Priebeh reči je generovaný zo samotných HMM založených na kritériu maximálnej pravdepodobnosti [87], [100].

## 2.4 Základná schéma syntetizátora

V tejto časti sa budeme stručne venovať základným blokom syntetizátora, a preto sme zaviedli pojem všeobecný syntetizátor. Na obrázku nižšie (Obr. 11) je zobrazená bloková schéma všeobecného syntetizátora. Táto schéma je samozrejme zjednodušená a funkcie ako napríklad spätná väzba u niektorých samoučiacich sa systémov, nie sú zobrazené.

Na nasledujúcich 5 základných častí sa dá rozdeliť každý syntetizátor. Sú to:

- Vstupný text, analýza
- Predspracovanie
- Syntéza
- Dodatočné spracovanie
- Syntetizovaná reč




***Obr. 11.*** Bloková schéma všeobecného syntetizátora

V bloku číslo 1 je text určený na syntetizovanie a analýzu. Vždy na začiatku syntézy musí používateľ zadať text, ktorý má byť vyslovený. Následne je tento text načítaný a začína jeho prvotné spracovanie. Táto časť má za úlohu identifikovať v danom texte slová a základné vyhovorenia, resp. časti textu, ktoré sú potom ďalej posunuté do bloku číslo 2. Vyhovorenia sú závislé od toho, aký typ syntetizátora používame [3].

Blok číslo 2 zahŕňa predspracovanie vstupu. Toto predspracovanie môže znamenať viacero techník, minimálne však fonetickú transkripciu. Tá prevádza vstupný text, ktorý je zrozumiteľný človeku do formy použiteľnej pre spracovanie syntetizátorom. V tomto kontexte spomenieme projekt SAMPA (viac v časti 5.4.1) Zjednodušene povedané, ide o mapovanie vstupného textu do počítačovo zrozumiteľného formátu, ktorý opisuje, ako sa má daný segment reči, v našom prípade fonéma, vyslovovať. Ide teda o druh vizualizá-

cie reči. Takto upravený vstupný text je jedným zo vstupov do samotného procesu syntézy (blok č. 3).

V bloku syntézy vyberáme na základe lingvistickej analýzy časti z rečovej databázy. Tá sa skladá z nahovorení slov/textu rečníkom, takže ide o zvukový záznam, ktorý je označený určitými parametrami. Takto vybrané časti databázy sú následne spojené do jedného zvukového záznamu, ktorý predstavuje výstup zo syntetizátora.

Výstup môže, ale nemusí, byť upravovaný dodatočným spracovávaním (blok č. 4). Najčastejšie sa v bloku číslo 4 uskutočňujú procesy na zlepšenie niektorých faktorov vplývajúcich na výslednú kvalitu syntetizovanej reči, t. j. na prirodzenosť a/alebo zrozumiteľnosť. V tomto kontexte možno prirodzenosť definovať ako vlastnosť syntetizovanej reči v zmysle jej vnímania ako reči reálneho človeka a zrozumiteľnosť ako schopnosť extrakcie informácie zo syntetizovanej reči. Pomer týchto dvoch vlastností sa obvykle mení vzhľadom na typ použitého syntetizátora. Dobrým príkladom pre techniky dodatočného spracovania je formantový filter, ktorý má za úlohu zlepšiť prirodzenosť syntetizovanej reči, a je vhodný pre použitie so syntetizátormi, ktoré síce syntetizujú zrozumiteľnú reč, ale neprirodzenú, t. j. umelo znejúcu reč.

Posledný blok už neobsahuje žiadne ďalšie spracovanie. V tejto časti sa zosyntetizovaná reč prehrá v závislosti od požiadaviek použitej aplikácie. V tomto bloku sa môže vykonať transformácia syntetizovanej reči do požadovaného audio formátu. Okrem zmeny formátu je možné podľa potreby zmeniť napr. vzorkovaciu frekvenciu alebo dynamiku (počet bitov na vzorku) výsledného rečového signálu.

### 2.4.1 SAMPA abeceda

SAMPA (Speech Assessment Methods Phonetic Alphabet) je počítačovo-čitateľný fonetický skript, ktorý používa 7 bitové ASCII znaky a je založený na projekte IPA – International Phonetic Alphabet. Jej autorom nie je iba jeden človek, ale vznikla po konzultácii výskumníkov reči v mnohých krajinách.

SAMPU vytvorila medzinárodná skupina fonetikov v rokoch 1987-89 pod projektom EPRIT 1541. Prvý krát bola použitá pre európske jazyky ako dánčina, holandčina, francúzština, nemčina a taliančina (1989), neskôr pre nórčinu a švédčinu (1992), nasledovala gréčtina, portugalčina a španielčina (1993). Pod projektom BABEL sa ju podarilo rozšíriť o bulharčinu, estónčinu, maďarčinu, poľštinu a rumunčinu. Existuje snaha, aby sa postupne rozšírila na všetky jazyky. Iniciatívou OrienTel projektu sa podarilo pridať aj jazyky ako arabčina, hebrejčina a turečtina. Posledné pridané jazyky boli kantončina, chorvátčina, čeština, ruština, slovinčina a thajčina. V najbližšej dobe sa plánuje pridať aj japončina a kórejčina [84].

| graféma | Ábel Kráľ | SAMPA Ivanecký | SAMPA SAV | SAMPA KTL | slovo |
|---|---|---|---|---|---|
| a | a | a | a | a | kapitola |
| e | e | E | e | E | meno |
| i | i | I | i | I | pivo |
| o | o | O | o | O | noha |
| u | u | U | u | U | bubon |
| ä | ä | { | { | { | päta |
| á | á | a: | a: | a~ | pohár |
| é | é | E: | e: | E~ | gén |
| í | í | I: | i: | I~ | víťaz |
| ó | ó | O: | o: | O~ | katalóg |
| ú | ú | U: | u: | U~ | múr |
| ia | i^a | I_^a | i_^a | I_^a | piatok |
| ie | i^e | I_^e | i_^e | I_^e | mier |
| iu | i^u | I_^u | i_^u | I_^u | paniu |
| ô | u^o | U_^O | u_^o | U_^O | kôň |

***Tab. 1.*** *Porovnanie slovenskej transkripcie samohlások podľa abecedy Kráľa, Ivaneckého, SAV a KTL*

X-SAMPA rozširuje základ pre každý symbol IPA, zahrňujúc celú diakritiku. To znamená, že dokáže vytvoriť strojom čitateľnú fonetickú transkripciu pre každý ľudský jazyk. Klávesnicovo-kompatibilné kódovanie pre celý set IPA symbolov zahŕňa všetko z IPA, teda aj diakritiku a tónové značky. Najpoužívanejšie symboly sú mapované do samostatných kláves ASCII rozsahu 33 až 126.

Fonetickou transkripciou v slovenskom jazyku sa začal ako prvý zaoberať Ábel Kráľ. Vychádzal z IPA abecedy a chýbajúce znaky nahradil svojimi vlastnými znakmi, ktoré boli vhodné pre strojový prepis. SAMPA abecedu pre slovenský jazyk ako prvý vytvorili Ivanecký a Nábělková [85], [86]. Ďalej sa v ich práci pokračovalo na Slovenskej akadémii vied (SAV). Bohužiaľ, SAMPA abeceda pre slovenčinu stále nie je štandardizovaná, takže sa stále nájdu odchýlky medzi jednotlivými pracoviskami. Porovnanie rôznych verzií pre slovenskú abecedu je v nasledujúcich tabuľkách (Tab. 1, Tab. 2).

| graféma | Ábel Kráľ | SAMPA Ivanecký | SAMPA SAV | SAMPA KTL | slovo |
|---|---|---|---|---|---|
| p | p | p | p | p | popol |
| b | b | b | b | b | žaba |
| t | t | t | t | t | vata |
| ť | ť | c | c | c | Maťo |

| graféma | Ábel Kráľ | SAMPA Ivanecký | SAMPA SAV | SAMPA KTL | slovo |
|---|---|---|---|---|---|
| d | d | d | d | d | voda |
| ď | ď | J\ | J\ | J- | háďa |
| k | k | k | k | k | páka |
| g | g | g | g | g | guma |
| c | c | ts | ts | ts | cena |
| dz | ʒ | dz | dz | dz | medza |
| č | č | tS | tS | t(S) | oči |
| dž | ǯ | dZ | dZ | d(Z) | džungla |
| f | f | f | f | f | figa |
| v | v | v | v | v | slovo |
| v | w | w | w | w | vdova |
| v | u^ | U_^ | u_^ | U_^ | kov |
| s | s | s | s | s | osa |
| z | z | z | z | z | zima |
| š | š | S | S | (S) | šek |
| ž | ž | Z | Z | (Z) | veža |
| ch | x | x | x | x | chata |
| h | h | h\ | h | h- | Praha |
| ch | y | G | G | G | vrch hory |
| j | j | j | j | j | jama |
| j | i^ | I_^ | i_^ | I_^ | kraj |
| r | r | r | r | r | para |
| r | r^ | r= | r= | r= | vrch |
| ŕ | ŕ^ | r=: | r=: | r=~ | vŕba |
| l | l | l | l | l | skala |
| l | l^ | l= | l= | l= | vlk |
| Ĺ | Ĺ^ | l=: | l=: | l=~ | vĺča |
| ľ | ľ | L | L | (L) | ľad |
| m | m | M | m | m | mama |
| m | ɱ | F | F | F | amfiteáter |
| n | n | n | n | n | rana |
| n | n^ | - | N\ | - | Slovensko |
| n | n | N | N | N | banka |
| ň | ň | J | J | (J) | vaňa |

***Tab. 2.*** *Porovnanie slovenskej transkripcie spoluhlások podľa abecedy Kráľa, Ivaneckého, SAV a KTL*

## 2.5 Najnovšie metódy v oblasti syntézy reči

Na syntetizovanú reč sa vyvinuli dve základné požiadavky, a to zrozumiteľnosť a prirodzenosť. Zatiaľ čo zrozumiteľnosť bola zvládnutá už asi pred štvrťstoročím pomocou formantových rezonátorov, prirodzenosť sa začala zlepšovať iba nedávno, vďaka syntetizátorom vytvárajúcim reč spájaním rečových jednotiek vyseknutých z pôvodných rečových nahrávok. V prípade, že tieto nahrávky neobsahujú ohraničenú množinu jednotiek bez kontextovej príslušnosti, ale tvorí ju veľké množstvo jednotiek z rôznych kontextov, hovorí sa o syntéze reči výberom jednotiek z rečovej databázy – korpusu [26], [27].

V súčasnosti dosahuje syntetizovaná reč spájania jednotiek korpusu dobré výsledky aj v rámci projektov na Fakulte elektrotechniky a informatiky STU. Pri syntéze reči výberom jednotiek z databázy je dôležité, aby obsahovala čo najviac akusticky rôznorodých jednotiek. Neplatí však, že čím väčšia databáza, tým lepší syntetizátor. Zo vzťahu medzi rečou a rozdelením veľkého množstva zriedkavých javov (LNRE - Large Number of Rare Events) vidíme, že niektoré jednotky sú zastúpené frekventovane, iné sa vyskytujú zriedka [29].

Vytvorenie korpusu, ktorý by pokrýval všetky vyhovorenia, nemusí viesť k úspechu. Napríklad zvýšenie počtu jednotiek z 50 000, ktoré pokrývali 75% možného japonského textu, na 80 000 zvýšilo kvalitu iba o 5% [28]. Jedným z možných prístupov ku korpusovej syntéze ľubovoľného vstupu je akceptovanie nižšej kvality reči menej frekventovaných jednotiek, keďže tie sa pri dobre vytvorenom korpuse málo vyskytujú ako v korpuse tak aj v syntetickej reči [30]. Algoritmus vyberajúci jednotky z databázy hodnotí ich vhodnosť pre syntézu na základe určitých parametrov, ako sú napríklad fonetická príslušnosť, doba trvania, intenzita a základná hlasivková frekvencia. Ďalšie parametre určujú kontext jednotky ako fonetickú príslušnosť susedov, pozíciu vo fráze, smerovanie úrovne základnej hlasivkovej frekvencie a iné. Ak je to možné, tak parametre sa udávajú v normalizovanej podobe, napr. ako odchýlka od priemernej hodnoty.

Veľmi známy proces výberu jednotiek z databázy je CHATR. Prvýkrát bol zadefinovaný v prácach [33], [34], v súčasnosti sa používa v mnohých syntetizátoroch (napr. [35]). Výsledná postupnosť jednotiek je s najmenším skreslením.

Pri výbere jednotiek reči z korpusu je možné použiť aj HMM. Postupuje sa tak, že k jednotlivým typom jednotiek reči sú priradené HMM so stavmi nesúcimi pravdepodobnosti pozorovaní – výstupov, ktoré sú v tomto prípade parametre jednotlivých rámcov jednotiek. To znamená, že každá jednotka v databáze je reprezentovaná postupnosťou pozorovaní – parametrov. Pravdepodobnosť pozorovania je možné pre model daného typu jednotky a známu postupnosť stavov (zistenú zo vstupného textu) vypočítať. To znamená, že výber jednotiek pomocou HMM je daný hodnotou pravdepodobnosti danej jednotky v modeli. Samozrejme, že tento „výber“ sa v skutočnosti uskutoční v čase prípravy databázy. Výsledkom tohto výberu môže byť jedna jednotka s najvyššou pravdepodobnosťou

SIS (Single Instance System) [37] alebo niekoľko najpravdepodobnejších jednotiek MIS (Multiple Instance System) [36]. Výhoda MIS je v tom, že výber jednotiek môže zohľadňovať ešte aj konkatenačné skreslenie alebo úpravu prozódie. Príkladom použitia HMM syntezý je aj článok [100].

Za zmienku ešte stojí pohľad na výber jednotiek cez informačno-teoretické kritériá, ktorý bol predostretý v práci [38]. V nich je výber jednotiek reprezentovaný ako prenosový kanál, do ktorého vstupuje požadovaný signál daný vstupnou špecifikáciou, a z ktorého vystupuje prenesený – skreslený signál – v tomto prípade zosyntetizovaný.

Syntéza, kde na vstupe sú priamo parametre požadovaného syntetického signálu a úlohou syntetizátora je čo najvernejšie sa priblížiť cieľu, sa nazýva syntéza nižšej úrovne (low-level) alebo DSP modul syntézy. Vstup do DSP modulu tvoria parametre, ktoré sú vygenerované v module vyššej úrovne NLP (Natural Language Processing). Pri TTS syntéze sa parametre vytvárajú priamo z textu. Takéto rozdelenie procesu syntézy bolo zadefinované vo viacerých prácach, napr. [31], [32].

## 2.6 Skvalitnenie syntézy a princíp učenia v procese syntézy reči

Pod požiadavkou prirodzenosti syntetickej reči sa chápe jej nerozlíšiteľnosť od reči ľudskej. Reč je vnímaná ako prirodzená, keď v kontexte, v akom sa nachádza, nie je možné určiť, či jej zdrojom bol človek alebo počítač. Práve na dosiahnutie prirodzenosti sa v posledných rokoch začala upriamovať najväčšia pozornosť v oblasti výskumu syntézy reči.

Jedným z príkladov je on-line implementácia inteligentného e-learningového systému pre modernú hovorenú čínštinu [69]. Je založená na princípe prirodzeného spracovania reči. Skladá sa zo štyroch klasických komponentov prispôsobených špecifikám dnešnej čínštiny: analýza textu, odhad prozódie, výber rečových jednotiek a samotný generátor reči.

Vzhľadom na nejednoznačnosti nachádzajúce sa v reči boli v trénovacej fáze na základe textov z webových stránok vygenerované n-gramové jazykové modely. Tieto modely zakladajú zanalyzované časti korpusovej databázy a pomáhajú efektívne odhadnúť pravdepodobnosť výskytu kategórie polyfónu. Tieto modely sú kombinované s tzv. fázovou schémou na vylepšenie presnosti. Použitý prístup podľa autorov dosahuje 95% presnosť.

Parametre prozódie, ako intonačné prestávky alebo hlasitosť, sú extrahované z reálnych nahrávok. Veta je rozdelená do častí na základe spracovania segmentácie slov. Niektoré slová však môžu byť spojené a predstavujú väčšiu lingvistickú jednotku – autori ju nazývajú prozodická sekcia. Medzi prozodické sekcie býva vložená intonačná pauza. Modul prozódie odhaduje aj dĺžku, hlasitosť a tempo. Výber rečových jednotiek nastáva po analýze textu. Autori majú nahraté mužské a ženské vzorky reči. Tieto vzorky boli pri

nahrávaní rozdelené na malé segmenty, na ktoré bola aplikovaná pre-enfáza za účelom zníženia SNR. Generátor reči využíva informácie z predradených modulov na syntézu čínskej reči a/alebo lexikologických informácií.

Jednotlivé modely systému používajú korpusovú databázu polyfónov, kde slovník obsahuje 80 000 záznamov (slovník obsahuje aj atribúty ako frekvencia a časť reči). Segmentácia slov je založená na slovníkovej metóde. Model využíva aj rečovú databázu obsahujúcu približne 4 000 čínskych znakov s piatimi tónmi, ktoré predstavujú jednotku rečovej syntézy. Nahratá reč prejde aj procesom normalizácie energie za účelom navodenia prirodzenosti. Model bol testovaný dvanástimi študentmi na náhodne vybratých vetách bežne používaných v čínskom jazyku, pričom sa brali do úvahy parametre energia reči a kvalita reči.

Model dosiahol parametre MOS (na stupnici od 1 po 10, 10 je najlepšie) 7,51, resp. 8,09 pri použití normalizácie energie. Na základe týchto výsledkov autori zhodnotili, že inteligentný systém dokáže poskytnúť syntetizovanú reč s prirodzeným tempom a kvalitou ako aj lexikologické informácie pre výučbu čínskeho jazyka.

Článok Eugenia Oancena a Adriana Badulescu [68] sa zaoberá určením slabiky s prízvukom pre rumunský jazyk v aplikáciách syntézy reči. Na kvalitu syntetizovanej reči vplýva aj určenie prízvuku každého slova vo vete. V prípade nesprávne určeného prízvuku (v horšom prípade jeho ignorovanie) dochádza k zhoršenej kvalite syntézy reči. Rumunčina nie je jazykom s pevne určeným prízvukom, preto je ťažké vytvoriť sústavu pravidiel, ktoré môžu určiť polohu prízvuku. Oancen navrhuje algoritmus, ktorý určí prízvuk v slove podľa počtu parametrov slova zahrňujúci morfológiu, fonetiku a lexikálny charakter slova. Metóda na určenie prízvuku v slove sa skladá z troch častí: príprava textu a spracovanie slova, klasifikácia slova, analýza tried a formulovanie pravidiel pre prízvuk.

V prvej fáze sa zhromaždili všetky slová z rumunského jazyka okrem prebratých z cudzích jazykov. Slovník tak predstavuje základný súhrn slov rumunského jazyka. Každé slovo zo slovníka bolo analyzované a boli extrahované parametre, pomocou ktorých sa vytvorili kritériá na zadelenie slov do tried. Kritériá boli nasledovné:

- určenie dĺžky slova *n*, ktorá je zhodná s počtom písmen v slove
- rozdelenie slov na slabiky podľa pravidiel platných pre rumunský jazyk
- určenie počtu slabík *r*
- *u* – sekvencia, ako nasledujú za sebou spoluhlásky a samohlásky v analyzovanom slove
- určenie prízvuku pre každé slovo

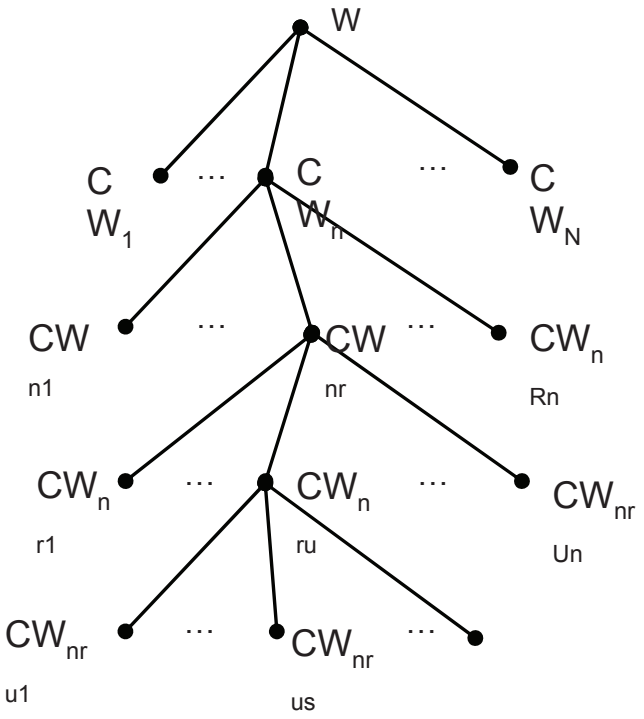
Počas druhej fázy boli slová slovníka *W* rozdelené do tried podľa rôznych kritérií, ako napríklad:

$CW_n$ – slová s dĺžkou *n*, $n = 1,2.N$, kde *N* je maximálna dĺžka slova.

$CW_{n,r}$ - trieda s dĺžkou slov n a počtom slabík $r$, $r = 1,2.R_n$ ($R_n$ je maximálny počet slabík pre dĺžku slova *n*). Podobným spôsobom sa vytvoria ďalšie podtriedy s kritériami ako fonetický kód slova, začiatočná sekvencia slov.

Platí, že $W$ je nadmnožinou $CW_n$ a tá je nadmnožinou $CW_{n,r}$ atď. až po poslednú triedu $CW_{n,r\ u\ sm}$, kde sú slová dĺžky $n$, s počtom slabík $r$, fonetickým kódom $u$, ktorý obsahuje $sm$ sekvenciu písmen. Analýza tried slov tak poskytne informácie o prízvuku v slabike.

V poslednej fáze, ktorá sa venuje analýze tried a vytvoreniu pravidiel pre prízvuk, sú triedy slov zoskupené do štruktúry Obr. 12. Vytvárajú tak hierarchickú stromovú štruktúru. Na najvyššej úrovni je úplná trieda $CW_n$, kým na najnižšej úrovni je trieda $CW_{n,r}$ $u_{sm}$ s najväčším počtom kritérií.




***Obr. 12.*** *Stromová štruktúra tried*

Analýza slov začala v najvyššej triede, kde sa analyzovali slová dĺžky $n$. Iba na základe dĺžky sa ale žiadne pravidlo na prízvuk nedalo vytvoriť. Ďalej sa analyzovala ďalšia trieda slov s dĺžkou $n$ a počtom slabík $r$. Tu bolo vytvorené prvé pravidlo **R1**: Slová dĺžky $n$ a počtom slabík $r$ musia mať prízvuk na pozícii $p$. Postupnou analýzou všetkých tried sa vytvorili aj ďalšie pravidlá na určenie prízvuku. **R2**: Slová dĺžky $n$, počet slabík $r$ a fonetický kód $u$ musia mať prízvuk na pozícii $p$. V závislosti od sekvencie písmen bolo pravidlo **R3** rozdelené na:

- **R31**: Slová dĺžky $n$, počet slabík $r$, fonetický kód $u$ a začiatočná sekvencia písmen $si$ musia mať prízvuk na pozícii $p$.
- **R32**: Slová dĺžky $n$, počet slabík $r$, fonetický kód $u$ a konečná sekvencia písmen $se$

musia mať prízvuk na pozícii *p*.

- **R33**: Slová dĺžky *n*, počet slabík *r*, fonetický kód *u* a vnútorná sekvencia písmen *sm* musia mať prízvuk na pozícii *p*.

Na základe týchto úvah a analýz bol vytvorený algoritmus, ktorý určí prízvuk pre každé slovo v rumunskom jazyku. Ako bolo spomenuté, slová sú klasifikované podľa začiatočnej, konečnej a vnútornej sekvencie. Tieto pravidlá sú príznačné, ale nie nezávislé, určenie prízvuku jedným pravidlom môže byť ovplyvnené výsledkom ďalších dvoch pravidiel. Akékoľvek pravidlo určené na začiatku sa stáva definitívnym až po zhodnotení výsledkov ďalších krokov v analýze.

Vstupom do programu na určenie prízvuku je ľubovoľný text v rumunskom jazyku. Všetky potrebné dáta na určenie prízvuku, ktoré sa získali analýzou slovníka, sú súčasťou databázy. Databáza je tvorená tabuľkami (Kategória, Kódovanie, Sekvencia) s určenou štruktúrou podľa vyššie spomenutého algoritmu. Na základe morfologického, fonetického a lexikálneho významu slova program vyberie z tabuľky pravidlo určitého typu. Ak má byť použité pravidlo **R1**, prízvuk je určený podľa tabuľky Kategória, ak má byť použité pravidlo **R2**, prízvuk je určený podľa tabuľky Kódovanie, ak ma byť použité pravidlo **R3**, prízvuk je určený podľa tabuľky Sekvencia.

V prvom teste boli testované všetky slová zo slovníka *W* (4 500 slov) a v druhom teste slová *W'*, ktoré sa predtým v slovníku nenachádzali. Globálna chyba bola 6%. Ako ukázali testy so slovníkom *W'*, zvyšovanie počtu slov v slovníku *W* za účelom pokrytia čo najväčšej množiny môže mierne zvýšiť chybovosť. Pridanie nového kritéria na určenie prízvuku môže, naopak, chybovosť znížiť. Nevýhodou je, že pravidlá nemôžu byť použité pre tie isté slová (rovnako napísané) s iným významom, kde význam je určený práve prízvukom.

## 2.7 Prozódia reči

Rečové signály vytvorené pomocou syntézy založenej na princípe spájania akustických jednotiek môžu byť kódované pomocou rozličných rečových modelov. Hlavnou požiadavkou, ktorá sa kladie na tieto modely, je zabezpečenie plynulého prechodu medzi spojenými akustickými jednotkami. Nespojitosť prozódie vo formantových frekvenciách a vo fáze zapríčiňujú neprirodzenosť syntetizovanej reči.

Prozódia reči sa zaoberá vlastnosťami reči, ktorých doménou nie je jednoduchý fonetický segment, ale väčšie jednotky skladajúce sa z viacerých segmentov, napríklad celé vety. Preto sa prozodické javy nazývajú supra-segmentálne. Tieto javy vnímame ako dôraz, akcent alebo rôzne modifikácie intonácie, rytmu a hlasitosti.

Prozodické javy môžeme rozdeliť do štyroch principiálnych úrovní: úmysel, artikulácia, akustická realizácia a vnímanie.

- **lingvistická úroveň** - v akomkoľvek jazyku môže rečník uplatniť prozodické kódovanie, ako aj iné zložky reči s určitým úmyslom. Tento úmysel môže ovplyvniť lingvistické aj paralingvistické vyjadrenie. Pod lingvistickým vyjadrením rozumieme rôzne ústne vyjadrenia s použitím rečových znamení. Paralingvistické zahŕňa neverbálny prejav ako onomatopoja a určité citové a emocionálne vyjadrenie, napríklad hnev, radosť, naliehavosť alebo iróniu. Prozódia hrá dôležitú úlohu v oboch typoch javov. Prozódia vo všeobecnosti znamená vzájomné spájanie rozličných lingvistických zložiek, predovšetkým zdôrazňovaním určitých zložiek textu pomocou označenia hraníc a definovaním prechodov medzi slovami, frázami alebo vetami. Prozodické javy sa z lingvistického hľadiska zvyčajne delia na tónové, intonačné alebo prízvukové (akcentové).
- **artikulačná úroveň** - na artikulačnej úrovni sa prozodické javy fyzicky prejavujú ako rad modifikácií artikulačných pohybov, ktoré sa dajú merať pomocou komplikovaných prístrojov (magnetografia, ultrazvuk, röntgen, atď.). Prozodické javy vnímame ako osobitnú vrstvu pridanú na normálnu artikulačnú vrstvu a prejavujú sa ako systematické modifikácie základných neutrálnych artikulačných vlastností. Fyzikálne merania prozodických prejavov obyčajne zahŕňajú zmeny amplitúd artikulačných pohybov, zmeny v tlaku vzduchu, alebo špecifické zobrazenia elektrických impulzov v nervoch.
- **akustická úroveň** - aktivita svalov v dýchacom systéme a pozdĺž vokálneho traktu spôsobuje vydávanie zvukových vĺn. Akustická realizácia prozodických javov môže byť meraná a vyčíslená pomocou akustickej analýzy signálu. Hlavnými akustickými parametrami vzťahujúcimi sa k prozódii sú fundamentálna frekvencia, intenzita a trvanie. Napríklad zdôraznené slabiky majú zvyčajne vyššiu fundamentálnu frekvenciu, väčšiu amplitúdu a trvajú dlhšie, ako nezdôraznené slabiky.
- **vnemová úroveň** - poslucháč uchom prijíma zvukové vlny a pomocou spracovania týchto vnemov si odvodzuje lingvistické informácie a paralingvistické informácie z prozodických javov. V tomto bode môžu psycholingvistické testy poskytnúť informácie o reakciách poslucháča na prozodické javy. Tento druh testov môže overovať významnosť rôznych prozodických značiek v reči, takisto ako významnosť akustických oddeľovačov potrebných na vyvolanie minimálnych vnemových rozdielov medzi rozličnými zámermi rečníka. Na vnemovej úrovni sa prozodické javy zvyčajne klasifikujú vo výrazoch poslucháčovej subjektívnej skúsenosti, ako pauzy, dĺžka, melódia a hlasitosť.

Prozódia má veľký význam pri syntéze reči, najmä z hľadiska zrozumiteľnosti a prirodzenosti zosyntetizovanej reči. Pri syntéze musíme dbať na charakteristické vlastnosti reči ako fundamentálna frekvencia, intenzita, dĺžka trvania, aby sa zabezpečili plynulé prechody medzi jednotlivými segmentmi reči. Prozódia výslednej reči je daná superpozíciou minimálnych jednotiek pomocou digitálneho spracovania signálov, ako je lineárna predikcia, OLA, PSOLA, MBROLA alebo sínusoidálne modely [2], [40].

Najjednoduchšia metóda, ktorá modifikuje signál v čase, je metóda súčtu s prekryvom OLA (Overlap-Add). Ak chceme signál roztiahnuť v čase, tak sa rámce prekrývajú v nových časových lokáciach. Vylepšený algoritmus PSOLA (Pitch-Synchronous Overlap-Add) [5] vykonáva tónovo-synchrónnu analýzu a syntézu rečového signálu pomocou priamej manipulácie so signálom. Počas analýzy sa nastavia tónové značky znelej časti signálu na tónovo-synchrónnu rýchlosť a neznelej časti sa nastavia na konštantnú rýchlosť. Analyzačné okná sú centrované okolo týchto značiek. Okná sa rozprestierajú cez dve periódy a v čase sa prekrývajú. Výška tónu sa počas syntézy mení pomocou zmeny časových odstupov medzi susednými oknami a ich sčítavaním. Trvanie sa mení pridávaním alebo odoberaním okien, čiže jedno okno je najmenšia jednotka modifikácie trvania [2], [40].

Algoritmus TD PSOLA (Time-Domain Pitch-Synchronous Overlap-Add) bol prvý raz publikovaný pánmi Moulinesom a Charpentierom v roku 1990. Nekladie vysoké nároky na výpočtový výkon hardvéru a prináša výrazný pokrok v oblasti syntézy reči. Po určitom čase sa však objavili niektoré nedostatky algoritmu. Preto v roku 1992 páni Dutoit a Leich prichádzajú s vylepšeným algoritmom MBR TD PSOLA (Multi-band Resynthesis Time Domain Pitch Synchronous Overlap Add), skrátene MBROLA. Podstatou syntézy je použitie algoritmu PSOLA na predspracovanú databázu. Túto databázu vytvoríme pomocou Multiband Excitation (MBE) resyntézy nahovorenej databázy. Jednotlivé časti nahovoreného textu sú označované automaticky generovanými značkami, ktoré môžu byť umiestnené na ľubovoľnom mieste v rámci periódy. Použitím fixnej hodnoty rozstupu medzi značkami sa môžeme vyhnúť nezhodám výšky tónu už v štádiu syntézy. Použitím fixného vzťahu medzi fázami dvoch databázových segmentov dosiahneme odstránenie fázových nezhôd. Používajú sa na to dve techniky: jednotlivé periódy sú transformované tak, aby začínali a končili s nulovou fázou, alebo s fixnou náhodnou fázou. Pri nastavení nulovej fázy však dosiahneme výrazne „plechovú“ syntetizovanú reč, preto sa odporúča používanie druhej metódy. Transformáciu je možné spraviť iba so znelými segmentmi – t.j. segmentmi, ktoré majú kvázi periodický charakter. Syntéza je v podstate zhodná so syntézou v prípade algoritmu PSOLA. Navyše robíme len priamu časovú interpoláciu spektrálnych obálok, aby sme zabránili ich nezhodám. Výsledkom algoritmu MBROLA by mala byť relatívne kvalitná syntetizovaná reč. Veľkou výhodou je, že syntéza so sebou neprináša výrazne zvýšené nároky na výpočtový výkon hardvéru pri spracovaní databázy ani následnej syntéze. Jeho najväčšou nevýhodou zatiaľ zostáva jemne „plechová“ reč, ktorá môže byť mierne zašumená [2], [40].

Veľmi silným a bohatým nástrojom na prozodické úpravy reči sú sínusoidálne modely. Sú to parametrické modely, pomocou ktorých je možné ľahko meniť veličiny, priamo súvisiace s prozodickými vlastnosťami, ako je zmena fundamentálnej frekvencie (čiže intonácie), tempa reči a jej intenzity (čiže prízvuku). Sínusoidálne modely využívame najmä na koncové úpravy, ktoré zabezpečia hladké napájanie rečových jednotiek a úpravu prozódie [87].

## 2.8 Úprava hlasu a emócií

V súčasnosti je dôležitým smerom vytvorenie variability ľudskej reči a zachytenie emócii človeka. Nárast požiadaviek na zrozumiteľnosť a prirodzenosť syntetickej reči spôsobili, že výskum sa začal zameriavať na prozódiu a napodobňovanie rozmanitostí prirodzeného hlasu a hlasu s emóciami.

Prvé systémy a experimenty v syntéze reči s emóciami boli založené na formantových syntetizátoroch [7]. Ďalšie systémy využívali na syntézu hlasu s emóciami syntézu založenú na spájaní akustických jednotiek. V tomto type syntézy sa využívali dve metódy. Prvá metóda bola založená na spájaní, z hľadiska emócií, neutrálnych akustických jednotiek [8]. Emócie sa ďalej upravovali pomocou automatického modelovania prozódie reči. Pri druhej metóde sa vytvorila databáza obsahujúca emočné vyhovorenia reči [9]. Z tejto databázy sa potom priamo syntetizovala reč s emóciami (ako šťastie, hnev, prekvapenie, smútok a neutrálne emócie) bez použitia akéhokoľvek explicitného matematického modelu. Sínusoidálne modely sú na modifikovanie reči veľmi vhodné a dajú sa pomocou nich meniť prozodické prvky, ako výška hlasu, tempo a dôraz, ktoré hrajú v reči s emóciami podstatnú rolu. Preto je prirodzené, že sa môžu využívať aj pri zmene emócií rečníka. Príklad takýchto modifikácií je popísaný v práci [10].

Techniky modifikácie hlasu sa pokúšajú transformovať rečové signály vyslovené jedným rečníkom tak, aby zmenili charakter hlasu, alebo aby vyslovená reč znela ako keby bola vyslovená druhým rečníkom. Tieto techniky sa nazývajú konverzia hlasu. Pretože vzťahy na nájdenie identifikačných vlastností rečníka ešte stále nie sú dostatočne prebádané, je bežné špecifikovať požadované modifikácie charakteristických veličín vo vzťahu k existujúcemu rečníkovi. Technológia modifikácie hlasu má mnoho aplikácií vo všetkých systémoch, ktoré využívajú pred nahratú reč, ako hlasové schránky alebo komplikovanejšie syntetizátory textu na reč založené na akustickom spájaní jednotiek. V takých prípadoch by mohla byť modifikácia hlasu jednoduchým a efektívnym spôsobom na vytvorenie požadovanej rozmanitosti hlasov, aby sme sa zároveň vyhli nahrávaniu rozdielnych rečníkov. K rečníkovej individualite prispievajú najmä vlastnosti ako rýchlosť hovorenia, poloha hlasu alebo trvanie odmĺk. V mnohých prípadoch sa tiež ukázalo, že špecifické charakteristiky vnímaného hlasu sú ovplyvnené lingvistickým štýlom reči. Všetky tieto okolnosti, najmä fakt, že na prozodické vlastnosti má veľký vplyv aj zmysel povedanej správy aj úmysel rečníka, znemožňujú spracovanie týchto vlastností reči automatickým systémom. Ukázalo sa, že aj priemerné hodnoty týchto vlastností (priemerná fundamentálna frekvencia, celková dynamika reči) nesú veľkú časť špecifických informácií o rečníkovi. Takisto existujú výskumy, že odlišní rečníci môžu byť rozlíšení porovnaním ich spektrálnych obálok. Z týchto dôvodov sa v praxi zatiaľ práce zameriavajú najmä na konverziu charakteristík spektrálnych obálok na segmentálnej úrovni.

Jednou z najskorších metód spektrálnej konverzie je metóda mapovania kódovou

knihou [11], ktorá je založená na aplikovaní vektorovej kvantizácie na spektrálne parametre zdrojového a cieľového rečníka. Ďalšou metódou je modifikovanie len špecifických aspektov spektrálnej obálky, ako napríklad umiestnenie formantov [12]. Na modifikovanie hlasu rečníka na báze spektrálnej konverzie bol použitý aj sínusoidálny systém publikovaný v [13].

Navrhnutá metóda je založená na využití modelu Gaussových zmesí (GMM), ktorý popisuje spektrálne obálky zdrojového rečníka. Konverzia sama je reprezentovaná spojitou parametrickou funkciou, ktorá berie do úvahy pravdepodobnostnú klasifikáciu danú GMM modelom. Metóda konverzie je implementovaná v prostredí HNM (Harmonic plus Noise) modelu, ktorý umožňuje vysokokvalitné modifikácie rečových signálov. Táto metóda odhaduje konverzné charakteristiky s využitím výrokov zdrojového a cieľového rečníka, ktoré boli časovo zarovnané aplikovaním algoritmu DTW (dynamickej časovej deformácie). Oblasť zdrojového rečníka je popísaná spojitou pravdepodobnostnou hustotou, ktorá zodpovedá parametrickému GMM modelu. Tým sa zvýši robustnosť konverzie. Parametre konverznej funkcie sa určujú minimalizáciou celkového kvadratického spektrálneho skreslenia medzi konvertovanými obálkami a cieľovými obálkami. V poslednom kroku, ktorý sa nazýva prírastkové učenie, sa vylepšuje časové zarovnanie pomocou aplikácie DTW algoritmu medzi konvertovanými a cieľovými obálkami, pretože sa prišlo na to, že značná časť reziduálnej chyby pochádza z lokálnych chýb v časovom zarovnaní. Spektrálna obálka sa určuje z parametrov HNM modelu pomocou výpočtu kepstrálnych koeficientov s využitím Barkovej frekvenčnej mierky. Táto metóda modifikácie hlasu rečníka sa ukázala robustnejšia a efektívnejšia ako metódy založené na vektorovej kvantizácii.

Popísane zlepšenie je dôsledkom kombinácie využitia spojitého pravdepodobnostného modelu spektrálnych obálok a HNM modelu pre reč, ktorý umožňuje vysokokvalitné modifikácie rečového signálu.

## 2.9 Možnosti modifikácie ľudskej reči

Základné možnosti úpravy reči spočívajú v úprave základnej hlasivkovej frekvencie, tempa reči a amplitúdy, teda úprave výšky hlasu vypovedania, dĺžok trvania jednotlivých foném a úrovne hlasitosti vypovedania. Týmto upravíme akustickú zložku reči. Ak však chceme zmeniť farbu hlasu rečníka, takýmto spôsobom sa to nepodarí. Vypovedanie bude znieť ako originál, ale spôsob vypovedania bude odlišný. Tento spôsob je vďaka svojej jednoduchosti vhodný na úpravu reči pre prenos pomocou syntetizátora, kde môže byť modifikácia implementovaná priamo v syntetizátore.

Pokročilejším postupom modifikácie je úprava sínusoíd daného vyhovorenia. Potom následnou spätnou syntézou je možné získať akékoľvek vyhovorenie. Miera podobnosti upraveného a originálneho vyhovorenia závisí od počtu sínusoíd použitých pri analýze,

správnom určení hraníc foném a samotnej podobnosti vyhovorení pred modifikáciou. Podrobnejšie rozobrané spôsoby modifikácie sú rozobraté v kapitolách 5.6 a 5.7.

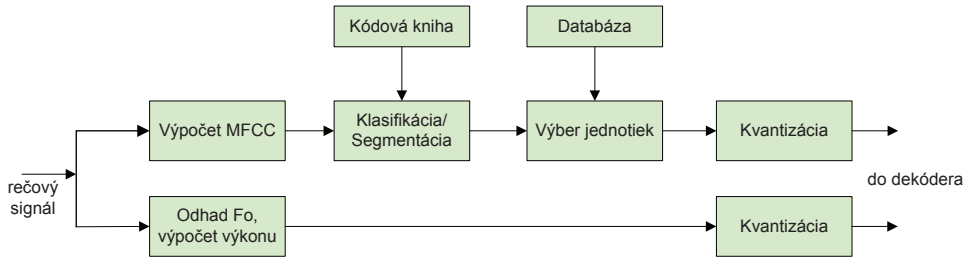
## 2.10 Kódovanie reči pomocou TTS

V súčasnej dobe existuje mnoho algoritmov (najznámejšie boli popísané v predošlých kapitolách) umožňujúcich zmenšenie množstva prenášaných dát. Väčšina z nich však spôsobuje straty na kvalite zvukového záznamu. Riešenie s využitím syntetizátora prináša nový spôsob prenosu reči s minimálnymi nárokmi na prenosné pásmo. Aj keď tento spôsob radíme medzi stratové algoritmy, strata nie je veľká a závisí od miery kompresie. V dnešnej dobe sa tejto téme venuje veľká pozornosť.

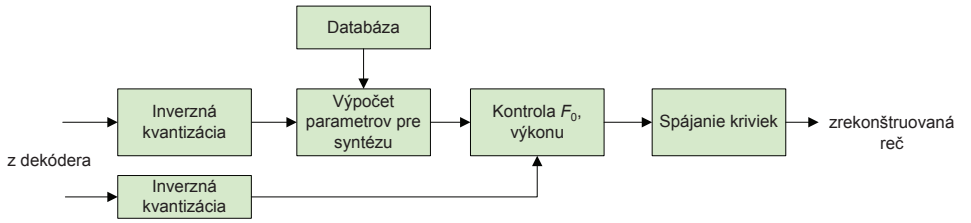
Riešeniu, kde sa na kódovanie reči využije TTS systém, sa venuje práca pána Coxa a Ki-Seung Leea [39]. Ich nízko-rýchlostný kodér (Obr. 13) je založený na princípe rozpoznávania a syntézy reči. Rečový signál je vytvorený pomocou syntézy spájaním jednotiek. Preto hlavné časti (štruktúra databázy, výber jednotiek, úprava prozódie) TTS systému, pracujúcom na princípe spájania jednotiek, sú využité na kódovanie reči. Rečové jednotky sa automaticky vyberajú z databázy na základe segmentačno-klasifikačnej schémy.

Pri výbere sa uplatňuje dynamické programovanie (dynamic programming) s dvoma základnými hodnotiacimi funkciami: cena akustického výsledku a cena spájania. Cieľom je vybrať jednotky čo najviac podobné vstupnému textu. Cieľ je okrem textu definovaný aj pomocou charakteristických vlastností reči, takže syntetizovaná reč sa viac podobá originálu. Rozdiely v prozódii medzi vybratými jednotkami a vstupným originálom sú minimalizované časovým posunom a modifikáciou základnej hlasivkovej frekvencie HNM metódou spomínanou v kapitole 2.8.

Pri prenose sa kóduje zvlášť text a zvlášť základná hlasivková frekvencia a obálka intenzity. Pri kódovaní $F_0$ sa berú do úvahy bity potrebné na zakódovanie aproximačnej chybovosti a bity požadované na zakódovanie $F_0$. Keďže sa nemôžu minimalizovať obe zložky, tak kritérium sa zvolilo také, že sa minimalizuje počet bitov potrebných na zakódovanie hlasivkovej frekvencie, pokým maximálna hodnota aproximačnej chybovosti je pod určitou hranicou. Na zakódovanie sa využíva dynamické programovanie. Najskôr sa nájde lokálne optimálna dráha každej $F_0$, ktorá má najmenší počet bitov a vyhovujúcu hodnotu aproximačnej chyby. Potom sa nájde globálna optimálna cesta. Taktiež verné napodobnenie intenzity reči prispeje k celkovému lepšiemu výsledku. Keďže kodér aj dekodér (Obr. 14) používajú tú istú databázu, ktorá obsahuje informácie o intenzite, stačí prenášať informácie o indexoch, z ktorých sa určí výsledná intenzita na strane prímača. V testoch pre jednotlivcov systém dosiahol veľmi dobre výsledky: pri rýchlosti 580 bit/s bola syntetizovaná reč prirodzená a zrozumiteľná.




***Obr. 13.*** *Bloková schéma kodéra*




***Obr. 14.*** *Bloková schéma dekodéra*

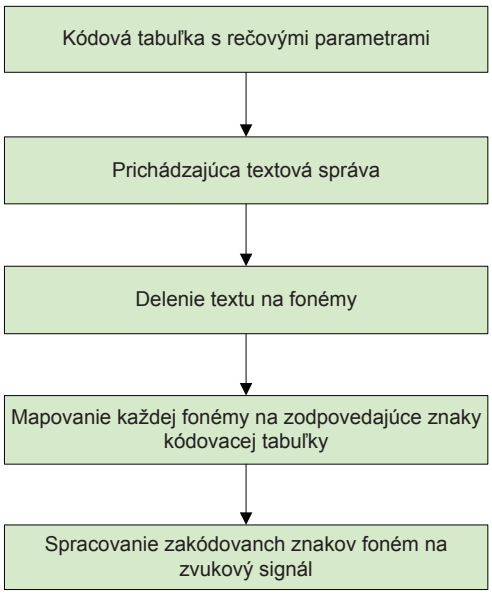
Ďalšia práca [41] pojednáva o vylepšenom TTS systéme, ktorý je menej náročný na spracovanie signálu pri syntéze reči využívajúc výhody digitálnych procesorov (DSP) a kódovacích algoritmov v dnešných mobilných telefónoch. Systém konvertuje textovú správu a na výstup pošle audio signál (syntetizovanú reč) využijúc pôvodný kódovací systém mobilného telefónu bez zvýšených nárokov na pamäť a procesor.

Po prijatí textovej správy, určený systém prekonvertuje signál do textového formátu (ASCII alebo Unicode). V prvom kroku sa text konvertuje na kódové znaky komunikačného systému, čím sa ušetria kroky potrebné na vytvorenie reči z textu a na spracovanie rečového signálu vokodérom. Kódové znaky má systém uložené v tabuľke, ktorá je uložená v pamäti, a zvyčajne obsahujú okrem iných hodnôt aj kód budenia lineárneho prediktora (Code Excitation Linear Predictors) a vektor sumy budenia lineárneho prediktora (Vector Sum Excited Linear Predictors). Kvôli vyššej efektivite obsahuje tabuľka aj skomprimované audio dáta reprezentujúce charakteristické parametre reči. Takže audio signál môže byť prenášaný tak, že sa kóduje a dekóduje pomocou tabuľky, čím sa ušetrí prenosové pásmo bez straty kvality rečového signálu.

V ďalšom kroku sa textová správa delí v audio serveri na akustické jednotky. Audio server sa nachádza v mikroprocesore mobilného telefónu. Presnejšie povedané, v audio serveri sa nachádzajú pravidlá prispôsobené na štruktúru a fonémy daného jazyka. Vety sa tu rozpoznaním medzier rozdelia na slová a tie potom na fonémy. Samozrejme, že vety môžu obsahovať aj iné znaky než písmená, napríklad čísla, skratky alebo iné špeciálne znaky jazyka. Preto sa pred samotným delením správy na slová a fonémy musia všetky tieto znaky a čísla prepísať slovne. Pri prepise textu sa niektoré ASCII znaky môžu využiť na označenie akcentu v slove. Tým sa dosiahne vyššia prirodzenosť syntetizovanej reči.

Po akustickej analýze sa vykoná mapovanie foném k príslušným znakom z kódovacej tabuľky spomenutej vyššie. Každá fonéma je mapovaná ku korešpondujúcemu digitálnemu zvuku, ktorý je v komprimovanej forme vlastnej danému mobilnému telefónu. Napríklad v GSM komunikačných systémoch sa používa formát zodpovedajúci polovičnej rýchlosti formátu vokodéra. V poslednom kroku sú všetky kódové parametre spracované v signálovom procesore (DSP) a odtiaľ rečový signál postupuje do audio obvodov telefónu, ktoré obsahujú prevodník. Keďže jednotlivé fonémy sú už zakódované v kódových parametroch (kódová tabuľka), procesor nerobí už žiadne modifikácie výsledného signálu.

Aby sa využili všetky výhody DSP, kódovací systém pre syntetizovanú reč by mal byť vlastný štandardu mobilného telefónu. Hlavným dôvodom je, že DSP a jeho software sú navrhnuté na dekompresiu určitého kódového formátu v existujúcom vokodéri. Takéto použitie TTS pomôže ušetriť dáta potrebné na prenos audio signálu, keď sa prenesie iba čistý text.




***Obr. 15.*** *Postup na TTS syntézu*

# 3 Využitie sínusoidálnych modelov na modifikáciu reči

Prvou aplikáciou sínusoidálnych modelov bolo kódovanie rečových signálov, neskôr kódovanie všeobecných audiosignálov. Hlavná výhoda spočíva v tom, že sínusoidálny model má schopnosť odstrániť nepodstatné dáta a kódovať signály s nižšou bitovou rýchlosťou. Výhody aplikácie využívajúcej tieto modely nie sú zanedbateľné, pretože v dnešnej dobe sa kladie dôraz na znižovanie šírky prenosového pásma bez zníženia kvality signálu.

Keďže sínusoidálny model je parametrickým modelom, je relatívne jednoduché pomocou neho meniť vlastnosti analyzovaného signálu. Z týchto dôvodov sa zvykne používať aj na zmenu dĺžky trvania, prípadne frekvenčnej výšky hudobných signálov. V posledných rokoch vznikla tendencia využívať sínusoidálne modely nielen pri kódovaní, ale aj pri syntéze reči pomocou spájania jednotiek z korpusu.

Sínusoidálne modely majú široké možnosti použitia od kódovania reči až ku kódovaniu všeobecných audiosignálov, keďže sínusoidálny model má schopnosť odstrániť zo signálu nepodstatné dáta a kódovať ho podstatne nižšou bitovou rýchlosťou. Tým znižuje šírku pásma potrebného na prenos bez zníženia kvality. Sínusoidálny model je možné použiť pre modifikáciu prozodických vlastností syntetizovanej reči.

V posledných rokoch sa využívajú sínusoidálne modely nielen pre kódovanie reči, ale aj pri syntéze reči spájaním jednotiek korpusu, z ktorej sa potom pri syntéze vyberajú jednotlivé fonémy v parametrickom tvare. Najčastejšie sa používajú na modifikáciu niektorých častí reči alebo na zlepšenie zrozumiteľnosti a prirodzenosti syntetizovanej reči. Priamo zo sínusoidálnych parametrov sa môže vytvoriť databáza, z ktorej sa v parametrickom tvare vyberú jednotlivé fonémy. Relatívne novou oblasťou, ktorou sa začali sínusoidálne modely v súvislosti s rečou zaoberať, je prozodická modifikácia reči, kedy sa modifikuje rytmus, intonácia a prízvuk reči. V súčasnosti sú sínusoidálne modely používané ako vedúce prístupy vo viacerých smeroch spracovania signálov [42].

Prvé pokusy o syntézu audiosignálov v tejto oblasti boli založené na sumácii sínusoidálnych zložiek meniacich sa v čase. Táto metóda sa nazýva aditívna syntéza. Aditívna syntéza umožňuje meniť základnú hlasivkovú frekvenciu a tempo reči nezávisle od seba. Je však vhodná len pre harmonické časti reči, nakoľko nerobí žiadne rozdiely medzi harmonickými a neharmonickými zložkami signálu a na reprezentáciu neharmonických zložiek potrebuje veľký počet sínusoíd. Preto je na analýzu vhodnejší štandardný sínusoidálny model plus šum (SN model).

## 3.1 Štandardný model sínusoíd plus šum

Na syntézu audiosignálov sa v zásade používa všeobecný audiosignál, ktorý môže byť modelovaný ako súčet deterministickej a stochastickej časti, alebo inak povedané ako suma množiny sínusoíd plus šumové rezíduum [2]. Sínusoidálne zložky sú zvyčajne harmonické a šumové rezíduum obsahuje energie nespôsobené periodickou vibráciou. Mohli by byť modelované sínusoidami rovnako ako periodické časti signálu, bolo by však potrebné veľké množstvo zložiek, a preto sa na modelovanie šumu zvyčajne používajú iné postupy. Najčastejšie je rezíduum reprezentované pomocou krátkodobých energií s pevnými frekvenčnými pásmami alebo pomocou filtrovaného bieleho šumu.

V tomto štandardnom SN modeli je deterministická zložka signálu $x(t)$ reprezentovaná ako suma sínusoidálnych trajektórií s parametrami meniacimi sa v čase. Trajektória je sínusoidálna zložka s frekvenciami, amplitúdami a fázami meniacimi sa v čase, ktorá v časovo-frekvenčnom spektrograme vyzerá ako krivka.

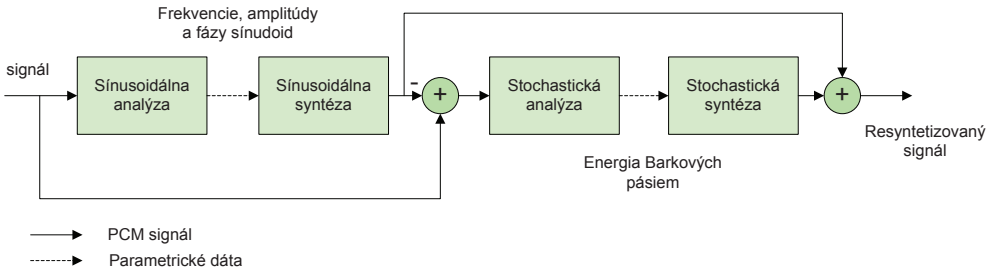
$$x(t)=\sum_{i=1}^{N} a_i(t)\cos\left(\Theta_i(t)\right)+r(t) \tag{1}$$

kde $a_i$ a $\Theta_i$ je amplitúda a fáza sínusoidy $i$ v čase $t$ a $r(t)$ je šumové rezíduum, ktoré môže reprezentovať stochastický model. Predpokladá sa, že amplitúdy sínusoíd nie sú vystavené rýchlym zmenám a fázy sú lokálne lineárne. Rezíduum $r(t)$ obsahuje všetky zložky signálu $x(t)$, ktoré nie sú modelované pomocou sínusoíd, teda aj tie sínusoidy, ktoré neboli detegované.

Ak predpokladáme, že rezíduum obsahuje iba stochastické zložky, môžeme ho reprezentovať pomocou filtrovaného bieleho šumu. Okamžitá fáza a amplitúda sa nezapamätajú, ale modelujú sa pomocou časovo sa meniaceho filtra obálky alebo pomocou krátkodobých energií s pevnými frekvenčnými pásmami (Barkove pásma). Sínusoidálny model sa preto považuje za model vychádzajúci z fyzikálnych vlastností. Okrem toho vychádza aj z vlastností fyziologických, pretože berie do úvahy fakt, že ľudské ucho nie je citlivé na detaily zvukovej obálky, ani na fázu neperiodických signálov.

Bloková schéma systému sínusoíd plus šum je znázornená na obrázku (Obr. 16). Najprv pomocou analýzy získame časovo sa meniace amplitúdy, fázy a frekvencie. Potom nasleduje syntéza sínusoíd, zosyntetizovaný signál sa odpočíta od originálneho signálu a dostaneme tak šumové rezíduum. Na toto rezíduum sa aplikuje stochastická analýza. Aby sme dostali celý resyntetizovaný signál, musí sa syntetizovať stochastický signál, ktorý sa musí pripočítať k syntetizovaným sínusoidám.

Syntéza je najzložitejšia časť celého systému. Vstupný signál je rozdelený na čiastočne sa prekrývajúce rámce. Potom sa pomocou krátkodobej Fourierovej transformácie (SFT) získa krátkodobé spektrum rámca, ktoré sa analyzuje a získavajú sa tak parametre sínusoíd: frekvencia, amplitúda a fáza.




***Obr. 16.*** *Základná štruktúra systému na analýzu/syntézu*

Stochastickú časť signálu dostaneme v časovej oblasti odpočítaním syntetizovaných sínusoíd z originálneho signálu. Toto rezíduum reprezentuje filtrovaný biely šum. Ako už bolo spomenuté, ľudské rozlišovanie zvukov nemôže rozlíšiť zmenu energie vnútri pevných frekvenčných pásiem nazývaných Barkove pásma, pre kvázišumové, stacionárne signály tak presná spektrálna obálka nie je potrebná. Jediná informácia potrebná pre kvázišumové signály sú krátkodobé energie vnútri Barkových pásiem. Pri syntéze sa generuje komplexné spektrum a v ňom náhodné fázy pre amplitúdy, ktoré sme dostali z energií Barkových pásiem. Susedné rámce sa kombinujú použitím aditívnej syntézy s prekrývaním. Ďalšou možnosťou modelovania šumu je modelovanie pomocou autoregresívnej metódy, kde sa z rezídua počíta autokovariancia (alebo autokorelácia v závislosti od použitej metódy). Na ich základe sa vypočítajú predikčné koeficienty, ktoré umožňujú výpočet obálky spektra rezídua a zosilnenie. Šum sa potom syntetizuje pomocou filtrácie bieleho Gaussovského šumu cez celopólový filter [2].

## 3.2 Analýza a odhad parametrov signálu pomocou sínusoidálneho modelu

Ľudské vnímanie audiosignálov sa dá reprezentovať ako séria úrovní: od nízkej až po vysokú. Reprezentácie na nízkej úrovni zodpovedajú signálom pred vnútorným uchom. Reprezentácie na vysokej úrovni sú také, pri ktorých rozpoznávame niekoľko rôznych zvukov naraz (napr. hrá rádio a v pozadí počuť hukot sanitky). Medzi týmito dvoma úrovňami existuje reprezentácia strednej úrovne [58], medzi ktoré patrí aj SN model.

Poznatky o reprezentácii strednej úrovne ľudského sluchového systému sú trochu obmedzené. Ellis a Rosenthal [58] zostavili nasledujúci zoznam vlastností pre reprezentáciu strednej úrovne:

- **Separácia zdroja zvuku**. Prirodzené zvuky sa prekrývajú s inými a náš sluch má schopnosť zaradiť zvuky k ich zdrojom.
- **Invertibilita**. Z parametrickej reprezentácie môžeme znovu vytvoriť originálny signál.

- **Redukcia zložiek**. Originálny signál, ktorý dosiahne ušný bubienok, možno považovať za súbor úrovní tlaku vzduchu. Pri reprezentácii by sa počet objektov mohol zmenšiť vzhľadom na dôležitosť klesajúcu s úrovňou tlaku.
- **Abstraktná dôležitosť charakteristických vlastností**. Vlastnosti, ktoré reprezentácia využíva, by mali viac korešpondovať s fyzikálnymi charakteristikami ako s algoritmickými detailmi.
- **Fyziologická vierohodnosť** z hľadiska fyziológie človeka.

Sínusoidálny model spĺňa druhé, tretie a štvrté kritérium. Model dovoľuje separáciu zdrojov, začiatok sínusoíd zodpovedá začiatku zvuku a frekvencie sínusoíd zodpovedajú rezonančným frekvenciám zdrojov zvuku. Fyziologická vierohodnosť modelu je malá. Model je orientovaný viac fyzikálne a produkuje zjednodušené dáta, čo sa dá považovať za výhodu. Ak je požadovaná informácia vyššej úrovne, dáta možno ľahko analyzovať použitím analýzy vyššej úrovne [2].

Sínusoidálnu analýzu môžeme zhrnúť do štyroch krokov:

- Detekcia špičiek signálu – najdôležitejšia časť, na ktorej sa celá analýza zakladá,
- Interpolácia špičiek – pre lepšie frekvenčné rozlíšenie,
- Odhad parametrov špičiek – amplitúd a fáz,
- Vytvorenie trajektórií z detegovaných špičiek.

Najdôležitejším krokom je detekcia špičiek, pretože na základe zdetegovaných špičiek sa vykonáva syntéza. Detekcia špičiek nie je triviálna a pri odhade ich parametrov existuje niekoľko elementárnych problémov, ktoré sú zväčša spojené s dĺžkou analyzačného okna. Krátke okno je spojité kvôli sledovaniu rýchlych zmien vo vstupnom signáli a dlhé okno je potrebné kvôli správnemu odhadu frekvencií sínusoíd a kvôli rozlíšeniu blízko ležiacich sínusoíd. Ak povolíme rýchle zmeny v amplitúde a frekvencii, bude modelovaná aj stochastická časť signálu (pri použití veľkého množstva sínusoíd). To sa zvyčajne od sínusoidálneho modelu neočakáva, hlavnou úlohou je reprezentovať iba harmonické zložky periodického signálu [2].

Takmer vo všetkých systémoch na analýzu sínusoíd sa odhad špičiek a parametrov vo frekvenčnej oblasti vykonáva použitím DFT: vychádza sa z predpokladu, že každej stabilnej sínusoide zodpovedá impulz vo frekvenčnej oblasti. Pretože prirodzené signály nie sú nikdy stabilné, musíme analyzovať signál v časovej oblasti na niekoľkokrát použitím posuvného okna a krátkodobej Fourierovej transformácie. Na zlepšenie frekvenčného rozlíšenia sa používa aj doplnenie nulami. Špička v magnitúde krátkodobej Fourierovej transformácie indikuje prítomnosť sínusoidy v okolí danej frekvencie. Veľmi rýchla a najjednoduchšia metóda na detekciu sínusoíd je taká, kde sa vyberie fixný počet lokálnych maxím v magnitúde. Táto metóda je výhodná pre audio kódovacie aplikácie, pre potreby analýzy však nie je veľmi praktická, pretože v prípade neharmonických zvukov vyberá metóda aj špičky spôsobené šumom, čo zvyčajne spôsobuje problémy pri ďalšej analýze [2].

Prirodzeným vylepšením metódy je použitie prahu na detekciu špičiek. Všetky lokálne maximá magnitúdy krátkodobej Fourierovej transformácie sa interpretujú ako

sínusoidné špičky. Táto metóda produkuje variabilný počet špičiek, no neodstránila problém, že niektoré špičky v magnitúde môžu byť spôsobené šumom alebo inými neharmonickými signálmi. Takisto neberie do úvahy celkovú spektrálnu obálku a amplitúdy harmonických tónov, ktoré v prípade prirodzených zvukov zvyčajne klesajú ako funkcia frekvencie. Následkom toho vyššie harmonické tóny zvyčajne spadnú pod fixný prah a nie sú detegované. Iným častým problémom je nedostatočné frekvenčné rozlíšenie DFT. Na zjemnenie spektra vo frekvenčnej oblasti sa používa metóda doplnenia nulami v časovej oblasti. To zabezpečí lepšie frekvenčné rozlíšenie a presnejší odhad parametrov sínusoidy. Predpokladá sa, že jedna zložka DFT môže obsahovať iba jednu sínusoidálnu zložku [2].

Na odhad parametrov sa často využíva metóda kvadratickej interpolácie. Každý DFT koeficient reprezentuje frekvenčný interval $f_s/N$, kde $f_s$ je vzorkovacia frekvencia a $N$ je dĺžka DFT. Kvôli vysokej kvalite vzorkovacích frekvencií je potrebná DFT s dĺžkou stoviek až tisícov vzoriek. To nie je príliš praktické, a preto bola na získanie presnejších frekvencií sínusoíd potrebná iná metóda. Táto metóda bola prvýkrát použitá v [63] a na získanie presnejších frekvencií využíva kvadratickú interpoláciu.

Na magnitúde spektra $|X(\omega)|$ naoknovaného signálu indikuje lokálne maximum prítomnosť sínusoidy na blízkej frekvencii. Ak je oknová funkcia $h(t)$ symetrická, tak kvadratická funkcia, ktorej stred je umiestnený v $\Omega$, dáva dobrú aproximáciu $|H(\omega-\Omega)|$ okolo bodu $\Omega$ [62]. Parametre sa potom dajú odhadnúť pomocou troch bodov DFT spektra. Často vyžívaným spôsobom kvadratickej interpolácie je metóda opísaná Smithom a Serrom [63]. Všeobecná podoba paraboly je daná ako

$$y(x)=a(x-p)^2+b \tag{2}$$

pre ktorú platí, že $y(-1)=\alpha$, $y(0)=\beta$, $y(1)=\gamma$. Konštanty $\alpha$, $\beta$ a $\gamma$ sa rovnajú hodnotám troch najvyšších DFT vzoriek:

$$\begin{aligned}\alpha&=20\log_{10}|X(\omega_1)|\\ \beta&=20\log_{10}|X(\omega_2)|\\ \gamma&=20\log_{10}|X(\omega_3)|\end{aligned} \tag{3}$$

Smith a Serra empiricky zistili, že frekvencie majú tendenciu byť dvojnásobne presnejšie, keď sa namiesto lineárnej magnitúdy použije magnitúda v dB. Umiestnenie extrému paraboly $p$ dostaneme ako

$$p=\frac{1}{2}\frac{(\alpha-\gamma)}{(\alpha-2\beta+\gamma)} \tag{4}$$

a odhad skutočného umiestnenia špičky $k^*$ (vo vzorkách DFT) bude

$$k^*=k^2+p \tag{5}$$

kde $k^2$ je vzorka DFT, ktorá má maximálnu magnitúdu. Frekvencia v Hz je potom

$$f = \frac{f_s k^*}{N} \tag{6}$$

Iný spôsob detegovania špičiek, nazvaný F-test, vynašiel Thompson [62] v geofyzike a úspešne sa používal aj v audio spracovaní [59], [60]. Táto metóda využíva množinu ortogonálnych okien. Odhad spektra sa vypočítava ako váhovaný priemer niekoľkých dátových okien, aby sa predišlo vyhladeniu a skresleniu. Pretože F-test potrebuje výpočet niekoľkých Fourierových transformácií, je výpočtovo náročnejší ako kroskorelačná metóda. V ideálnych podmienkach je schopný detegovať sínusoidy bez toho, aby detegoval aj špičky šumu. V neideálnych podmienkach, ako napríklad pri sínusoidách príliš blízko seba alebo pri rýchlo sa meniacich amplitúdach alebo frekvenciách, dáva horšie výsledky ako kroskorelačná metóda.

Metóda kroskorelácie [61] sa s úspechom používala aj v kódovaní reči. Vypočítava kroskoreláciu medzi krátkodobým spektrom signálu a spektrom ideálnej sínusoidy. Avšak táto metóda je schopná detegovať veľký počet sínusoíd za rozličných podmienok. V štandardnom prístupe sa získavajú stredné hodnoty parametrov v jednom krátkom časovom rámci. Novším prístupom je snažiť sa odhadnúť priebeh zmeny amplitúdy sínusoidálnej krivky priamo vnútri rámca [61]. Táto zmena amplitúdy sa modeluje polynómom nižšieho stupňa. Ak poznáme strednú frekvenciu, vieme pomocou zložitých maticových operácií odhadnúť aj koeficienty polynómu, ktorý modeluje priebeh amplitúdy. Okrem toho, že použitie takéhoto modelovania lepšie popisuje priebeh signálu, slúži aj ako parameter pri lepšom napájaní špičiek do trajektórií pri použití skrytých Markovových modelov.

Ďalším vylepšením klasickej sínusoidálnej analýzy je algoritmus iteratívnej analýzy zvyškového signálu. Algoritmus je kombinovaný s algoritmom zlučovania parametrov. Tento postup odhadu parametrov má dve výhody: znižuje počet sínusoidálnych zložiek a dáva presnejšie parametre jednotlivých sínusoíd. Nevýhodou je výpočtová náročnosť, pretože vyžaduje niekoľko realizácií tradičnej analýzy [2].

Iteratívna analýza funguje nasledovne. Najprv sa detegujú sínusoidy z originálneho signálu pomocou nejakej jednoduchej detekčnej metódy. Detegované sínusoidy sa syntetizujú a odpočítajú sa od pôvodného signálu v časovej oblasti, aby sme dostali zvyškový signál. Ďalej sa detegujú sínusoidy vo zvyškovom signáli a opäť sa syntetizujú a odpočítajú. Tento postup sa môže opakovať buď fixným počtom iterácií, alebo kým nedostaneme požadovaný počet sínusoíd, prípadne kým signál neobsahuje žiadne zmysluplné harmonické zložky. Hoci tento algoritmus produkuje dobré výsledky, je počet získaných sínusoíd pomerne veľký. Parametre získané počas jednej iterácie zvyčajne nie sú úplne presné, čo má za následok chyby odhadu prejavujúce sa vo zvyškovom signáli. Častými chybami odhadu sínusoidálnych zložiek sú tiež sínusoidy, ktoré sú frekvenčne blízko k originálu, a amplitúdy, ktoré sú menšie ako originálne amplitúdy. V nasledujúcich iteráciách sa tieto chyby odhadu detegujú, takže každú harmonickú zložku pôvodného signálu nakoniec

reprezentujú viaceré sínusoidy. Keďže toto z hľadiska počtu sínusoidálnych zložiek nie je príliš vhodné, algoritmus môže sínusoidy po každej iterácii zlučovať [62].

Ďalšou metódou je tónovo-synchrónna analýza, ktorá sa používa najmä pri monofónnom audio signáli a reči. Vtedy sa dá dĺžka okna synchronizovať s fundamentálnou frekvenciou. Túto výhodu vidieť najlepšie vo frekvenčnej oblasti, kde frekvencie harmonických zložiek presne korešpondujú s frekvenciami DFT koeficientov. Koeficienty sínusoíd dostaneme priamo z komplexného spektra. Táto metóda dovoľuje takú malú dĺžku okien, ako je perióda zvuku. Pre každý nový rámec sa vypočíta odhad základnej hlasivkovej frekvencie a podľa nej sa prispôsobí dĺžka okna (čím je tón nižší, tým je okno dlhšie a naopak). Toto zaručuje, že sa nájdu všetky frekvencie nad fundamentálnou frekvenciou. Metódu nemožno použiť, ak sa prekrýva niekoľko zvukov s rozličnou hlasivkovou frekvenciou.

Problém dobrého frekvenčného aj časového rozlíšenia rieši multirozlišovacie sínusoidálne modelovanie. Na vyriešenie problému dĺžky okna sa používa prístup, keď sa vstupný signál rozdelí do niekoľkých pásmovo ohraničených kanálov. Dĺžka okna je pre každý kanál navrhnutá inak. Tento spôsob umožňuje dobré frekvenčné rozlíšenie v pásme nízkych frekvencií (so slabým časovým rozlíšením) a zároveň dobré časové rozlíšenie v pásme vyšších frekvencií (so slabým frekvenčným rozlíšením). Na určenie parametrov potrebujeme na vstupný signál aplikovať niekoľko posuvných FFT s rozdielnymi dĺžkami okna. Pri malej dĺžke okna sa odhadujú parametre pre vyššie frekvencie, pri veľkej dĺžke okna nižšie frekvencie. Zložitosť tohto algoritmu však lineárne stúpa s počtom oktáv, respektíve s počtom FFT [2].

SN model sa ťažko aplikuje na zvuky s dočasnou zmenou vlastností (prechodovými javmi - transient), ako sú veľmi rýchle zmeny parametrov audio signálu, napr. rýchly nábeh amplitúdy. Na takéto modelovanie je potrebný veľký počet sínusoíd. Tomuto sa dá vyhnúť, ak sa budú prechodové javy detegovať a budú reprezentované napr. pomocou neparametrického transformačného kódovania.

Keď sú sínusoidné špičky a ich parametre odhadnuté, spájajú sa do medzirámcových trajektórií. Pre každý rámec sa algoritmus spájania špičiek snaží pripojiť sínusoidné špičky do už existujúcich trajektórií z predchádzajúceho rámca takým spôsobom, aby výsledkom bola hladká krivka frekvencií a amplitúd. Špičky sa spájajú na základe kriteriálnej funkcie – buď sa porovnávajú iba na základe frekvencií, alebo aj ostatných parametrov. Problémy nastanú v prípade, že sú v signáli prítomné viacnásobné harmonické štruktúry, neharmonické zložky alebo veľké frekvenčné zmeny. Tento problém je riešiteľný pomocou HMM (skrytých Markovových modelov). Optimalizácia sa robí v danom časovom intervale, ktorý zodpovedá štatistickému kritériu pokračovania kriviek pre všetky sínusoidálne parametre. Optimálna množina trajektórií sa určí ako sekvencia stavov s najvyššou pravdepodobnosťou pomocou Viterbiho algoritmu. Používa sa koncept „narodenia“ a „smrti“, čiže ktorá trajektória nemá pokračovanie, tak „umrie“, a ak nejaká špička ostane nenapojená na trajektóriu, tak sa „narodí“ nová trajektória [63].

Sínusoidné trajektórie obsahujú všetky informácie potrebné na rekonštrukciu har-

monických častí signálu: amplitúdy, frekvencie a fázy každej trajektórie v každom rámci. Aby neboli hranice rámcov nespojité, amplitúdy, frekvencie a fázy sa z jedného rámca do ďalšieho interpolujú. Amplitúdy sa interpolujú lineárne, frekvencie a fázy pomocou kubickej interpolačnej funkcie [62].

Iným spôsobom syntézy harmonickej časti signálu je už spomínaná metóda aditívnej syntézy s prekryvom. Aditívna syntéza s prekryvom má menšiu výpočtovú náročnosť ako sumácia sínusoidálnych trajektórií. Na spektrum sa aplikuje požadovaná modifikácia, napríklad prenásobenie prenosovou funkciou filtra, aby sme dostali požadovaný $m$-tý rámec spektra. Potom sa vykoná inverzná Fourierova transformácia a dostaneme naoknovaný výstupný rámec. Výstup je daný súčtom takýchto naoknovaných rámcov s prekryvom. Analýza a resyntéza pomocou sčítania s prekryvom sú identické operácie za predpokladu, že súčet prekrytých analyzačných okien je jednotkový. Konštrukcia okien s touto vlastnosťou je opísaná v literatúre. Navrhlo sa niekoľko rekonštruujúcich okien – pravouhlé, trojuholníkové, Hannovo, Hammingovo. Tieto okná sa nazývajú aj okná s vlastnosťou aditívneho prekryvu. Treba si však uvedomiť, že každá oknová funkcia spĺňa túto podmienku, ak je krok posúvania okna o jednu vzorku. V prípade, že krok posúvania okna je rovný dĺžke okna, jediné okno, ktoré spĺňa podmienku, je pravouhlé okno [2].

Ako už bolo spomenuté vyššie, reziduálny signál dostaneme ako rozdiel originálneho signálu a syntetizovaných sínusoíd v časovej oblasti. V ideálnom prípade obsahuje rezíduum len neharmonické zložky. Jediná potrebná informácia na reprezentáciu rezídua je v čase sa meniaca spektrálna obálka, pretože ľudské vnímanie zvuku nie je citlivé na fázu (platí len pre monofónne signály). Túto spektrálnu obálku dostaneme pomocou parametrov krátkodobých energií vnútri pevných frekvenčných pásiem (najčastejšie Barkove pásma). V tom prípade sa pri syntéze generuje komplexné náhodné spektrum a susedné rámce sa kombinujú pomocou aditívnej syntézy s prekryvom. Pri psychoakustických experimentoch sa zistilo, že ucho nie je citlivé na zmeny energie vo vnútri Barkových pásiem. Platí to pre signály podobné šumu. Za predpokladu, že rezíduum je podobné šumu, je možné ho modelovať pomocou výpočtu krátkodobých energií v každom Barkovom pásme. V pásme medzi 0-20 kHz existuje dvadsaťpäť nelineárne rozložených Barkových pásiem.

Inou možnosťou na získanie spektrálnej obálky je autoregresívne modelovanie, kde sa z rezídua počíta autokovariancia alebo autokorelácia. Z toho sa ďalej určia predikčné koeficienty a zosilnenie. Cez celopólový filter sa filtruje biely Gaussovský šum, a tak dostaneme výsledný šum v signáli.

V porovnaní sínusoidálnej a stochastickej analýzy a syntézy vidieť, že spracovanie stochastickej časti signálu je oveľa jednoduchšie. Pri stochastickej analýze môžeme upraviť iba dĺžku okna a počet rámcov za sekundu.

Po zosyntetizovaní sínusoíd aj stochastickej časti možno obe časti spočítať, čím dostaneme resyntetizovaný signál. V niektorých systémoch sa obe časti syntetizujú vo frekvenčnej oblasti.

## 3.3 Harmonický plus šumový model

HNM – Harmonic plus Noise model vychádza z predpokladu, že rečový signál sa skladá z harmonickej a šumovej časti. Harmonické časti zodpovedajú kvázi-periodickým zložkám reči, šumu zodpovedajú neperiodické zložky. Tieto dva druhy zložiek sú vo frekvencii navzájom delené časovo sa meniacim parametrom, nazývaným maximálna znelá frekvencia $F_m$. Nižšie pásmo spektra (pod $F_m$) je v tomto modeli reprezentované výhradne harmonickými sínusoidami, zatiaľ čo horné pásmo (nad $F_m$) je reprezentované modulovanými šumovými zložkami. Hoci tieto predpoklady z hľadiska produkcie reči nie sú celkom platné, sú užitočné z hľadiska vnímania reči – vedú k jednoduchému modelu reči, ktorý poskytuje vysokokvalitnú syntézu a modifikovanie rečového signálu [66].

Ako bolo spomenuté, HNM predpokladá, že rečové spektrum je rozdelené do dvoch pásiem. Tieto pásma sú oddelené maximálnou znelou frekvenciou, ktorá sa v čase mení. Nižšie pásmo, čiže harmonická časť, je modelované ako suma harmonických zložiek

$$s_h(t) = \sum_{k=-L(t)}^{L(t)} A_k(t)\, e^{jk\omega_0(t)t} \tag{7}$$

kde $L(t)$ znamená počet harmonických zložiek zahrnutých v harmonickej časti, $\omega_0$ znamená fundamentálnu frekvenciu a $A_k(t)$ môže nadobúdať jednu z nasledujúcich podôb:

$$\begin{aligned} A_k(t) &= a_k(t_i) \\ A_k(t) &= a_k(t_i) + t b_k(t_i) \\ A_k(t) &= a_k(t_i) + t c_k(t_i) + t^2 d_k(t_i) \end{aligned} \tag{8}$$

kde $a_k(t_i)$, $b_k(t_i)$, $c_k(t_i)$, $d_k(t_i)$ sú komplexné čísla s $\arg\{a_k(t_i)\} = \arg\{c_k(t_i)\} = \arg\{d_k(t_i)\}$ (predpoklad konštantnej fázy, avšak $b_k(t_i)$ môže mať odlišnú fázu), kde arg znamená fázový uhol komplexného čísla. Tieto parametre sa merajú v čase $t = t_i$, ktorý sa nazýva analyzačný časový okamih. Počet harmonických zložiek $L(t)$ závisí od fundamentálnej frekvencie $\omega_0(t)$ a od maximálnej znelej frekvencie $F_m(t)$. Ak je $|t - t_i|$ malé, tak HNM model predpokladá, že $\omega_0(t) = \omega_0(t_i)$ a $L(t) = L(t_i)$ [66].

Použitím prvého výrazu pre $A_k(t)$ dostaneme jednoduchý stacionárny harmonický model (nazývaný HNM1), zatiaľ čo dva ďalšie výrazy vedú ku komplikovanejším modelom nazývaným HNM2 a HNM3. Posledné dva modely sa snažia modelovať dynamické charakteristiky rečového signálu. Ukázalo sa, že HNM2 a HNM3 sú presnejšie modely pre reč, navyše HNM3 sa správa robustnejšie pri prítomnosti aditívneho šumu. Avšak HNM1 dokáže napriek svojej jednoduchosti produkovať reč, ktorá je pri vnímaní skoro neodlíšiteľná od originálneho rečového signálu. Na druhej strane, vďaka svojej jedno-

duchosti nebude pri HNM1 vyhladzovanie parametrov cez napájacie body akustických jednotiek zložitou úlohou. Zvážením všetkých týchto skutočností sa pre syntézu reči javí ako najvhodnejší model HNM1 (ďalej naňho budeme odkazovať ako na HNM) [66].

HNM model predpokladá, že v hornom pásme znelej reči prevažuje modulovaný šum. V skutočnosti sa vysoké frekvencie znelej reči preukazujú špecifickou časovo-frekvenčnou štruktúrou z hľadiska lokalizácie energie (šumových dávok). Energia tejto vysoko priepustnej informácie sa nerozprestiera cez celú periódu reči a toto sa pri HNM zohľadňuje. Šumová časť je popísaná vo frekvencii pomocou časovo sa meniaceho autoregresívneho (AR) modelu $h(\tau,t)$. Jeho štruktúra v časovej oblasti je predpísaná parametrickou obálkou $e(t)$, ktorá moduluje šumovú zložku. Šumová časť $s_n(t)$ je teda daná:

$$s_n(t)=e(t)[h(\tau,t)*b(t)] \tag{9}$$

kde $*$ je konvolúcia a $b(t)$ je biely Gaussovský šum. Celý syntetizovaný signál dostaneme ako

$$\hat{s}(t)=s_h(t)+s_n(t) \tag{10}$$

Veľmi dôležité je, aby šumová časť bola synchronizovaná s harmonickou časťou. Ak to tak nie je, tak šumová časť nie je vnemovo integrovaná s harmonickou časťou a javí sa ako odlišný zvuk oddelený od harmonickej časti [66].

## 3.4 Analýza reči pomocou HNM modelu

Prvým krokom analýzy pomocou HNM modelu je odhad fundamentálnej frekvencie a odhad maximálnej znelej frekvencie [2]. Tieto dva parametre sa odhadujú každých 10 ms. Dĺžka okna závisí od minimálnej fundamentálnej frekvencie, ktorá je povolená. Najprv sa zistí úvodný odhad fundamentálnej frekvencie pomocou prehľadania minimálnych hodnôt chybovej funkcie cez predšpecifikovanú množinu periód výšky tónov. Chybová funkcia je daná ako

$$E(P)=\frac{\sum_{t=-\infty}^{\infty} s^2(t)w^2(t)-P\sum_{t=-\infty}^{\infty} r(lP)}{[\sum_{t=-\infty}^{\infty} s^2(t)w^2(t)][1-P\sum_{t=-\infty}^{\infty} w^4(t)]} \tag{11}$$

kde $s(t)$ je rečový signál, $w(t)$ je analyzačné okno a $r(k)$ je definované ako

$$r(k)=\sum_{t=-\infty}^{\infty} s(t)w^2(t)s(t+k)w^2(t+k) \tag{12}$$

Úvodný odhad fundamentálnej frekvencie sa používa na určenie znelosti reči v časovej aj vo frekvenčnej oblasti a takisto na ďalšie zjemnenie odhadu fundamentálnej frekvencie. Rozhodovanie o znelosti/neznelosti je založené na kritériu, ktoré berie do úvahy, ako blízko je harmonický model k originálnemu rečovému signálu. Takže s použitím úvodného odhadu fundamentálnej frekvencie sa vygeneruje syntetický signál $\hat{s}(n)$ ako suma harmonických sínusoíd s amplitúdami a fázami odhadnutými priamo z DFT. Ak $\hat{S}(\omega)$ je syntetické spektrum a $S(\omega)$ je originálne spektrum, potom sa robí rozhodnutie o znelosti/neznelosti porovnaním normalizovanej chyby cez prvé štyri harmonické s daným prahom (typicky –15 dB)

$$E = \frac{\int_{0,7\omega_0}^{4,3\omega_0} \left(|S(\omega)| - |\tilde{\hat{S}}(\omega)|\right)^2}{\int_{0,7\omega_0}^{4,3\omega_0} |S(\omega)|^2} \tag{13}$$

kde $\omega_0$ je úvodný odhad fundamentálnej frekvencie. Ak je chyba $E$ pod daným prahom, tak sa rámec označí ako znelý, inak ako neznelý [2], [66].

Pre znelé rámce je odhad maximálnej znelej frekvencie $F_m$ založený na nasledujúcom algoritme vyberania špičiek. Nájde sa najväčšia špička vo frekvenčnom rozsahu $[\omega_0/2; 3\omega_0/2]$. Nech $\omega_c$ označuje umiestnenie frekvencie špičky a nech $A(\omega_c)$ označuje jeho amplitúdu (v dB). Pre lepšie rozlíšenie skutočných a falošných špičiek sa používa druhé meranie amplitúdy, nazývané kumulatívna amplitúda $A_c(\omega_c)$. Táto amplitúda je definovaná ako suma amplitúd všetkých vzoriek od predchádzajúceho minima po nasledujúce minimum špičiek. Do úvahy sa berú aj špičky z frekvenčného rozsahu $[\omega_c - \omega_0/2; \omega_c + 3\omega_0/2]$ a pre každú špičku sa vypočítajú dva typy amplitúd. Nech $\omega_c$ určuje frekvencie týchto špičiek a nech $A(\omega_i)$ a $A_c(\omega_i)$ sú amplitúda a kumulatívna amplitúda v $\omega_i$. Označme strednú hodnotu týchto kumulatívnych amplitúd $\overline{A(\omega_i)}$ a počet najbližších harmonických pri frekvencii $\omega_c$ označme $l$. Na špičku na frekvencii $\omega_c$ sa potom aplikuje harmonický test:

ak $\frac{A_c(\omega_c)}{A_c(\omega_i)} > 2$ alebo $A(\omega_c) - \max\{A(\omega_i)\} > 13$ dB, potom

ak $\frac{|\omega_c - l\,\hat{\omega}_0|}{l\,\hat{\omega}_0} < 10\,\%$ , tak frekvencia $\omega_c$ sa prehlási za znelú, inak na neznelú.

Keď máme určenú frekvenciu $\omega_c$, tak sa prehľadáva interval $[\omega_c - \omega_0/2; \omega_c + 3\omega_0/2]$, hľadá sa najväčšia špička a aplikuje sa ten istý harmonický test. Tento proces pokračuje cez celé rečové pásmo [66].

V mnohých prípadoch nie sú znelé regióny reči čisto separované od neznelých regiónov. Rieši sa to tým, že sa vytvorí vektor binárnych rozhodnutí s pravidlom, že znelé frekvencie sa označujú ako 1 a neznelé ako 0. Filtrovaním tohto vektora trojbodovým

mediánovým vyhladzovacím filtrom sa tieto dva regióny separujú. Najvyššia nenulová hodnota vo filtrovanom vektore potom určuje maximálnu znelú frekvenciu $F_m$ [2], [66].

Na redukovanie chýb modelovania pri reprezentácii znelej reči pomocou HNM je potrebný presný algoritmus odhadovania frekvencie. S použitím úvodného odhadu fundamentálnej frekvencie môžeme dostať zjemnený odhad $\hat{\omega}_0$ ako hodnotu, ktorá minimalizuje chybu

$$E(\hat{\omega}_0) = \sum_{i=1}^{L} \left| \omega_i - i\,\hat{\omega}_0 \right| \tag{14}$$

kde $L$ je počet detegovaných znelých frekvencií $\omega_0$.

Moderné aplikácie na spracovanie reči vyžadujú operácie so signálom, ktorý je kontaminovaný vysokou úrovňou šumu. Tieto metódy často využívajú na odhad fundamentálnej frekvencie štatistické metódy sledovania frekvencie [66].

## 3.5 Syntéza pomocou HNM modelu

Pri syntéze predpokladáme, že výber jednotiek pre syntetizovaný výraz už prebehol. Ďalším predpokladom je, že priebeh základnej hlasivkovej frekvencie a dĺžky trvania úsekov sú známe.

### 3.5.1 Modifikácia prozódie akustických jednotiek

Pri modifikovaní prozódie sa berú do úvahy hlavne dve veci. Po prvé, odhad syntetizačných okamihov, po druhé, odhad harmonických amplitúd a fáz pre zmenené, nové harmonické [67].

Syntetizačné časové okamihy určuje $t_s^i$ rekurzívny algoritmus z analyzačných časových okamihov $t_a^i$, z faktorov modifikácie tónu $\alpha(t)$ a z faktorov modifikácie časovej mierky $\beta(t)$. Za predpokladu, že originálna kontúra výšky tónu $P(t)$ je spojitá a že poznáme syntetizačný časový okamih $t_s^i$, môžeme určiť syntetizačný časový okamih $t_s^i$ ako

$$t_s^{i+1} = t_s^i + \frac{1}{t_v^{i+1} - t_v^i} \int_{t_v^i}^{t_v^{i+1}} P\,\frac{(t)}{a(t)}\,dt \tag{15}$$

kde $t_v^i$ znamená virtuálnu časovú konštantu, ktorá má takýto vzťah so syntetizačnými časovými konštantami

$$t_s^i = D(t_v^i) \tag{16}$$

kde $D(t)$ je mapovacia funkcia daná ako

$$D(t) = \int_0^t \beta(\tau)\, d\tau \tag{17}$$

Analyzačná časová os je mapovaná do syntetizačnej časovej osi pomocou mapovacej funkcie $D(t)$. Virtuálne časové okamihy sú definované na analyzačnej časovej osi a vo všeobecnosti sa nezhodujú s reálnymi analyzačnými časovými okamihmi. Z toho dôvodu máme pre daný virtuálny časový okamih dve možnosti – buď interpolovať HNM parametre od $t_a^i$ do $t_a^{i+1}$, alebo posunúť $t_v^i$ do najbližšieho časového okamihu ($t_a^i$ alebo $t_a^{i+1}$). Bežne sa v implementáciách využíva druhá možnosť.

Integrály môžu byť ľahko aproximované, ak je splnený predpoklad, že $P(t)$, $\alpha(t)$ a $\beta(t)$ sú po častiach konštantné funkcie. Špeciálna pozornosť sa musí venovať napájaciemu bodu, kde majú modifikačné faktory a kontúra výšky tónu veľké nespojitosti.

Keď sú už určené syntetizačné časové okamihy, ďalším krokom je odhad amplitúd a fáz tónovo modifikovaných harmonických. Najpriamejším prístupom, ktorý sa často používa, je prevzorkovanie komplexného rečového spektra. Treba si uvedomiť, že komplexné spektrum rámca $t_i$ je nadvzorkované do maximálnej znelej frekvencie $F_m(t_i)$. Harmonická časť pred a po tónovej modifikácii teda zaberá to isté frekvenčné pásmo [0 Hz - $F_m(t_i)$] [2], [66].

## 3.5.2 Spájanie akustických jednotiek

Počas spájania akustických jednotiek predstavujú HNM parametre nespojitosti v bodoch spájania. Z hľadiska vnímania nie sú parametre šumovej časti (AR koeficienty a variancia) dôležité. Z toho dôvodu sa pri vyhladzovaní nespojitosti berú do úvahy iba HNM parametre harmonickej časti (frekvencie, amplitúdy a fázy). Počas analyzačného procesu sú odstránené fázové nespojitosti medzi znelými rámcami, takže vyhladzovací algoritmus pozostáva len z odstraňovania frekvenčných nespojitostí a spektrálnych nespojitostí. Pre jednotky, ktorých prozódia nebola modifikovaná, sa môžu aj tak vyskytovať nespojitosti v bodoch spájania [66].

Spektrálne aj frekvenčné nespojitosti sa odstraňujú pomocou jednoduchej lineárnej interpolačnej techniky v okolí bodu napájania $t_i$. Najprv sa merajú rozdiely medzi hodnotami frekvencií a amplitúd pre každú harmonickú zložku v bode $t_i$. Potom sa tieto rozdiely váhujú a šíria sa doľava a doprava od bodu $t_i$. Počet rámcov, ktorý sa používa pri interpolačnom procese, závisí od zmeny počtu harmonických zložiek a veľkosti (v rámcoch) základných jednotiek (foném) pozdĺž bodu napájania [2], [66].

Nech $u^l$ a $u^r$ označujú ľavé a pravé akustické jednotky pozdĺž bodu napájania. Nech $\omega_0^i$ a $A_k^i$ označujú fundamentálnu frekvenciu a amplitúdu $k$-tej harmonickej zložky

z posledného rámca $u^l$, respektíve nech $\omega_0^{i+1}$ a $A_k^{i+1}$ označujú fundamentálnu frekvenciu a amplitúdu $k$-tej harmonickej zložky z prvého rámca $u^r$. Potom sú frekvenčné nespojitosti vyhladené pre $L$ rámcov v $u^l$ a pre $R$ rámcov v $u^r$ pomocou vzťahov

$$\Delta\omega_0 = \frac{\omega_0^{i+1} - \omega_0^i}{2} \tag{18}$$

$$\tilde{\omega}_0^l = \omega_0^i + \Delta\omega_0 \frac{i}{L} \tag{19}$$

$$\text{pre } i = L, L-1, \ldots 1$$

$$\tilde{\omega}_0^r = \omega_0^r + \Delta\omega_0 \frac{n}{L} \tag{20}$$

$$\text{pre } n = R, R-1, \ldots 1$$

$$\text{kde } r = i + 1 - R - n \tag{21}$$

Harmonické amplitúdy sú vyhladzované podobným spôsobom a používajú rovnaký počet rámcov $R$ a $L$ ako v predchádzajúcich vzťahoch (pre každú harmonickú $k$)

$$\Delta A_k = \frac{A_k^{i+1} - A_k^i}{2} \tag{22}$$

$$\tilde{A}_k^l = A_k^i + \Delta A_k \frac{i}{L} \tag{23}$$

$$\text{pre } n = L, L-1, \ldots 1$$

$$\tilde{A}_k^r = A_k^r + \Delta A_k \frac{n}{L} \tag{24}$$

$$\text{pre } n = R, R-1, \ldots 1$$

$$\text{kde } r = i + 1 - R - n \tag{25}$$

Táto jednoduchá lineárna interpolácia spektrálnych obálok robí formantové nespojitosti menej počuteľné. Ak sú ale formantové frekvencie naľavo a napravo veľmi odlišné, tak nie je problém úplne vyriešený. Použitie algoritmu výberu jednotiek by malo vybrať a spájať také akustické jednotky, ktoré nemajú veľké nespojitosti vo formantových frekvenciách [66].

### 3.5.3 Generovanie tvaru vlny v časovej oblasti

Syntéza sa často uskutočňuje využitím synchronizácie okien s fundamentálnou frekvenciou a pomocou aditívnej syntézy s prekryvom. Pre syntézu harmonickej časti sa využíva už spomenutý vzťah (7).

Pre výpočet amplitúd šumovej časti prvou metódou sa používa ten istý postup ako pri harmonickej časti. $F_0$ = 100 Hz a fázy sú nastavené náhodne. Pri druhej metóde dostaneme šumovú časť filtrovaním jednotkovej variancie bieleho Gaussovského šumu cez normalizovaný celopólový filter. Výstup z LP filtra je prenásobený obálkou variancií, ktoré boli odhadnuté počas analyzačného procesu. Ak je rámec znelý, šumová časť je filtrovaná cez hornopriepustný filter s hraničnou frekvenciou, ktorá je rovná maximálnej znelej frekvencii zodpovedajúcej danému rámcu. Na záver je šumová časť modulovaná obálkou v časovej oblasti (parametrická trojuholníková obálka), ktorá je synchronizovaná s periódou fundamentálnej frekvencie. V časovej oblasti sa potom šum a harmonické sínusoidy spočítajú [2], [66].

# 4 Kompresia a modelovanie reči

Napriek vzostupu prenosovej šírky pásma v optických sieťach je potrebné zvyšovať efektivitu prenosu a šifrovania v bezdrôtových a mobilných komunikáciách. Mobilná komunikácia zaznamenala celosvetový rozmach v rozvoji privátnej mobilnej komunikačnej siete. Na druhej strane je trend rozvoja zvukových aplikácií, buď samostatne alebo ako časť multimediálnej aplikácie, na PC alebo mobilných telefónoch. Väčšina týchto aplikácií pracuje s digitálnym signálom, pretože v digitálnej forme je jednoduchšie signál modifikovať, ukladať alebo prenášať. Hoci digitálny rečový signál prináša flexibilitu a možnosti šifrovania, v nekomprimovanej podobe má väčšie nároky na šírku pásma a prenosovú rýchlosť.

Pri prenose reči nie je základnou podmienkou verný prenos pôvodného signálu, ale v prvom rade dobrá zrozumiteľnosť reči. Tu nastupujú metódy, ktoré sa nesnažia zachovať pôvodný priebeh rečového signálu, ale signál modelujú s ohľadom na vlastnosti ľudského vnímania reči. Existuje veľa prístupov, ako ľudskú reč kódovať a efektívne prenášať. Oblasť, ktorá je zameraná na dosiahnutie kompaktnej digitálnej reprezentácie na účely efektívneho prenosu alebo uloženia dát, sa nazýva kompresia reči.

Rečové signály sú analógové, a preto sa musí pred digitálnym spracovaním uskutočniť transformácia zo spojitej do diskrétnej oblasti. Transformácia využíva časovú diskretizáciu, čiže vzorkovanie, po ktorom nasleduje kvantizácia. Vzorkovanie musí byť robené s dostatočne veľkou frekvenciou, aby sa nestratila frekvenčná informácia signálu. Kvantizácia do signálu vždy prináša určité skreslenie. Z toho dôvodu je kvantizácia kritická pre celú výkonnosť systému. Zatiaľ čo signál sa skoro vždy vzorkuje rýchlosťou rovnou alebo väčšou ako dvojnásobok pásma analógového signálu, veľká variabilita je v navrhovaných metódach reprezentácie navzorkovaného signálu. Cieľom kódovania reči je reprezentovať reč minimálnym počtom bitov so zachovaním vnemovej kvality reči. Kvantizácia alebo binárna reprezentácia môže byť priama alebo parametrická. Priama kvantizácia navrhuje binárnu reprezentáciu vzoriek signálu sebou samými, zatiaľ čo parametrická kvantizácia zahŕňa binárnu reprezentáciu rečového modelu a/alebo spektrálnych parametrov. Pre kódovanie reči v digitálnych rečových komunikáciách je reč vo všeobecnosti obmedzená do 4 kHz a navzorkovaná pomocou 8 kHz vzorkovacej frekvencie.

Najjednoduchšou neparametrickou kódovacou technikou je pulzná kódová modulácia (PCM). Reč kódovaná 64 kbit/s pomocou logaritmickej PCM sa považuje za „nekomprimovanú“ a často sa používa ako referenčná pre porovnávanie s inými typmi kódovaní. Ako signály so strednou rýchlosťou sa označujú signály kódované s prenosovými rýchlosťami v rozsahu od 8 do 16 kb/s, s nízkou rýchlosťou sú signály kódované v rozsahu od 2,4 po 8 kbit/s a s veľmi malou rýchlosťou sa označujú signály pod 2,4 kbit/s. Kódovanie reči v stredných prenosových rýchlostiach a nižších sa dosahuje pomocou analyzačno-syntetizačného procesu [1].

V analyzačnom stupni je reč reprezentovaná pomocou pevnej množiny parametrov, ktoré sú efektívne kódované. V stupni syntézy sú tieto parametre dekódované a sú použité v kombinácii s rekonštrukčným mechanizmom na spätné vytvorenie reči. Analýza môže byť robená pomocou otvoreného alebo uzavretého cyklu. V analýze s uzavretým cyklom sa parametre vyberajú a kódujú pomocou minimalizovania explicitnej miery (zvyčajne strednej kvadratickej hodnoty) rozdielu medzi originálnym a rekonštruovaným signálom. Z toho dôvodu zahŕňa analýza s uzatvoreným cyklom aj syntézu a tento proces sa nazýva aj analýza pomocou syntézy.

Parametrické reprezentácie môžu, ale nemusia, byť špecifické pre reč. Kodéry, ktoré nie sú špecifické pre reč (kodéry tvaru vlny), sa zúčastňujú na vernej reprezentácii signálu v časovej oblasti a vo všeobecnosti operujú na stredných prenosových rýchlostiach. Rečovo špecifické kodéry – vokodéry (voice coder) sa spoliehajú na rečové modely a sú zamerané na tvorenie vnemovo zrozumiteľnej reči bez nutnosti kopírovať tvar vlny [1]. Vokodéry sú schopné operovať na veľmi malých prenosových rýchlostiach, ale majú tendenciu produkovať syntetickú kvalitu reči. Existujú však aj kodéry, ktoré kombinujú vlastnosti z oboch kategórií, napríklad rečovo špecifické kodéry tvaru vlny ako ATC (Adaptive Transform Coder) a tiež hybridné kodéry, ktoré sú založené na lineárnej predikcii analýzy pomocou syntézy. Hybridné kodéry kombinujú efektívnosť kódovania vokodérov s potenciálom vysokej kvality kodérov tvaru vlny pomocou modelovania spektrálnych vlastností reči a využívaním vnemových vlastností ľudského ucha, za súčasnej snahy zachovania tvaru vlny [1]. Moderné hybridné kodéry môžu dosiahnuť prenosové rýchlosti 8 kbit/s a nižšie pod podmienkou vyššej zložitosti a výpočtovej náročnosti.

Podrobnejšie budú najznámejšie metódy popísané v nasledujúcich kapitolách a potom budú popísané špecifiká kódovania reči pomocou sínusoidálneho modelu.

## 4.1 História kódovania reči

Výskum kódovania reči začal pred viac než päťdesiatimi rokmi prácou Dudleyho z Bellových laboratórií. V tom čase sa venovala veľká pozornosť výskumu systémov na prenos reči cez úzkopásmové telegrafné káble, čo bolo veľkou motiváciou na výskum kódovania reči. Dudley predstavil svoj vokodér založený na prvej analyzačno-syntetizačnej metóde na kódovanie reči. Základnou myšlienkou tohto vokodéra bolo analyzovať reč podľa jej spektra a výšky a syntetizovať ju budením banky desiatich analógových pásmovo-priepustných filtrov, ktoré reprezentovali hlasový trakt. Budenie sa realizovalo dvoma spôsobmi, v prípade znelých zvukov to bol bzukot, v prípade neznelých zvukov to bol náhodný signál, šum. Tomuto kanálovému vokodéru sa dostalo najväčšej pozornosti počas druhej svetovej vojny vďaka jeho potenciálu na efektívny prenos šifrovanej reči [43].

Počas päťdesiatych a šesťdesiatych rokov sa používali formantové [44] a pattern-mat-

ching [45] vokodéry so zlepšenými analógovými implementáciami kanálových vokodérov. Vo formantovom vokodéri sledovali pohyb formantov rezonančné charakteristiky banky filtrov.

V pattern-matching vokodéroch sa určila najlepšia zhoda medzi krátkodobým spektrom reči a množinou uložených vzoriek frekvenčných odpovedí. Reč bola produkovaná budením kanálového filtra pridruženého k určenej vzorke. Bol to prvý vokodér s implicitným uplatnením vektorovej kvantizácie.

Prvé implementácie vokodérov boli založené na analógovej reprezentácii reči. Neskôr sa však v pomerne krátkom čase zvýšil záujem o digitálne reprezentácie, pretože sa objavovali veľké výhody šifrovania a vysoká úroveň kvality prenosu a uloženia dát. Záujem sa zvýšil najmä v štyridsiatych rokoch po publikovaní PCM metódy [48] (4.2.1.1). PCM je metóda na aproximáciu analógových signálov diskrétnymi signálmi s diskrétnou amplitúdou a nemá žiaden mechanizmus na odstránenie redundancií. Neskôr boli navrhnuté kvantizačné metódy, ktoré využívajú koreláciu signálu, ako sú DPCM (Diferenčná PCM) (4.2.1.2), DM (Delta Modulácia), ADPCM (Adaptívna DPCM) (4.2.1.3). Kódovanie reči podľa PCM s rýchlosťou 64 kbit/sa ADPCM s rýchlosťou 32 kbit/ssa stali štandardmi podľa CCITT [49].

Vďaka flexibilite digitálnych počítačov sa počiatočné úsilie koncentrovalo na digitálnu implementáciu vokodéra. Koncom päťdesiatych rokov sa činnosť koncentrovala na lineárny model produkcie reči systémom zdroja, ktorý bol vyvinutý Fantom [50]. Tento model sa neskôr vyvinul na dobre známy model produkcie reči, ktorý pozostáva z lineárneho, v čase pomaly sa meniaceho systému (pre hlasový trakt a model hlasiviek), ktorý je budený periodickým radom impulzov (pre znelú reč) alebo náhodným budením (pre neznelú reč).

Z tohto modelu sa spojením s autoregresívnymi metódami vyvinul systém, kde sa hlasový trakt modeluje pomocou celopólového filtra. Parametre filtra sa získavajú pomocou LPC (Lineárnej Predikčnej Analýzy) [51] – procesom, kde je aktuálna vzorka rečového signálu predikovaná pomocou lineárnej kombinácie predchádzajúcich vzoriek. Okrem lineárnej predikcie využíva analýzu zdroja systému aj homomorfná analýza. Je to metóda, ktorá môže byť použitá na separáciu signálov, ktoré boli kombinované konvolúciou [52]. Jednou z výhod homomorfnej analýzy reči je dostupnosť informácie o výške tónu priamo z kepstra.

V šesťdesiatych a sedemdesiatych rokoch vznikli podmienky pre implementáciu nových a vylepšených riešení kódovania reči. Prispel k tomu objav VLSI technológií a pokrok v teórii číslicového spracovania signálov. Vznikol fázový vokodér [53], analyzačno-syntetizačný systém, ktorý využíval krátkodobú Fourierovu transformáciu (SFT). Vytvoril sa teoretický základ pre časovo-frekvenčnú analýzu reči pomocou SFT. Do konca sedemdesiatych rokov sa paralelne venovalo úsilie aj systémom lineárnej predikcie, transformačnému kódovaniu a podpásmovému (sub-band) kódovaniu.

Na vývoj robustných rečových kodérov s nízkymi prenosovými rýchlosťami produkujúcich vysokokvalitnú reč pre komunikačné aplikácie bolo zamerané úsilie osemdesiatych

a deväťdesiatych rokov. Hlavnou motiváciou tohto úsilia bola potreba úzkopásmových a bezpečných prenosov rečových dát v bezdrôtových bunkových a vojenských komunikáciách. Vďaka tomu vzniklo niekoľko metód, ako napríklad sínusoidálna analýza-syntéza reči [47], vokodéry multipásmového budenia (multi-band excitation) [54], systémy multi-pulzového a vektorového budenia pre LPC [55] a vektorovú kvantizáciu [46], [56]. Vektorová kvantizácia sa ukázala ako veľmi užitočná pri kódovaní LPC parametrov. Najznámejší algoritmus lineárnej predikcie so stochastickým vektorom budenia sa nazýva CELP [57] (Code Excited Linear Prediction).

V súčasnosti sa naďalej vyvíja veľká snaha znížiť prenosovú rýchlosť a súčasne zachovať kvalitu reči. Existujú tendencie súčasne využívať viacero algoritmov na kódovanie reči – tzv. hybridné kodéry. Podrobnejšie ich popíšeme v ďalšej kapitole.

Najnovší trend v sínusoidálnej analýze a syntéze reči je modelovanie reči pomocou HNM (harmonického a šumového) modelu, ktorý modeluje reč ako sumu harmonických sínusoíd (ich frekvencia je násobkom hlasivkovej frekvencie) a šumu.

## 4.2 Vlnové kodéry

Kodéry na modelovanie tvaru vlny sú navrhované tak, že sú nezávislé od signálu, a preto môžu byť použité na širšie spektrum signálov, nie iba na rečový signál. Veľkou výhodou je malá degradácia pri prítomnosti šumu a pri prenosových chybách. Ich nevýhodou je, že kvôli efektivite sa môžu využívať len pri stredných prenosových rýchlostiach. Modelovanie tvaru vlny sa môže vykonávať v časovej aj frekvenčnej oblasti.

### 4.2.1 Modelovanie v časovej oblasti

#### 4.2.1.1 PCM (Pulse Code Modulation)

PCM je najjednoduchší spôsob kódovania tvaru vlny. V podstate je to len kvantizačný proces. Každá vzorka je kvantizovaná do jednej z konečného množstva úrovní. Každá úroveň je označená jedinečnou postupnosťou binárnych čísiel. V tejto schéme sa môže využívať ľubovoľná schéma skalárnej kvantizácie, ale najpoužívanejšou formou pre kódovanie telefónnej reči je 8-bitová logaritmická kvantizácia [2].

#### 4.2.1.2 DPCM (Differential Pulse Code Modulation)

Pretože PCM nerobí predpoklady o povahe signálu, ktorý sa kóduje, veľmi dobre funguje aj pre nerečové signály. Pri kódovaní reči však existuje veľmi veľká korelácia medzi susednými vzorkami. Táto korelácia sa môže použiť na redukciu bitovej rýchlosti.

Jednoduchou metódou je prenášať len rozdiely medzi susednými vzorkami. Tieto rozdiely budú mať oveľa menší dynamický rozsah ako pri originálnej reči, takže môžu byť efektívne kvantizované použitím kvantizéra s menším počtom úrovní. V tejto metóde sa predchádzajúca vzorka používa na predikciu nasledujúcej. Táto metóda sa zlepší, ak sa na predikciu použije väčší blok predchádzajúcich vzoriek. Technika sa nazýva DPCM. Predikovaná hodnota je lineárnou kombináciou konečného počtu posledných vzoriek. Rozdiel medzi predikovanou a skutočnou hodnotou sa nazýva rezíduum, ktoré sa kvantizuje a prenáša. Koeficienty prediktora sú nastavené tak, aby sa minimalizovala stredná kvadratická chyba medzi originálnym a predikovaným signálom [2].

#### 4.2.1.3 ADPCM (Adaptive Differential Pulse Code Modulation)

Pri DPCM sa ani prediktor, ani kvantizér v čase nemenia. Lepšiu efektívnosť môžeme dosiahnuť v prípade, ak sa bude kvantizér adaptovať na zmenu štatistiky rezídua. Ďalšie zlepšenie môžeme dosiahnuť, ak sa bude sám prediktor adaptovať na rečový signál. To zabezpečí, že sa stredná kvadratická chyba bude neustále minimalizovať nezávisle od rečového signálu. ADPCM sa často používa na kódovanie reči pri stredných prenosových rýchlostiach [2].

### 4.2.2 Modelovanie vo frekvenčnej oblasti

Pri modelovaní tvaru vlny vo frekvenčnej oblasti sa signál rozdelí do množstva oddelených frekvenčných zložiek, ktoré sa nezávisle kódujú.

#### 4.2.2.1 Pásmová filtrácia

Pásmová filtrácia je najjednoduchšou technikou vo frekvenčnej oblasti. Signál prejde bankou pásmovo priepustných filtrov. Každé frekvenčné pásmo je potom podvzorkované podľa Nyquistovho kritéria. Pásma sa potom kódujú pomocou jednej z techník popísaných vyššie. Hlavnou výhodou pásmovej filtrácie je, že kvantizačný šum produkovaný v jednom pásme je obmedzený na toto pásmo. Zabraňuje to kvantizačnému šumu maskovať frekvenčné zložky v iných pásmach. Pre každé pásmo potom môžu byť použité odlišné úrovne kvantizácie [2].

#### 4.2.2.2 Transformačné kódovanie

Transformačné kódovanie je zložitejšia technika, ktorá využíva blokovú transformáciu naoknovaného segmentu vstupného signálu. Myšlienkou tohto kódovania je, že signál je transformovaný do frekvenčnej, alebo inej podobnej oblasti takým spôsobom, že vzorky sú vzájomne nekorelované. Kódovanie sa potom robí tak, že sa viac bitov priradí dôležitejším transformačným koeficientom. V prijímači sa potom signál dekóduje pomocou inverznej transformácie. Môže sa používať viacero transformácií, ako diskrétna Fourierova transformácia (DFT), Karhunen-Loeveho transformácia (KLT), alebo najčastejšie používaná diskrétna kosínusová transformácia (DCT).

Transformačné kódovanie sa široko využíva pri kódovaní širokopásmových audio a video signálov, avšak kvôli zložitosti sa obvykle nevyužíva na kódovanie rečových signálov [2].

### 4.2.3 Modelovanie založené na vlastnostiach reči

Pri modelovaní tvaru vlny sa nerobia predpoklady o povahe signálu, ktorý sa modeluje. Ak je signálom len reč, tak bude efektívnejšie, ak sa to v metóde, ktorá signál modeluje, vezme do úvahy.

Vokodéry predpokladajú explicitný model rečovej produkcie. Tento model predpokladá, že reč je produkovaná budením lineárneho systému (hlasového traktu) sériou periodických impulzov (ak je reč znelá), alebo šumom (ak je reč neznelá). Pri znelej reči pozostáva budenie z periodického sledu impulzov. Vzdialenosti medzi impulzmi sa rovnajú perióde výšky tónu. Ak je reč neznelá, budením je postupnosť náhodného šumu, ktorá zodpovedá šumu, ktorý vzniká tým, že vzduch prúdi cez prekážky hlasového traktu. Lineárny systém, ktorý modeluje hlasový trakt a jeho parametre, môžeme určiť viacerými technikami. Vokodéry sa pokúšajú produkovať signál, ktorý znie ako originálna reč, aj keď sa priebeh signálu originálnemu signálu nepodobá. Reč sa analyzuje, aby sa určili parametre modelu a budenie, z ktorých sa potom reč spätne syntetizuje. Výsledkom je, že vokodéry môžu produkovať zrozumiteľnú reč pri veľmi malých bitových rýchlostiach.

Takto syntetizovaná reč znie neprirodzene, preto sa vokodéry využívajú najmä tam, kde je na prvom mieste požiadavka malej prenosovej rýchlosti [2].

#### 4.2.3.1 Kanálový vokodér

Kanálový vokodér využíva necitlivosť ľudského ucha na krátkodobú fázu. Pre rečové segmenty dĺžky 20 ms stačí vziať do úvahy len magnitúdu spektra. Spektrum sa odhaduje pomocou banky filtrov. Je zrejmé, že čím bude väčšia banka filtrov, tým bude lepší odhad, ale zároveň aj vyššia bitová rýchlosť. Výstup každého filtra je potom usmernený a dol-

nopriepustne filtrovaný, aby sa našla obálka signálu. Tá je navzorkovaná a prenesená do prijímača. V prijímači sa potom robí opačný postup, ako vo vysielači. Pásma filtrov, ktoré sa používajú v banke filtrov, sa so zvyšujúcou frekvenciou zväčšujú. Je to možné vďaka tomu, že ľudské ucho má približne logaritmické vnímanie frekvencií [2].

### 4.2.3.2 Homomorfný vokodér

Homomorfné spracovanie signálov patrí do všeobecnej triedy nelineárnych techník spracovania signálov, ktoré môžu veľmi efektívne pracovať so zloženými signálmi. Vokodér predpokladá, že rečový signál vznikol konvolúciou (v čase) impulzovej odpovede hlasového traktu a budenia. Táto konvolúcia zodpovedá násobeniu vo frekvenčnej oblasti. Ak vezmeme do úvahy logaritmus spektra, z násobenia sa stane sčítanie. Pretože ľudské ucho je necitlivé na fázu, berieme do úvahy len magnitúdu spektra:

$$\log\left(\left|S\left(e^{j\omega}\right)\right|\right)=\log\left(\left|B\left(e^{j\omega}\right)\right|\right)+\log\left(\left|V\left(e^{j\omega}\right)\right|\right) \tag{26}$$

kde $S(e^{j\omega})$ je spektrum reči, $B(e^{j\omega})$ je spektrum budenia a $V(e^{j\omega})$ je spektrum hlasového traktu.

Kepstrálne koeficienty dostaneme inverznou Fourierovou transformáciou logaritmického spektra. Pri produkcii reči sa impulzová odpoveď hlasového traktu mení v čase len pomaly, zatiaľ čo budenie sa mení rýchlo. Tento rozdiel sa uchováva v kepstre a je zjavný ako časový posun. V kepstre sa môžu príspevky oboch zložiek reči ľahko od seba oddeliť pomocou dolnopriepustného filtra. Je ľahké detegovať presnú informáciu o výške tónu. V homomorfnom kodéri sa potom na popis reči uchovávajú kepstrálne koeficienty spolu s informáciou výšky tónu [2].

### 4.2.3.3 LPC vokodér

Lineárny prediktívny vokodér (LPC vokodér) využíva ten istý model produkcie reči, ako predchádzajúce vokodéry. Líši sa len metódou používanou na modelovanie hlasového traktu. Tento vokodér predpokladá, že hlasový trakt môže byť popísaný pomocou celopólového IIR filtra, ktorého prenosová charakteristika je

$$H\left(z\right)=\frac{G}{1+a_1 z^{1}+a_2 z^{2}+K+a_p z^{p}} \tag{27}$$

Inými slovami, predpokladá sa, že každá vzorka rečového signálu je lineárnou kombináciou predchádzajúcich vzoriek. Koeficienty filtra sa počítajú tak, aby sa mini-

malizovala chyba medzi predikciou a skutočnou vzorkou. Vo vokodéri sa na analýzu najčastejšie používajú bloky dĺžky 20 ms (reč môžeme považovať za stacionárnu, ak trvá od 10 do 30 ms), z ktorých sa určia predikčné koeficienty. Úlohou prediktora je odstrániť koreláciu medzi susednými vzorkami. Na syntézu je potom potrebné poznať predikčné koeficienty a budenie [2].

### 4.2.4 Hybridné modelovanie

Ako sme mohli vidieť, existujú dve hlavné triedy rečových modelov – modely tvarovania vlny a vokodéry. Modely tvarovania vlny sa snažia modelovať priebeh vlny signálu. Sú schopné dosiahnuť vysokú kvalitu reči pri stredných prenosových rýchlostiach (okolo 32 kbit/s), avšak nemôžu kódovať reč pri nízkych prenosových rýchlostiach.

Na druhej strane sa vokodéry snažia produkovať signál, ktorý znie ako pôvodný signál. Môžu dosahovať veľmi malé prenosové rýchlosti (menej ako 4 kbit/s), ale výsledná reč znie synteticky, aj keď je zrozumiteľná. Z týchto dôvodov sa vytvorili modely, ktoré kombinujú obidve techniky a dosahujú prenosové rýchlosti okolo 8 kbit/s. Tieto hybridné modely sa snažia zachovať z hľadiska vnímania dôležité časti rečového signálu [2].

#### 4.2.4.1 RELP (Residual Excited Linear Prediction)

Keď prejde signál cez lineárny prediktor, odstráni sa medzivzorková korelácia. Ak je predikcia dostatočne dobrá, tak bude výstup prediktora aproximovať biely šum, takže rezíduum bude spektrálne hladké. Rezíduum obsahuje celú informáciu budenia a všetky informácie, ktoré lineárny prediktor nezachytil. Myšlienkou RELP modelovania je, že stačí preniesť malú časť rezídua, z ktorej sa dá v prijímači vytvoriť celé rezíduum. Pri RELP prejde rezíduum cez dolnopriepustný filter s hraničnou frekvenciou 1000 Hz. Výstup filtra sa kóduje pomocou niektorej metódy tvarovania vlny. V prijímači je rezíduum rekonštruované kopírovaním základného pásma rezídua do ostatných pásiem, ktoré berieme do úvahy.

Hlavnou nevýhodou RELP modelovania je, že LPC filter neodstráni jemnú spektrálnu štruktúru, ktorá je daná výškou tónu hlasu. Je vysoko nepravdepodobné, že by hraničná frekvencia dolnopriepustného filtra zodpovedala niektorej harmonickej frekvencii výšky hlasu. To znamená, že pri rekonštruovaní rezídua bude vnemovo dôležitá informácia o výške tónu pre vyššie pásma nekorektná. Tento problém sa dá zmierniť prispôsobením hraničnej frekvencie dolnopriepustného filtra frekvencii výšky hlasu [2].

### 4.2.4.2 MPC (Multi-Pulse Coding)

Hlavným problémom pri LPC modelovaní je modelovanie budenia. Vokodér delí reč na znelú alebo neznelú a neexistuje nič medzi tým. Preto takto vyprodukovaná reč znie veľmi synteticky.

MPC sa tento problém snaží riešiť. Keď reč prejde cez lineárny prediktor, odstráni sa korelácia medzi susednými vzorkami. Pre znelú reč však výška hlasu zavádza dlhodobú koreláciu, ktorá vyplýva z kvázi-periodicity spomenutej vyššie. Táto periodickosť nie je odstránená lineárnym prediktorom a produkuje v rezíduu značné špičky. Dlhodobá korelácia môže byť odstránená tým, že rezíduum prejde cez druhý lineárny prediktor – prediktor výšky tónu. Tento prediktor je navrhnutý tak, aby odstraňoval koreláciu medzi susednými periódami signálu. Je to dosiahnuté zavedením oneskorenia do prediktora, ktoré zodpovedá perióde výšky tónu:

$$P(z)=1-\sum_{i=1}^{M}\beta_i z^{-M-i} \tag{28}$$

kde $M$ je perióda výšky tónu.

Výstup prediktora bude aproximovať biely Gaussovský šum. Pri syntéze potom MPC kodér budí kaskádu dvoch lineárnych prediktorov pomocou série impulzov. Vo všeobecnosti sa ako budenie používa štyri až šesť impulzov. Pozícia a amplitúda impulzov sa určuje procedúrou analýzy pomocou syntézy (čiže pre každú množinu impulzov sa syntetizuje reč a vyberie sa tá množina, ktorá produkuje najmenšiu chybu medzi originálnou a rekonštruovanou rečou). Polohy impulzov sa určujú postupne (najprv sa hľadá optimum pre prvý impulz, potom pre druhý, atď.) [2].

### 4.2.4.3 CELP (Codebook Excited Linear Prediction)

Pri CELP modelovaní prejde reč cez kaskádu dvoch prediktorov - hlasového traktu a prediktora výšky hlasu. Výstupom je dobrá aproximácia Gaussovského šumu. Táto postupnosť musí byť kvantizovaná - CELP kodéry ju kvantizujú pomocou vektorovej kvantizácie. Uloží sa ten index kódového slova, ktoré produkuje najlepšiu kvalitu reči.

Vyhľadávanie v kódovej knihe sa uskutočňuje technikou analýzy pomocou syntézy. Reč sa syntetizuje pre každý údaj v kódovej knihe. Ako budenie sa vyberie kódové slovo, ktoré produkuje najmenšiu chybu. Meranie chyby je vnemovo váhované, takže sa vyberie to kódové slovo, ktoré znie najlepšie. Vyhľadávanie v kódovej knihe je výpočtovo veľmi náročné, ale boli vyvinuté rýchle algoritmy, takže CELP modelovanie môže byť implementované aj v reálnom čase [2] .

### 4.2.4.4 MBE (Multiband Excitation)

MBE model využíva ten istý dvojstavový model ako tradičné vokodéry, teda reč je produkovaná excitáciou modelu hlasového traktu budením. Model hlasového traktu je ten istý ako pri ostatných vokodéroch, ale metóda modelovania budenia úplne iná. Pri tradičných vokodéroch je budenie definované ako znelé/neznelé. Ak je znelé, tak je potrebná aj fundamentálna frekvencia a budením je rad periodických impulzov; pre neznelú reč je budením sekvencia šumu.

MBE model nahradzuje jednoduché rozhodnutie znelá/neznelá reč sériou takýchto rozhodnutí. Budenie je rozdelené do niekoľkých pásiem. V každom pásme je spektrum reči analyzované a určí sa, či je znelé alebo neznelé. Znelé segmenty sú potom podobne ako pri iných vokodéroch kódované radom periodických impulzov a neznelé segmenty šumom. Vďaka tomu je možné, aby mala modelovaná reč súčasne charakteristiky znelej aj neznelej reči [2].

### 4.2.4.5 STM (Sinusoidal Transform Modeling)

Sínusoidálny transformačný kodér sa podobá transformačným kodérom spomenutým vyššie. STM kodéry reprezentujú rečový signál ako lineárnu kombináciu L sínusoíd s časovo sa meniacimi frekvenciami, amplitúdami a fázami:

$$\hat{s}(n)=\sum_{k=1}^{L} A_k \cos\left(\Omega_k n+\phi_k\right) \tag{29}$$

kde $A_k$ je amplitúda, $\Omega_k$ je frekvencia a $\phi_k$ je fáza.

Počet sínusoíd sa môže v čase meniť. Znelá reč má tendenciu byť vysoko periodická, takže môže byť reprezentovaná malým počtom sínusoíd. Neznelá reč môže byť tiež zachovaná pomocou malého počtu sínusoíd s náhodnými fázami. Preto je možné, aby bol rečový signál popísaný použitím malého počtu parametrov.

Samotné sínusoidy môžeme získať pomocou krátkodobej Fourierovej transformácie (SFT). Vyberajú sa tie sínusoidy, ktoré zodpovedajú špičkám SFT. Na reč sa zvyčajne aplikuje Hammingovo okno s dĺžkou 2,5-násobku periódy priemernej výšky tónu. Je to potrebné pre dostatočné frekvenčné rozlíšenie. STC modelovanie sa zároveň dá veľmi dobre využiť aj pri modelovaní nerečových signálov. Pre aplikácie, ktoré vyžadujú malé prenosové rýchlosti, môžu byť frekvencie obmedzené na násobky fundamentálnej frekvencie. Takýto typ modelu sa nazýva HNM (Harmonic plus Noise Model). Posledným dvom spomenutým kodérom sa budeme ešte v ďalších kapitolách venovať podrobnejšie [2].

## 4.3 Kompresné algoritmy

V dnešnej dobe má digitálne spracovanie zvuku veľký význam, najmä z hľadiska zachovávania omnoho väčšej vernosti zvuku, napr. pri archivovaní. Pri prenose audio signálu analógovým prenosom nie je ľahké zabezpečiť jeho vysokú kvalitu. Preto je audio v digitálnej podobe lepšou voľbou. Pri prenose a archivovaní audio signálov (teda aj reči) sa kladie dôraz na čo najmenší objem dát. Hlavné dôvody sú veľmi jednoduché: menšie náklady, praktickejšie archivovanie, úspora prenosového pásma.

Na tento účel sa používa metóda takzvanej stratovej kompresie, pri ktorej dochádza k bezstratovej kompresii, ale len ohraničeného informačného obsahu vstupných dát. Pre kódovanie (stratovú kompresiu) audio signálu sa používa tzv. „perceptual audio co¬ding" metóda, ktorá využíva fakt, že ľudský sluchový systém vie zachytiť len obmedzené množstvo informácií. Táto metóda sa snaží o čo najvernejšie zakódovanie. Prípustné sú straty, ktoré sú najmenej potrebné pre presnosť vnemu [101].

V súčasnosti je väčšina audio kodekov založená na práci Expertnej skupiny pre pohyblivý obraz Motion Picture Experts Group (MPEG), ktorá je súčasťou Medzinárodnej štandardizačnej organizácie (International Standards Organization, ISO). Počas jej existencie skupina uviedla niekoľko audio formátov, ktoré sa celosvetovo používajú.

Ako bude ďalej zrejmé, kodeky z rodiny MPEG sú založené na stratovom kódovaní, čo znamená, že modifikujú pôvodný audio signál a rekonštruovaný signál nie je nikdy zhodný s pôvodným [101].

Najnovšie trendy v kódovaní audia sú metódy, ktoré s postupujúcim vývojom využívajú nielen nedokonalosť ľudského vnemu, ale aj časovú štruktúru signálu s dôrazom na správne pochopenie zvuku a nie nevyhnutne jeho vysokú podobnosť s originálom.

Medzi najčastejšie používané patria predovšetkým algoritmy MPEG-1 layer 3 (MP3), MPEG-2 a MPEG-4 Advanced Audio Coding (AAC), ktoré boli vyvinuté organizáciou MPEG (Motion Picture Experts Group) [15]. Hoci AAC predsta¬vuje následníka algoritmu MP3, majú toho veľmi veľa spoločného. Medzi tieto algoritmy sa dá zaradiť ešte jeden podobný, ktorý však pochádza od iných tvorcov. Nazýva sa Ogg-Vorbis (Ogg) [18] a podobnosťou leží niekde uprostred medzi MP3 a AAC. Veľmi známy je algoritmus Twin VQ [16]. Od ostatných sa líši najmä tým, že pou¬žíva vektorovú kvantizáciu, a hoci bol zverejnený už dávno, nestal sa veľmi populárnym, aj keď jeho kódovacie jadro sa nachádza v štandardoch MPEG-4 [17], [101].

Príkladom bezstratového kódovania sú dáta zakódované PCM postupnosťou. Hodnoty tejto postupnosti nie sú reálne čísla s nekonečnou presnosťou, ale sú reprezentované len konečným množstvom bitov. Pri stratovom kódovaní sa v algoritmoch MP3, AAC a Ogg používa iná reprezentácia ako PCM. Presnejšie povedané, MPEG a OGG používajú transformačné kódovanie, teda reprezentujú dáta v inej oblasti než PCM v časovej oblasti.

Všetky tri algoritmy používajú MDCT [19], [20] transformáciu (Modified Discrete Cosine Transformation), ktorá sa používa na zmenu reprezentácie zo zvukovej vlny do spektrálnej oblasti. MDCT sa neaplikuje na signál ako na celok, ale na bloky PCM postupnosti dĺžky n. Tieto sa potom transformujú na postupnosť hodnôt dĺžky n (teda MDCT sa vykonáva na n-ticiach hodnôt), ktoré sú už spektrálnou reprezentáciou. Na MDCT možno pozerať ako na transformáciu dimenzie n, pričom táto transformácia je zmenou bázy elementárnych vektorov e1,e2.,en na bázu skladajúcu sa z ortogonálnych sínusoíd rôznych vlnových dĺžok. Reprezentácia v spektre je veľmi podobná reprezentácii v sluchovom vneme človeka.

MP3, AAC aj Ogg umožňujú použiť v zásade dva rôzne prístupy kódovania stereo infor¬mácie (nad spektrálnou reprezentáciou). Prvý spôsob používaný pre väčšie šetrenie je nazývaný spojené stereo (Joint Stereo). Využíva vlastnosť, že ak počujeme v oboch ušiach rovnaký tón v tvare sínusoidy, ale s určitým fázovým posunom, pri vysokých tónoch nie sme schopní túto fázu rozlíšiť. Ďalším faktom vyplývajúcim z povahy ľudského ucha je, že panoramatickú polohu vnímame lepšie pre vyššie tóny. Preto princíp pri tomto type kódovania spočíva v tom, že signál sa ako dva plnohodnotné kanály kóduje len vo frekvenčnom pásme od 0 Hz do cca 2000 Hz. Zvyšné pásmo, 2000 Hz až 16000 Hz, je rozdelené do niekoľkých pásiem a kóduje ako mono signál (súčet kanálov) spolu s niekoľko málo informáciami o pomeroch intenzít jednotlivých kanálov pre každé pásmo. Pri dekódovaní je potom tento mono signál umiestnený medzi kanály tak, aby bol pomer intenzít v každom pásme rovnaký, ako bol v pôvodnom signáli.

Druhý spôsob kódovania sterea je súčtovo-rozdielne kódovanie (Middle/Side Coding). Na rozdiel od predošlého je toto kódovanie bezstratové, ale snaží sa využiť koreláciu kanálov v prospech kompresie. Preto namiesto kódovania hodnôt každého kanála sa kóduje súčet (middle) a rozdiel (side) kanálov. Predpokladá sa, že väčšia časť zvuku je v panoramatickom strede ľavo-pravej bázy. Preto bude informácia rozdielu (side) celkovo nadobúdať menšie hodnoty koeficientov spektra ako súčet (middle).

Ogg v metóde nazývanej bezstratové stereo (Lossless Stereo) používa spojenie oboch techník. V princípe to vyzerá tak, že meria intenzity v každom pásme a následne transformuje priestor hodnôt celého pásma pootočením súradnicovej sústavy (ľavý kanál ako os x, pravý kanál ako os y, spektrálne koeficienty ako dvojice) tak, aby sústredil intenzitu do jedného kanála. Výsledok je o to lepší, čím je korelácia medzi pôvodnými kanálmi väčšia.

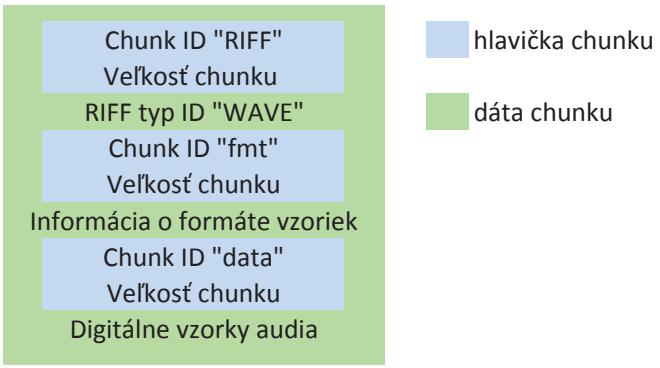
Ďalšími spoločnými črtami vyššie spomenutých kompresných metód je použitie kvantizácie spektrálnych koeficientov. Tu sa vy¬užíva psycho-akustický model, ktorý pre každý koeficient (resp. každé pásmo) určí hodnotu q – kvantizačný krok – ktorá je potrebnou presnosťou (lepší model znamená presnejšie určené q) kódovania spektrálnych koeficientov a mala by stačiť na to, aby bol zvuk vnímaný rovnako. Koeficienty sú potom kódované ako zaokrúhlené celočíselné násobky tohto q, čím sa dosahuje veľké šetrenie. Tieto q čísla bývajú tak malé, že ak by sme vypočítali priemer absolútnych hodnôt týchto celých čísel, dostali by sme číslo približne 1,5 alebo aj menej. To znamená, že veľa q čísel má hodnotu nula. Na kódovanie vzniknutých postupností celých čísel sa používa

kód s variabilnou dĺžkou slova (prípadne nejaká kombinácia s iným kódom) využívajúc entropiu a závislosť susedných koeficientov (v čase alebo vo frekvencii). Na dosiahnutie požadovaného bitového toku sa kvantizačný krok q prenásobí určitou hodnotou tak, aby sa dosiahla príslušná kompresia. Tým sa riadi globálna kvalita v čase, ktorá potom prislúcha danému bitovému toku.

### 4.3.1 Wave

Tento formát patrí do rodiny Microsoft RIFF (Resource Interface File Format) špecifikácií pre uchovávanie súborov rôznych typov dát z rôznych vstupno-výstupných zariadení:

- Bitmapy (.RDI)
- MIDI informácie (.RMI)
- Palety farieb (.PAL)
- Multimediálne filmy (.RMN)
- Animované kurzory (.ANI)
- Ostatné (.BND)
- Audio/Video dáta (.AVI)
- Zvukové dáta (.WAV)




***Obr. 17.*** *Základná štruktúra Wave*

Wave súbory používajú štandardnú RIFF štruktúru, ktorá zoskupuje obsah súboru do separátnych „chunkov". Každý obsahuje vlastnú hlavičku a dáta. Hlavička špecifikuje typ a veľkosť dát. Každý chunk môže obsahovať sub – chunky. Vidieť to aj na obrázku (Obr. 17). „Fmt" a „data" chunky sú sub – chunky RIFF. Veľkosť chunkov musí byť zarovnaná, to znamená že ich celková veľkosť musí byť násobkom 2 bajtov. Ak chunk obsahuje nepárny počet bajtov dátovej časti, pridáva sa jeden extra bajt na koniec dátovej časti s nulovou hodnotou.

Wave hlavička nasleduje RIFF štandard. Prvých 8 bytov je chunk ID a veľkosť wave súboru, ktorá nezahŕňa týchto 8 bytov. Potom nasleduje formát chunk „fmt“, ktorý obsahuje informácie o wave ako sú typ kompresie, počet kanálov, frekvenciu vzorkovania a iné. Ďalšie chunky, ktoré môže RIFF obsahovať, sú: data, fact, wavl, slnt, cue, plst, list, labl, note. Podrobná štruktúra hlavičky je znázornená na obrázku (Obr. 18) nižšie [87].

| endian | Posun (bajty) | Meno poľa | Veľkosť poľa (bajty) | |
|---|---|---|---|---|
| malý | 0 | Chunk ID | 4 | |
| malý | 4 | Veľkosť chunku | 4 | Definovanie "RIFF" chunku |
| veľký | 8 | Formát | 4 | |
| veľký | 12 | Subchunk1 ID | 4 | |
| malý | 16 | Veľkosť subchunku1 | 4 | |
| malý | 20 | Audio formát | 2 | |
| malý | 22 | Počet kanálov | 2 | |
| malý | 24 | Vzorkovacia frekvencia | 4 | Subchunk "fmt" |
| malý | 28 | Bajtová rýchlosť | 4 | |
| malý | 32 | Zarovnanie | 2 | |
| malý | 34 | Počet bitov na vzorku | 2 | |
| veľký | 38 | Subchunk2 ID | 4 | |
| malý | 40 | Veľkosť subchunku2 | 4 | Subchunk "data" |
| malý | 44 | Dáta | Veľkosť subchunku2 | |

***Obr. 18.*** *Štruktúra wave hlavičky*

## 4.3.2 MPEG-1

Štandard MPEG-1 predstavuje flexibilnú kódovaciu techniku, ktorá využíva viacero metód, napr. Subpásmové kódovanie, analýzu bankou filtrov, transformačné kódovanie, entropické kódovanie a psychoakustickú analýzu. Pracuje so vzorkovacími frekvenciami 32, 44,1 alebo 48 kHz so 16 bitmi/vzorka a výstupný dátový tok sa pohybuje od 32 do 192 kbit/s na jeden kanál. Štandard ponúka 4 režimy kódovania kanálov: mono, stereo, duálne mono a spojené stereo (iba vrstva III).

Architektúra štandardu obsahuje 3 vrstvy, ktoré sa líšia výpočtovou náročnosťou, oneskorením a kvalitou výstupu. Vrstvy I (mp1) a II (mp2) sú si podobné a líšia sa iba

v niekoľkých detailoch. Obe používajú rýchlu Fourierovu transformáciu (fast Fourier transform, FFT), avšak vrstva I využíva okno s veľkosťou 512 vzoriek, zatiaľ čo vrstva II používa 1024-vzorkové okno. Maximálne podporované rozlíšenie v subpásmovej kvantizácii je v prípade vrstvy I 15 bitov/vzorka a v prípade vrstvy II 16 bitov/vzorka. Hoci sa tieto rozdiely zdajú byť minimálne, ukázalo sa, že vrstva II poskytuje rovnakú či dokonca vyššiu kvalitu výstupu pri bitovom toku 128 kbit/s než vrstva I s bitovým tokom 192kbit/s na kanál.

Proces kompresie v oboch vrstvách I a II pracuje so vstupným PCM signálom, ktorý rozkladá na 32 subpásiem. Počas rozkladu sa vykoná FFT, ktorej výstup prejde psychoakustickou analýzou a určením jnd (just-noticeable distortion alebo práve pozorovateľné skreslenie), ktoré vyjadruje mieru zmien v signáli, ktoré sú ešte akceptovateľné bez pozorovateľného rozdielu oproti pôvodnému signálu.

V závislosti od prahu maskovania sa pre každé subpásmo stanoví najvhodnejší krok kvantovania tak, aby boli dodržané požadovaný dátový tok a úroveň maskovania. Výstup kodéra san a záver zakóduje Huffmanovým entropickým kódovaním.

Hoci MPEG-1 vrstva II poskytuje prijateľné výsledky, prevládajúcim formátom je MPEG-1 vrstva III, všeobecne známa svojou skratkou mp3. Vychádza z vrstiev I a II, pridáva však viaceré nové techniky, ktoré vedú k nižšiemu dátovému toku (okolo 64 kbit/s na kanál) pri zachovaní kvality svojich predchodcov [89].

### 4.3.3 MP3

MP3, ako bolo spomenuté vyššie, je založený na kompresnom algoritme MPEG-1. Štandardizovaný bol v roku 1991. Pri zachovaní vysokej kvality umožňuje zmenšiť veľkosť hudobných súborov v PCM kvalite približne na desatinu až dvanástinu.

Každý z doteraz uvedených algoritmov pracuje tak, že audio signál sa rozdelí na bloky a komprimuje spektrum v jednotlivých blokoch. Takéto spracovanie má niekoľko výhod, ako je nižšia výpočtová náročnosť, nižšie nároky na čas a na hardvér systému. Taktiež aj degradácia spektra, ktoré patrí príliš veľkému časovému intervalu, spôsobuje neželané efekty, ako napríklad rozmazávanie v čase.

MP3 kóduje navzájom nezávisle bloky dlhé približne 26 ms (38 blokov za sekundu). Z tohto dôvodu platí, že ak signál obsahuje zvuk ako napr. dlhý tón, ktorý sa vôbec nemení, tak MP3 nedokáže túto jeho vlastnosť využiť mimo rámec 26 ms. Algoritmus Ogg je na tom rovnako s tým rozdielom, že veľkosti blokov sú trochu iné ako u MP3.

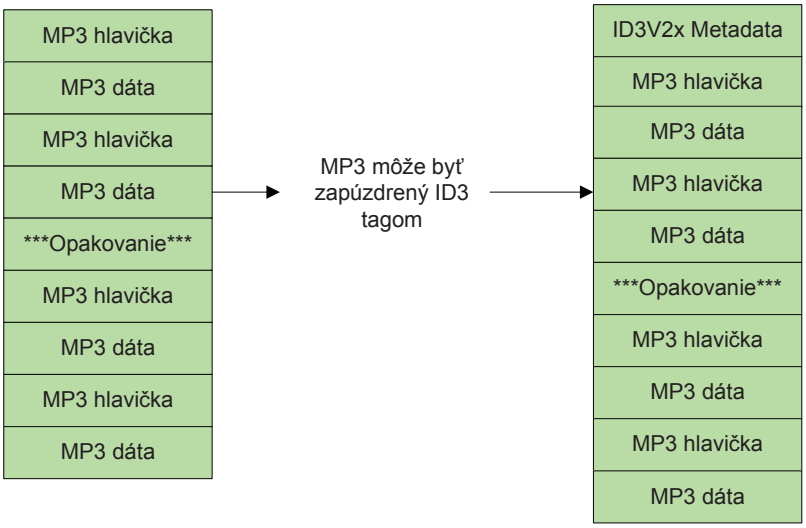
MPEG-1 štandard nemá však presnú špecifikáciu pre MP3 kodér. Preto je veľa rôznych MP3 kodérov, z ktorých každý produkuje výsledný súbor s inou kvalitou. Vo všeobecnosti môžeme povedať, že MP3 používa viacstupňovú hybridnú transformáciu (je tam aj MDCT). Na druhej strane, dekódovanie je v štandarde veľmi presne zadefinované. Väčšina dekodérov pracuje na princípe zvanom súhlasný bitstream (compliant bitstream). To znamená, že dekódovaný výstup z pôvodného MP3 súboru bude rovnaký (alebo

v povolenej hranici tolerancie chyby) ako výstup definovaný matematicky v štandarde ISO-IEC [24]. MP3 má štandardný formát, ktorého rámec obsahuje 384, 576 alebo 1156 vzoriek v závislosti od použitej verzie a vrstvy MPEG. Všetky rámce majú informácie uložené v hlavičke (32 bitov) a časť vyhradenú pre vedľajšie (side) informácie (9, 17 alebo 32 bytov v závislosti od použitej MPEG verzie a stereo-mono typu). Informácie uložené v hlavičke slúžia dekodéru pri dekódovaní dát, ktoré sú najčastejšie zakódované Huffmanovým kódom. Dekodéry sa porovnávajú z hľadiska zaťaženia pamäte a procesora.

Najjednoduchšie typy MP3 súborov používajú jednu bitovú rýchlosť (bit rate) pre celý súbor. Toto kódovanie je známe pod názvom Constant Bit Rate (CBR). Výhoda spočíva najmä v rýchlom a jednoduchom dekódovaní. Súbory, v ktorých sa bitová rýchlosť mení, sú zakódované pomocou Variable Bit Rate (VBR). Celá myšlienka spočíva v tom, že niektoré časti audia sme schopní zakódovať rýchlejšie a jednoduchšie (ako ticho alebo zvuk iba pár hudobných nástrojov) a iné zložitejšie. Celková kvalita súboru tak môže vzrásť použitím nižšej rýchlosti pre menej komplexné časti a vyššej pre komplexné časti. U niektorých kodérov je možné nastaviť požadovanú kvalitu a kodér vyberie vhodné rýchlosti [24].

V štandarde MPEG-1 je špecifikovaných niekoľko bitových rýchlostí: 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160, 192, 224, 256 a 320 kbit/s, a k nim prislúchajúce vzorkovacie frekvencie sú 32, 44,1 a 48 kHz. Najčastejšie používaná vzorkovacia frekvencia je 44,1 kHz, pretože sa používa pre audio CD, čo je hlavný zdroj pre vytváranie MP3. Najbežnejšia bitová rýchlosť je 128 kbit/s, pretože ponúka relatívne dobrú kvalitu a spotrebuje málo pamäte [24].

Šum je náhodná postupnosť a z in¬formatického hľadiska obsahuje najviac nikdy sa neopakujúcich informácií. Z tohto dôvodu je signál šumu pre algoritmus MP3 (aj Ogg) najnáročnejším prípadom pre účinnosť kompresie, pretože sa ho snaží kódovať exaktne.




***Obr. 19.*** *Štruktúra MP3 súboru*

MP3 sa skladá z viacerých MP3 rámcov, z ktorých každý sa skladá z hlavičky a dátovej časti (Obr. 19). Táto sekvencia rámcov sa nazýva elementárne vlákno. Rámce nie sú navzájom nezávislé, a preto nie je možné vybrať iba ľubovoľné rámce. Na detekciu začiatku rámca sa používajú synchronizačné slová. Hlavička MP3 súboru obsahuje tiež synchronizačné slovo, ako je vidieť na obrázku (Obr. 20). Za synchronizačným slovom nasleduje bit určujúci, že ide o MPEG štandard a dva bity, ktoré špecifikujú 3. vrstvu. Ďalšie hodnoty sú rozdielne podľa druhu MP3 súboru. ISO-IEC 11172-3 definuje rozsah hodnôt pre každý bit [24]. Väčšina MP3 súborov dnes obsahuje ID3 metadáta, ktoré nasledujú MP3 štruktúru, ako je to znázornené na obrázku (Obr. 19).

| Bit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Označenie | MP3 synchronizačné slovo | | | | | | | | | | | | Verzia | Vrstva | | Ochrana proti chybám | Bitová rýchlosť | | | | Frekvencia | | Pridaný bit | Privátny bit | Mód | | Rozšírený mód | | Kópia | | Originál | | Dôraz | |
| Hodnota | Synchronizačné slovo | | | | | | | | | | | | 1=MPEG | 01=Vrstva 3 | | 1=nie | 1010=160 | | | | 00=44100 Hz | | 0=rámec nie je pridaný | Neznáma | 01=Spojené stereo | | sterea vypnutý | stereo vypnuté | 0=bez práva na kopírovanie | | 0=kópia originálu | | 00=žiaden | |

***Obr. 20.*** *Hlavička MP3*

## 4.3.4 MPEG-2

Štandard MPEG-2 je formálnym pokračovateľom MPEG-1. Zahŕňa 2 režimy, jeden spätne kompatibilný s MPEG-1 (Backward Comptible, MPEG-2 BC) a druhý spätne nekompatibilný (Non-Backward Comptible, MPEG-2 NBC), ktorý prináša nové metódy a techniky kódovania.

Jedinými zmenami MPEG-2 BC oproti MPEG-1 je podpora pre nižšie vzorkovacie frekvencie (LSF) a viackanálové kódovanie podobné rozšíreniu mp3 surround. Formát MPEG-2 NBC sa označuje aj ako pokročilé kódovanie zvuku (Advanced Audio Coding, AAC) a je zostavený ako súprava nástrojov na efektívne kódovanie. Čím viac nástrojov sa použije, tým lepšia kompresia sa dosiahne, pričom kvalita výstupu zostane zachovaná. Cenou je však vyššia výpočtová náročnosť a oneskorenie. Na rozdiel od MPEG-1 formát MPEG-2 NBC nepoužíva na analýzu signálu hybridnú banku filtrov, ale iba MDCT v kombinácii s novými oknovými funkciami. Formát MPEG-2 sa stal súčasťou rodiny štandardov MPEG-4 [89].

## 4.3.5 MPEG-4 ACC

AAC obsahuje všetky techniky ktoré boli použité v MP3, ale obsahuje aj techniky nové. V porovnaní s MP3 dokáže o niečo lepšie využiť štruktúru signálu, ale aj ďalšie nedostatky ľudského sluchu, resp. kognitívne postačujúcej kvality. Okrem toho sa snaží aj odstraňovať niektoré nedostatky MP3, a to lepším spracovaním, alebo maskovaním

neželaných artefaktov spôsobených kvantizáciou spektrálnych koeficientov.

AAC tiež používa kódovanie v blokoch (rovnako ako aj MP3), ale umožňuje využívať vzájomné podobnosti sused¬ných blokov, a to tak, že pomocou LTP (Long Term Prediction) [21] algoritmus odhaduje tvar spektra práve kódovaného bloku z predošlého, už zakódovaného bloku. Zakóduje sa iba rozdiel medzi práve spracovávaným blokom a odhadom. Táto technika je veľmi účinná hlavne pri zvukoch ako sú sláčikové nástroje, píšťala, orchestrálne nástroje, dlho znejúce samohlásky vokálov.

ACC pri kódovaní zvuku využíva fakt, že človek vníma informácie zo štruktúry zvuku a nie priamo zo zvukovej vlny. Takže dva vygenerované šumy znejú pre človeka stále rovnako. Dôraz sa pri kódovaní kladie nie na samotný tvar zvukovej vlny šumového signálu, ale na jeho energetické rozdelenie v spektre, napr. ozvena hromu je šum v nízkych frekvenciách, šum vetra v stredných a zvuk činely je v prevažne vysokých. ACC pomocou PNS (Perceptual Noise Substitution) techniky deteguje vo zvuku zložku s charakteristikou šumu a odpočíta ju zo signálu. Šum sa nekóduje cez koeficienty spektra, ale sa zakóduje iba obálka jeho spektrálnej intenzity, podľa ktorej sa pri dekódovaní šum generuje v spektre [21], [22].

Ďalšia technika, ktorá je vyžívaná v ACC, je TNS (Temporal Noise Shaping). TNS je technika predikovania spektrálnych koeficientov z predošlých, teda na základe za¬kódovaných j koeficientov spektra odhaduje hodnotu pre koeficient j + 1. Kóduje sa iba rozdiel odhadu od skutočnej hodnoty. Keďže tieto kódované hodnoty sú už menšie, môže sa zvoliť jemnejšia kvantizácia, čím sa zmierni efekt rozmazávania zvuku (Pre/Post Echo) [22], [23].

AAC prináša podporu pre vzorkovacie frekvencie od 8 po 96 kHz, 1 až 48 audio kanálov plus 15 basových a 15 dátových kanálov s rozlíšením 8, 16, 24 alebo 32 bitov/vzorka. Formát AAC s nízkou zložitosťou (Low Complexity (LC) AAC) predstavuje pôvodný kodek MPEG-2 AAC a je vhodný na kódovanie reči pri dátovom toku 8 – 12 kbit/s. Formát AAC s vysokou efektivitou (High Efficiency (HE) AAC) prináša podporu technológie SBR (Spectral Band Replication) (verzia 1) a kanálového režimu parametrické stereo (verzia 2), ktorý je založený na profile spojené stereo štandardu MPEG-1 vrstva III. SBR je technológia vyvinutá spoločnosťou Coding Technologies [15], ktorá výrazne zlepšuje účinnosť perceptuálnej stratovej kompresie. Princíp spočíva v tom, že kóduje len dolnú polovicu celého frekvenčného pásma signálu. Horná polovica frekvenčného pásma sa „rekonštruuje“ z dolnej polovice pri dekódovaní signálu. Využíva sa tu tá vlastnosť, že polovice spektra sú v bežnom signáli korelované. Dolná polovica obsahuje väčšinu energie celého signálu, a je teda podstatne dôležitejšia ako horná polovica. V hornej polovici spektra väčšina audio signálov obsahuje už len harmonické frekvencie tónov z dolnej polovice. SBR teda na strane kódovania audio signálu nachádza dané vzťahy a pridáva ich kód (nájdené parametre zhody) k už zakódovanému dolnému pásmu. Rekonštrukcia horného pásma je v princípe analogická, ak poznáme prvých niekoľko členov číselnej postupnosti, tak vieme určiť, ako bude táto postupnosť pokračovať.

Existujú signály, ktoré do tohto modelu nezapadajú, preto má SBR ešte ďalšie prídavné techniky ako Sinusoidal Regeneration (SR) a Adaptive Noise Addition (ANA). SR pridáva do signálu tóny, ktoré sú iba v hornom pásme a nie sú harmonické k žiadnemu tónu v dolnom pásme. Tieto tóny nie je možné vyjadriť vzťahom, ale je potrebné ho do signálu jednoducho pridať. ANA slúži na podobný účel ako PNS v AAC. Ide o pridávanie generovaného šumu do hornej polovice spektra na strane dekodéra tak, aby doplnil energiu tohto spektra do energie, akú mal pôvodný kódovaný signál. Teda je nutné taktiež kódovať informácie o energii. SBR umožňuje zmenšenie veľkosti kódu na 70% oproti kódovaniu bez použitia SBR pri rovnakej kvalite.

PS (Parametric Stereo) je ďalšia technológia vyvinutá spoločnosťou Coding Technologies [16], ktorá sa zaoberá spracovaním stereo signálu. PS bola inšpirovaná myšlienkou parametrického kódovania (PK) audio signálu, ktoré je v súčasnosti iba v ranom štádiu. Principiálne je PS technika, ktorá zo stereo signálu vytvorí mono signál s niekoľkými prídavnými paramet¬rami navyše. Pri dekódovaní z tohto mono signálu a parametrov spätne poskladá (približne) rovnaký stereo signál ako bol pôvodný.

### 4.3.6 Ogg

Najčastejšie používaným kódovacím formátom v Ogg súboroch je vorbis. Je to otvorený, voľne šíriteľný a bezpatentový kompresný formát pre vysokokvalitné (8 kHz – 48 kHz, 16 bit, multi kanál) audio a hudbu s rovnakou alebo rôznou bitovou rýchlosťou (16 až 256 kbit/s/kanál). Toto postavilo vorbis na rovnakú úroveň ako audio reprezentované cez MPEG-4 štandard (ACC) a poskytuje vyšší výkon ako MP3. Dosahuje transparentnosť približne 150-170 kbit/s (-q5). Kodéry nie sú ešte veľmi optimalizované, takže v budúcnosti sa očakáva vylepšenie, teda lepšia kompresia.

Pri kódovaní sa používa viacnásobná veľkosť bloku na prepínanie okien vrátane prekrytia (vždy iba mocnina dvoch: 128/1024, 256/2048, 512/4096). Na kódovanie šumu sa používa MDCT, tóny sú kódovane pomocou rýchlej Fourierovej transformácie (Fast Fourier Transformation, FFT). Obálka (úroveň 1) sa určuje pomocou globálnych kriviek a lineárnej aproximácie rozdelenej spektrom (pre väčšie detaily). Levinson-Durbin LPC model v úrovni 0 sa už takmer nepoužíva, výnimkou sú niektoré staršie kódy [25].

Najväčší úspech mal vorbis vo video hrách (Epic Games‘ Unreal Tournament 2003 a Unreal Tournament 2004). Vorbis (Ogg) je tiež oficiálna časť OpenAL API (Application Programming Interface) knižnice, používanej v počítačových hrách [25]. V roku 2006 RAD Games integrovali vorbis do ich MSS (Miles Sound System) systému, ktorý bol použitý na celom svete vo viac ako 3200 hrách. Vorbis je veľmi vhodný pre internetové streamovanie.

### 4.3.7 Windows Media Audio

Windows Media Audio (WMA) je proprietárny kodek vytvorený spoločnosťou Microsoft ako odpoveď na licenčné obmedzenia formátu MP3. Existuje viacero verzií kodeku: WMA 9 je priamym súperom MP3 s podporou pre vzorkovacie frekvencie do 48 kHz so 16 bitmi/vzorka a výstupným dátovým tokom od 64 do 192 kbit/s, podporujúc CBR aj VBR.

Verzia WMA 10 Professional rozširuje možnosti kodeku v súboji s MPEG-4 AAC pridaním vzorkovacej frekvencie 96 kHz s 24 bitmi/vzorka pre 7.1 kanálov. Ak zariadenie nie je schopné prehrať 7.1 nahrávku, signál sa automaticky degraduje (vzorkovacia frekvencia, počet bitov na vzorku, zníženie počtu kanálov) na úroveň vhodnú pre zariadenie.

WMA 10 poskytuje tiež režim pre kompresiu reči nazývaný WMA 10 Voice, ktorý poskytuje bitový tok od 4 až do 20 kbit/s. Jeho zaujímavosťou je schopnosť dynamicky prepínať medzi hlasovou a štandardnou verziou kodeku, ak je spracúvaný signál príliš zložitý. Okrem toho WMA 10 poskytuje bezstratový režim, ktorý údajne dokáže zredukovať veľkosť pôvodného PCM signálu na polovicu až tretinu [89].

### 4.3.8 Porovnanie kompresných algoritmov

Je známe že o MP3 pri 128 kbit/s sa niekedy hovorilo, že jej kvalita sa vyrovná resp. blíži ku kvalite CD Audio. Ukázalo sa, že to nie je celkom pravda a že existujú skladby, ktoré môžu znieť pri 128 kbit/s dobre, ale ako záruka kvality táto bitová rýchlosť nestačí. MP3 sa dnes používa najčastejšie pri 192 kbit/s, ktorý už zaručuje veľmi dobrú kvalitu závislú už len od použitého kodéra.

Algoritmy Ogg a AAC dosahujú pri kompresii 112 kbit/s zhruba rovnaké výsledky, aj keď AAC v závislosti od kodéra možno o čosi lepšie. V oboch prípadoch možno povedať, že znejú lepšie ako MP3 pri 128 kbit/s a že AAC možno aj pri 96 kbit/s. Hlavný rozdiel je v tom, že ich neželané efekty sú pri nižšej bitovej rýchlosti (hlavne v prípade Ogg) iné ako pri MP3 a nepôsobia tak rušivo.

O „AAC Plus“ sa tvrdí, že kvalitu CD Audio zabezpečí už pri 48 kbit/s, a kvalitu blízku CD pri 32 kbit/s. Ako bolo spomenuté, SBR umožňuje ozajstné zacho¬vanie kvality len pri zmenšení na 70%, teda asi 64 kbit/s. SBR má aj svoje tienisté stránky, napr. pri niektorých skladbách totiž vyrába neželaný „krochkavý“ efekt vo zvuku, ktorý je spôsobený tým, že aj šumy z dolného pásma sa replikujú do horného a vzniká tak korelácia, ktorá by vo zvuku byť nemala.

Tomuto by sa pri MP3 Pro, a pri „AAC Plus“ dalo zabrániť, ak by bola technika PNS použitá na dekódovaný signál až po SBR. Toto by spôsobilo nárast výpočtovej zložitosti AAC Plus. AAC Plus v2 dosahuje podobné výsledky ako AAC Plus, ale zlepšuje kvalitu kódo¬vania stereo pri veľmi nízkych bitových rýchlostiach ako sú 32 kbit/s alebo 24 kbit/s. Pri rýchlosti 24 kbit/s, ktorú autori označujú za excelentnú, už zvuk síce znie mierne synteticky, ale nepôsobí nepríjemným dojmom.

## 4.4 Vyhodnotenie kódovacích algoritmov

Algoritmus kódovania reči sa hodnotí na základe prenosovej rýchlosti, kvality rekonštruovaného signálu, zložitosti algoritmu, uvádzaného oneskorenia a robustnosti algoritmu voči chybám v kanále a akustickej interferencii. Vo všeobecnosti platí, že vysokokvalitné kódovanie reči na nízkych prenosových rýchlostiach sa dosahuje využitím vysoko zložitých algoritmov.

Napríklad, implementácia hybridného algoritmu s nízkou prenosovou rýchlosťou v reálnom čase sa musí robiť na digitálnom signálovom procesore schopnom vykonávať 12 alebo viac miliónov inštrukcií za sekundu (MIPS). Jednosmerné oneskorenie (čiže len oneskorenie kódovania plus dekódovania) pri takomto algoritme je zvyčajne medzi 50-60 ms. Robustné systémy kódovania reči včleňujú aj algoritmy korekcie chýb, aby zabezpečili vnemovo dôležité informácie voči chybám v kanále. Navyše, v niektorých aplikáciách sa musia kodéry vyrovnať s rečou, ktorá je poškodená šumom v pozadí, nerečovými signálmi (ako napr. signály modemov) a s rozmanitosťou jazykov a prízvukov [102].

V digitálnych komunikáciách sa kvalita reči klasifikuje do štyroch všeobecných kategórií, konkrétne: rozhlasová, sieťová, komunikačná a syntetická. Rozhlasová širokopásmová reč zodpovedá vysokokvalitnej reči, ktorú môžeme vo všeobecnosti dosiahnuť pri rýchlostiach nad 64 kbit/s. Sieťová kvalita zodpovedá kvalite klasickej analógovej reči (300 – 3400 Hz) a môže byť dosiahnutá pri rýchlostiach nad 16 kbit/s. Pod komunikačnou kvalitou reči myslíme reč, ktorá znie prirodzene, je vysoko zrozumiteľná a primeraná telekomunikáciám. Syntetická reč je zvyčajne zrozumiteľná, ale môže byť neprirodzená a spojená so stratou schopnosti rozpoznať rečníka. Komunikačná kvalita reči môže byť dosiahnutá pri rýchlostiach nad 4,8 kbit/s. Rečové kodéry s rýchlosťou pod 4,0 kbit/s majú tendenciu produkovať reč syntetickej kvality.

Najviac používaným objektívnym kritériom kvality na meranie výkonu algoritmu je SNR (signal-to-noise ratio), ktorý je daný ako:

$$SNR = 10\log\left[\frac{\sum_{n=0}^{M} s^2(n)}{\sum_{n=0}^{M}\left(s(n)-\hat{s}(n)\right)^2}\right] \tag{30}$$

kde $s(n)$ sú dáta originálnej reči a $\hat{s}(n)$ sú kódované dáta. SNR je dlhodobá miera na presnosť rekonštrukcie reči, preto môže skryť prechodný rekonštrukčný šum, najmä pre nízkoúrovňové signály. Prechodné zmeny výkonu môžu byť lepšie detegované a vyhodnotené pomocou krátkodobého SNR, napr. výpočtom SNR pre každý segment reči dĺžky N vzoriek. Miera výkonnosti, ktorá odhalí slabú výkonnosť algoritmu, je segmentálna SNR (SEGSNR) daná ako

$$SEGSNR = \frac{10}{L}\sum_{i=0}^{L}\log\left[\frac{\sum_{n=0}^{N-1} s^2(i.N+n)}{\sum_{n=0}^{N-1}\left(s(i.N+n)-\hat{s}(i.N+n)\right)^2}\right] \tag{31}$$

Pretože výpočet priemeru sa robí až po logaritme, SEGSNR sa využíva pre kodéry, ktorých výkonnosť je premenlivá. Iné objektívne kritériá kvality sú artikulačný index, spektrálna vzdialenosť (log spectral distance) a Euklidovská vzdialenosť. Okrem týchto hodnotení sú potrebné aj subjektívne hodnotenia ako napríklad DRT (Diagnostic Rhyme Test), DAM (Diagnostic Acceptability Measure) a MOS (Mean Opinion Score). Sú založené na známkovaní poslucháčov. DRT je miera zrozumiteľnosti, kde je úlohou subjektu rozlíšiť jedno z dvoch možných slov v množine rýmujúcich sa párov. DAM body sú založené na výsledkoch testovacích metód, ktoré hodnotia kvalitu komunikačného systému podľa prijateľnosti reči, ako je vnímaná trénovaným poslucháčom.

MOS (Tab. 3) je miera, ktorá sa všeobecne používa na vyčíslenie kvality kódovanej reči. MOS zvyčajne vyžaduje 12 až 24 poslucháčov, ktorí sú inštruovaní, aby ohodnotili foneticky vyvážené vety päťúrovňovou stupnicou kvality (od 1-zlá až po 5-vynikajúca). Vynikajúca kvalita reči znamená, že kódovaná reč je nerozlíšiteľná od originálnej a je bez poznateľného šumu. Na druhej strane, zlá (neakceptovateľná) kvalita znamená prítomnosť extrémne obťažujúceho šumu a umelých signálov v kódovanej reči. Pri MOS testoch dostaneme klasifikáciu pomocou priemeru číselných bodov cez niekoľko stoviek rečových záznamov. MOS rozsah zodpovedá kvalite reči nasledovne: MOS v hodnotách od 4 do 4,5 znamená sieťovú kvalitu, body medzi 3,5 a 4 znamenajú komunikačnú kvalitu a MOS medzi 2,5 a 3,5 znamenajú syntetickú kvalitu. MOS hodnotenie sa však môže významne líšiť, a preto nie je absolútnym kritériom pri porovnávaní dvoch rozličných kodérov. Formálne subjektívne hodnotenia môžu byť zdĺhavé a veľmi nákladné. Preto vznikli miery kvality alebo zrozumiteľnosti, ktoré sa snažia automaticky predikovať subjektívnu kvalitu reči [3].

| MOS stupnica | Kvalita reči |
|---|---|
| 1 | Zlá |
| 2 | Slabá |
| 3 | Slušná |
| 4 | Dobrá |
| 5 | Vynikajúca |

***Tab. 3.*** *MOS miera*

Miera zrozumiteľnosti popisuje, ako je reč pochopiteľná a zrozumiteľná. Do úvahy sa pritom berú rôzne vlastnosti, napríklad hlasnosť reči, nelineárne skreslenie, úroveň

šumu na pozadí, ozveny a dozvuky a mnohé iné. Na porovnanie zrozumiteľnosti existujú dve hlavné stupnice: index prenosu reči (Speech Transition Index, STI) a všeobecná škála zrozumiteľnosti (Common Intelligibility Scale, CIS), ktoré majú rozsah od 0 (najhoršie) po 1 (najlepšie), resp. 0% až 100%. Vo všeobecnosti sa reč považuje za zrozumiteľnú, ak algoritmus dosiahne na stupnici skóre aspoň 0,5 (alebo 50%) [89].

Výsledkom snahy nahradiť subjektívne testy objektívnymi je algoritmus PEAQ (Perceived Evaluation of Audio Quality, ITU-R BS.1387-1) [93]. Metóda PEAQ používa pravdepodne najpresnejší a najdetailnejší percepčný model aký sa dnes používa. PEAQ určuje mieru rozdielu v kvalite dvoch signálov (referenčného a testovaného signálu). Metóda je vhodný najmä pre moderné nelineárne a nestacionárne kodeky. Výstupným parametrom z PEAQ je objektívny stupeň rozdielu (Objective Difference Grade, ODG) a index skreslenia (distortion index DI). ODG sa uvádza s presnosťou na jedno desatinné miesto. DI a ODG majú rovnaký význam, avšak porovnávať ich možno len kvantitatívne a nie kvalitatívne. Podľa všeobecného pravidla, ODG sa používa na hodnotenie kvality pre hodnoty väčšie ako je -3,6. Ak je ODG hodnota menšia, odporúča sa použiť DI [90].

Do PEAQ modelu vstupujú dva signály: referenčný audio signál a testovaný audio signál. Proces merania je aplikovateľný aj na analógové aj na digitálne signály. PEAQ je založený na všeobecne akceptovaných psycho-akustických princípoch. Súbežné rámce oboch signálov sú transformované do výstupu sluchového modelu. V nasledujúcom kroku algoritmus modeluje akustické skreslenie prítomné v testovanom signáli a porovnáva ho so sluchovým modelom. Informácie získané týmto procesom sú obsiahnuté v niekoľkých hodnotách, nazývaných MOVs (Model output variables, výstupné premenné modelu). Cieľom je získať jediné číslo opisujúce počuteľné skreslenie prítomné v testovanom signáli. Základom pre výpočet tejto hodnoty je MOVs. PEAQ algoritmus používa na výpočet neurónové siete [90].

PEAQ reprezentuje veľmi dôležitý zdroj v objektívnom hodnotení audio kodekov. Porovnanie štyroch najbežnejších stratových kodekov pomocou PEAQ (mp3, AAC, OGG Vorbis a WMA 9.1) je nasledovné: kodeky boli testované pri 4 štandardných bitových rýchlostiach pri rovnakej vzorkovacej frekvencii 48 kHz. Pre každú bitovú rýchlosť bola určená ODG hodnota.

Všetky kodeky sa správajú podobne pri vyšších bitových rýchlostiach (≥ 192 kbit/s), rozdiely sú minimálne s výnimkou AAC. Pri nižších bitových rýchlostiach (≤ 128 kbit/s) (ktoré sa používanú najmä v internetovom audiu) sa kodeky správajú rôzne. Najlepší je OGG, WMA je na druhom mieste a prekvapivo najhoršie výsledky má mp3. Z porovnania vidíme, že vybrať správny kodek je dôležité pri nižších bitových rýchlostiach, ale pri vyšších to už z pohľadu audio kvality také dôležité nie je [90].

Medzi objektívne metódy patrí aj algoritmus PESQ (Perceptual Evaluation of Speech Quality, ITU-T P. 862) [92]. Princíp porovnávania je veľmi podobný metóde PEAQ: základom sú referenčný signál a testovaný (syntetizovaný) signál. Posledná verzia algoritmu PESQ však zahŕňa aj kompenzáciu oneskorení, čo znamená, že referenčné a testované signály nemusia byť časovo zarovnané. Táto vlasnosť sa s výhodou používa,

najmä ak k degradácii referenčného signálu prichádza pri prenose IP sieťou, ktorá takýto časový posun môže spôsobiť [65]. Vďaka tejto kompenzácii časových oneskorení sa tiež zdá, že by PESQ mohol korektne vyhodnocovať aj umelú reč. Metódy PEAQ a PESQ patria medzi state-of-the-art techniky pre vyhodnotenie vnímanej kvality reči a hudby. Nevýhodou PESQ je, že nie je vhodný pre aplikácie v reálnom čase. V takomto prípade sa odporúča použiť jeho predchodcu, metódu PSQM (ITU-T P. 861).

| **kodek/bitová rýchlosť** | **64 kbit/s** | **128 kbit/s** | **192 kbit/s** | **256 kbit/s** |
|---|---|---|---|---|
| MP3 | -3,506 | -1,179 | -0,146 | -0,038 |
| AAC | -3,183 | -0,975 | -0,255 | -0,243 |
| Ogg | -2,411 | -0,485 | -0,145 | 0,125 |
| WMA | -2,774 | -0,661 | -0,163 | -0,120 |

***Tab. 4.*** *ODG hodnoty pre rozdielne kodeky*

Existuje aj niekoľko štandardizovaných metód na predikciu zrozumiteľnosti reči. Patria medzi ne hlavne SII (Speech Intelligibility Index) a STI (Speech Transmission Index). Algoritmy na vyhodnotenie kvality reči v telekomunikačných sieťach patria do ďalšej skupiny použiteľných objektívnych meraní.

Použitie spomínaných prístupov môžeme vo všeobecnosti charakterizovať nasledovne:

- SII na vyhodnotenie zrozumiteľnosti reči v prítomnosti stacionárneho šumu.
- STI na vyhodnotenie zrozumiteľnosti reči degradovanej nelineárnym skreslením.
- PESQ na vyhodnotenie kvality reči bez obmedzenia.

Objektívne merania SII a STI sa zvyčajne používajú na pôvodné rečové signály, nie na umelú reč. V korpusovej syntéze reči však na segmentálnej úrovni pracujeme so segmentami pôvodného rečového sinálu, takže tieto metódy sú vhodné minimálne na predikciu segmentálnej zrozumiteľnosti umelej reči. SII a STI požadujú na vstupe oddelený rečový a šumový signál, algoritmus PESQ potrebuje na vstupe referenčný rečový signál a zašumený rečový signál [65].

# 5 Skvalitnenie syntézy a princíp učenia v procese syntézy reči

Pod požiadavkou prirodzenosti syntetickej reči sa chápe jej nerozlíšiteľnosť od reči ľudskej. Reč je vnímaná ako prirodzená vtedy, keď v kontexte, v ktorom sa nachádza, nie je možné určiť, či jej zdrojom bol človek alebo počítač. Práve na dosiahnutie prirodzenosti sa v posledných rokoch začala upriamovať najväčšia pozornosť v oblasti výskumu syntézy reči. Ako hlavná príčina umelosti syntetickej reči bola označená silná parametrizácia (napr. len 3 formanty) generovaná na základe manuálnych empirických nastavení (syntéza podľa pravidiel). Tým sa dospelo k presvedčeniu, že potlačenie neprirodzenosti môže priniesť používanie rečových segmentov vyseknutých z pôvodných rečových nahrávok, ktoré sa pri syntéze pospájajú do požadovaného poradia. Kvalita syntézy však nezávisí len od veľkosti segmentov, resp. počtu spojení v syntetickej reči, ale aj od charakteru týchto spojení. Pri syntéze totiž dochádza k spájaniu segmentov z rôznych kontextov, a tieto viac či menej ovplyvňujú prinajmenšom okraje segmentov a vytvárajú tak neprirodzené spoje pre iné kontexty.

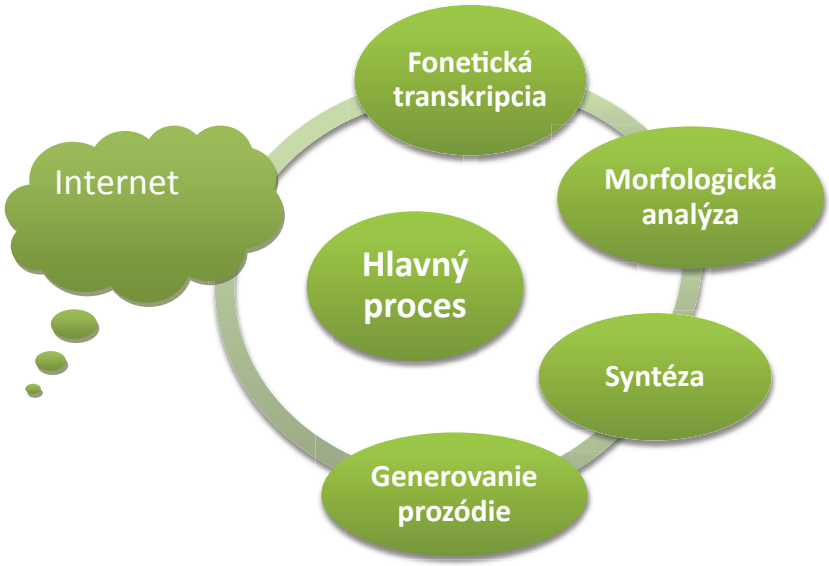
V posledných desaťročiach sa začal výskum venovať vplyvu prozódie na spracovanie reči (Allen 1992 [94], Boite 2000 [95], Eefting a Notebook 1993 [96], Larreur at al. 1989 [97]). Pri rozpoznaní reči je prozódia dôležitá, pretože poskytuje informáciu k segmentácii vyhovorení. Prozódia je významná aj pre syntézu reči, nakoľko je jedným z hlavných kritérií pre prirodzenosť syntetizovanej reči.

Okrem toho je potrebné, aby syntetizátor správne vyslovoval slová, vedel čítať skratky, číslovky a skloňovať (musia byť zohľadnené nepravidelnosti a výnimky vo výslovnosti v danom jazyku) a mal správnu prozódiu. Ako zdokonaliť syntetizátor, aby ovládal všetky spomenuté veci, je popísané v nasledujúcich kapitolách. Návrh je konkretizovaný a vysvetlený pre slovenský jazyk.

## 5.1 Inteligentný systém pre syntézu reči

Na pracovisku autorov (STU Bratislava) sa vyvinulo a vyvíja niekoľko rečových syntetizátorov. Syntetizátory pracujú nad rôznymi databázami a fungujú na rôznych platformách: desktopová, mobilná alebo webová. Sú využívané na rôzne účely. Podstatou každého syntetizátora je vyberanie rečových jednotiek z databázy a ich spájanie do zadaného výstupu (jadro tvorí konkatenačná syntéza). Ak chceme mať výstup čo najprirodzenejší, tak za týmto jednoduchým popisom musíme vidieť aj ďalšie procesy, ktoré

prispievajú k dokonalejšiemu výsledku. Jednotlivé procesy je najlepšie implementovať ako nezávislé samostatné celky a v prípade potreby zapojiť do syntézy iba tie, ktoré sú momentálne potrebné. Tento modulárny prístup navyše ponúka možnosť samostatného vývoja jednotlivých častí. Základný princíp modulárneho syntetizátora je znázornený na obrázku (Obr. 21).




***Obr. 21.*** *Modulový návrh syntetizátora*

V tomto príklade pristupuje používateľ k syntetizátoru cez web rozhranie. V aplikácii zadá text štandardným spôsobom. Do hlavného procesu, ktorý môže bežať na ľubovoľnom serveri, sa odošle požiadavka na syntézu. Hlavný proces riadi komunikáciu s ostatnými modulmi syntetizátora, v ktorých prebieha analýza, predspracovanie a samotná syntéza. Moduly musia dodržiavať stanovené formáty vstupného a výstupného protokolu na komunikáciu. Musia byť teda stanovené presné formáty vstupu a výstupu každého modulu. V našom prípade sa na komunikáciu jednotlivých modulov používa XML formát, a to z viacerých dôvodov:

- prehľadný a ľahko kontrolovateľný zápis vymieňaných informácii
- široká podpora zo strany knižníc vývojových nástrojov
- jednoduché dopĺňanie nových dát
- ľahká a rýchla práca s XML súbormi

XML je multiplatformový jazyk slúžiaci na opis štruktúrovaných údajových elementov. Štruktúra XML dokumentu umožňuje jednoduché a rýchle medziplatformové vyhľadávanie, flexibilné pridávanie nových parametrov a informácii. [82]. V moduloch syntetizátora slúži ako médium, ktoré nesie informácie a meta znaky prvkov predspracovania syntézy reči. Každý element v XML formáte (napr. `sentence`, `word`, `val`) má

presný význam. Nie všetky tagy sú povinné. Jednotlivé moduly pracujú s informáciami ľubovoľne, ale tak, aby dodržali dohodnuté pravidlá. Moduly môžu doplniť nové údaje, modifikovať existujúce údaje, v prípade chyby ich opraviť alebo aj zmazať. Na začiatku obsahuje XML dokument iba vstupný text, zadaný z internetového rozhrania. Na konci procesu obsahuje XML dáta zo všetkých modulov. Syntetizátor využije pri syntéze všetky údaje, ktoré sú k dispozícii.

K celkovému zlepšeniu syntetizovanej reči prispeje aj to, ak sa syntetizátor bude vedieť neustálym používaním zdokonaľovať a učiť sa nové pravidlá (napr. výnimky pri fonetickej transkripcii, nové skratky a pod.). Takúto učiacu sa funkcionalitu je nutné implementovať do rozhrania samotného syntetizátora. K procesu výučby sú potrební najmä používatelia, ktorí chybné slová alebo vety opravia. Je viacero možností, ako proces výučby môže prebiehať, od tých najjednoduchších, ako je ručná oprava, až po tie najzložitejšie, akými je akustická oprava. Analýza jednotlivých možností, ich výhody a nevýhody sú rozobraté nižšie.

- **Ručný prepis.** Rozhranie syntetizátora umožní používateľovi ručne prepísať chybný text (napr. nesprávne vyslovené slovo, zle rozvitú skratku alebo nesprávne vyskloňovaný text). Po prepise si používateľ môže znovu vypočuť syntetizovaný text už s opraveným slovom (alebo viacerými slovami).
- **Výber slova s nesprávnou syntézou** s ponukou iných možností. Rozhranie syntetizátora ponúkne používateľovi (napr. po kliknutí na nesprávne slovo) všetky možnosti, ako môže syntetizovaný text znieť. Menu môže obsahovať rôzne tvary pre fonetickú transkripciu, rôzne modely pre prozódiu vety a pod. Používateľ si vyberie z ponúkaných možností správnu a môže si znovu vypočuť syntetizovaný text už s opraveným textom. Ak ani jedna z ponúkaných možností nevyhovuje, mala by byť možnosť ručnej opravy textu. Na základe ručnej opravy textu je možné rozšíriť zoznam alternatívnych možnosti.
- **Automatické/inteligentné prehľadávanie** iných možnosti. Táto možnosť vychádza z predošlej možnosti, je však oveľa inteligentnejšia. Ponúkané možnosti by nemali byť zoradené náhodne, ale podľa pravdepodobnosti výskytu a aj určitej logiky. Rozhranie syntetizátora ponúkne používateľovi (napr. po kliknutí na nesprávne slovo) všetky možnosti s tým, že na začiatku sú alternatívy s vyššou pravdepodobnosťou a posledné sú najmenej pravdepodobné možnosti. Obsah zoznamu sa líši aj v závislosti od typu opravy. Pri označení vety sa jedná a zmenu prozódie vety, označenie slova môže znamenať prepis jeho fonetickej transkripcie alebo skloňovania. Takéto vytvorenie zoznamu si však môže vyžiadať viac času.
- **Akustická oprava výslovnosti.** Tento spôsob opravy si vyžaduje dokonalý rečový rozpoznávač. Používateľ po vypočutí zosyntetizovaného textu opraví text tak, že ich sám prečíta. Môže prečítať aj celú vetu, v ktorej sa chybné slovo alebo slová nachádzajú. Rozpoznávač vetu zapíše a syntetizátor si porovná, ktoré slovo alebo slová boli nesprávne, a nahradí ich. V prípade prozódie sa porovná typ vety. Používateľ si môže znovu vypočuť syntetizovaný text už s opraveným slovom

a v prípade chyby sa proces zopakuje.

- **Oprava rovnakých slov** (alebo rovnakého kmeňa slov). Pri syntéze dlhšieho textu sa dá predpokladať, že sa niektoré slová vyskytujú v texte viacnásobne. Preto v prípade akejkoľvek opravy sa prehľadá syntetizovaný text a ak sa nájde opravované slovo/slová/časť slova, označí sa (farebne vysvieti alebo podčiarkne). Používateľ bude môcť prejsť označenými slovami a rozhodne, či opraviť niektoré alebo všetky označené slová (implementuje sa funkcia *prepíš slovo* alebo *prepíš všetky slová*). Takáto funkcionalita by významne urýchlila opravu slov, keďže používateľ by nemusel sám prehľadávať všetky slová a vyberať správne možnosti.

Celý proces opravy by nebol veľmi užitočný, ak by sa museli vždy opravovať tie isté slová a syntetizátor by sa „neučil" opravené verzie textov. Je potrebné nájsť optimálny spôsob učenia. Proces učenia môžeme zhrnúť do dvoch základných bodov:

- ukladanie výnimiek (opravených slov) do slovníka
- štatistické vyhodnotenie zmien výslovnosti z pohľadu času a pretrénovanie pravidiel

Najdôležitejšie je nájsť optimálnu hranicu pre vytváranie nových pravidiel a ponechávanie slov v slovníku výnimiek. Jednou z možností vytvárania pravidiel a triedenia slov do slovníka je manuálny proces. Tento proces je síce veľmi pomalý, ale bol by presný. Chybovosti sa dá predísť aj tým, že by pravidlá nezávisle od seba vytvárali viacerí ľudia. Ich čiastkové pravidlá porovnáme a vybrieme výsledné pravidlá a konečný slovník výnimiek. Druhou možnosťou je tento proces zautomatizovať, čo povedie k vyššej efektivite. Jadrom takéhoto riešenia je algoritmus na prehľadávanie databázy. Proces vytvárania nových pravidiel je nutné v určitých časových intervaloch opakovať.

Podrobnejšie sú navrhované možnosti rozobraté v kapitolách 5.3.2, 5.3.5, 5.4.1, 5.4.7, 5.5.1, 5.5.7, 5.6.2, 5.6.4 pre každý modul syntetizátora.

## 5.2 Spracovanie prirodzeného textu

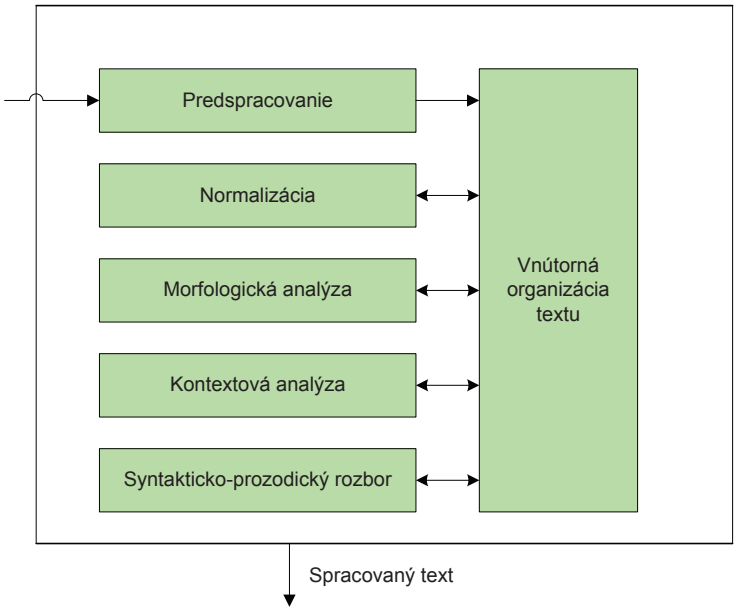
Pre človeka je prirodzené, že dokáže jednoducho pochopiť text z novín, e mail, recept v novinách. V ľudskom mozgu prebieha veľa procesov, ktoré sú zložité. Pri snahe aplikovať tieto procesy do syntézy reči a dosiahnuť tak ideálny výstup narazíme na množstvo problémov. Napriek tomu je potrebné, aby sa aspoň časť týchto procesov implementovala pri syntéze. Okrem samotného procesu syntézy ku kvalite reči prispieva aj proces predspracovania textu.

Ako bolo v teoretickej časti spomenuté, TTS systému je na vstupe zadaná monotónna informácia a na výstupe očakávame, že bude systém generovať prirodzenú reč

s prirodzeným prednesom bez toho, aby musel poslucháč venovať počúvaniu zvýšenú pozornosť. Tento proces je teda podstatne zložitejší ako obyčajná syntéza. Pre prirodzene znejúcu syntetizovanú reč na výstupe by mal syntetizátor obsahovať tri základné moduly, a to morfologicko-syntaktickú analýzu, fonetickú transkripciu a generátor prozódie.

### 5.2.1 Morfologicko-syntaktická analýza

Predspracovanie textu, inak povedané proces vyššej syntézy alebo morfologicko-syntaktická analýza, je súbor nezávislých, paralelných procesov, ktorými prechádza vstupný text pred samotnou syntézou (Obr. 22). Procesy môžu byť volané v ľubovoľnom poradí. Každý komponent procesu pridáva do výsledného textu prídavnú informáciu. Na texte môže v danom čase pracovať preto vždy len jeden komponent. Po skončení úlohy a zapísaní nových údajov sa môže vykonávať ďalší proces. Komponenty môžu využívať pridané informácie. Z tohto dôvodu je dôležité riadiť komunikáciu procesov a ich prístup.




***Obr. 22.*** *Bloková schéma morfologicko-syntaktickej analýzy*

V prvej fáze dochádza k načítaniu textu, tokenizácii. To zabezpečuje rozdelenie textu do logických celkov: kapitol, odsekov, viet, a nakoniec určuje hranice jednotlivých slov. Zabezpečuje odstránenie niektorých nečitateľných znakov, prázdnych medzier a pod. Je to predprípravа textu pre ďalšie operácie. Do tejto fázy sa môže zaradiť aj detekcia textu (napr. zoradenie čítania webovej stránky alebo spôsob čítania tabuľky). Procesom tokenizácie, ktorý je v tomto prípade rovnako dôležitý ako pri strojovom prepise, sa vytvára

databáza skratiek. Tá sa neskôr používa pri normalizácii textu. Skratky, ktoré sa končia bodkou, zohrávajú hlavnú úlohu v ďalšom kroku spracovania textu – segmentácii viet. Ak sa vo vete nachádza skratka, za ktorou nasleduje vlastné podstatné meno, môže indikovať koniec vety. Program musí vyhodnotiť túto možnosť a určiť, či daný reťazec predstavuje skratku. Ak sa reťazec nachádza v databáze, nevykoná sa segmentácia vety, ak sa nenachádza, veta sa rozdelí podľa základného pravidla [82].

Proces normalizácie vykonáva prepis do plnej slovnej formy so správnou gramatickou kategóriou. Tento proces musí poskytnúť informácie o tom, ako prečítať správne čísla (1, 2, 3. na „jedna, dva, tretí“), dátumy (1. 8. 2010 na „prvého ôsmy dvetisícdesať“), časové údaje (12:00 na „dvanásť nula nula“), peniaze (€ 200 na „dvesto EUR“), skratky (Bc. na „bakalár“), akronymy (USA na „Spojené štáty americké“) a pod. Normalizácia a slovenské znenie výrazov sú o niečo zložitejšie, a to z dôvodu flektívneho charakteru jazyka a nedostatku dostupných a účinných jazykových nástrojov pre analýzu, ako sú morfologické a syntaktické analyzátory [82].

Tretia fáza určuje slovné druhy jednotlivých slov vo vete a ich gramatické kategórie. Rozdeľuje slová na ohybné a neohybné. Pri ohybných sa určujú rôzne gramatické kategórie, ako je rod, číslo, pád a vzor. Takáto informácia pomôže v iných častiach predspracovania textu, ako sú fonetická transkripcia a určovanie prozódie. Z tohto dôvodu sa jej budeme venovať podrobnejšie v kapitole 5.4. Pridávanie morfologickej informácie k slovám sa nazýva značkovanie (tagovanie). Určovanie slovných druhov je komplexný problém a jeho riešenie nie je triviálne. Neohybných slovných druhov je menej ako ohybných, preto sa určujú podľa slovníka. Ohybné slovné druhy sa vyskytujú v rôznych tvaroch (rôzny rod, číslo, pád, čas) a môžu sa k nim pridávať rôzne prípony, predpony, vytvárajú sa rôzne zdrobneniny a negácie. Prehľadávanie slovníka, ktorý by obsahoval všetky tvary slov, by bolo náročne a veľmi ťažko by sa dalo stihnúť v reálnom čase [83].

Kontextová analýza na jednotlivé slová pozerá v kontexte okolitých slov a napomáha tak k určeniu morfologického charakteru slov a ich gramatických kategórií. V predošlom kroku sa získali všetky najpravdepodobnejšie vlastnosti slova, v tomto bloku sa na základe okolitých slov upresňujú a získavajú nové informácie (napr. slovo „perie“ môže byť slovesom, ale aj podstatným menom). Použitím týchto metód sa redukuje množstvo nejednoznačností. Existujú dva prístupy kontextovej analýzy: pravdepodobnostná a deterministická. Pravdepodobnostná metóda je založená na prechodových pravdepodobnostiach medzi susediacimi slovnými druhmi a gramatickými kategóriami dvoch slov. V súčasnosti je trend používať slovníky založené na stochastických metódach. Ich hlavnou výhodou je jednoduché zabudovanie do systému a ľahké získanie štatistík [74]. Najznámejšie príklady využitia štatistickej metódy sú modely založené na skrytých Markovových modeloch (HMM) nazývané Maximum Likelihood Tagger alebo Markov Model Tagger. V dnešnej dobe sa používa aj N-gram. Deterministické metódy sú založené na pravidlách (typ áno/nie). Pravidlá sú vlastne hypotézy o slovnom druhu alebo gramatickej kategórii. Podľa vyhodnotenia pravidla sa vlastnosť slova zmení alebo nie. Príkladom takejto techniky je Brillov značkovač. Obe z týchto metód potrebujú množstvo manuálne

označkovaných slov na trénovanie; pri prvej metóde na vytvorenie štatistiky a pri druhej metóde na vytvorenie pravidiel [83].

V poslednej fáze sa vykonáva syntakticko-prozodický rozbor. Tento proces definuje frázy. Frázy pomáhajú určovať prozodické vlastnosti vety ako je melódia, intonácia, intenzita a časovanie jednotlivých slabík. Poznáme tri základné skupiny na syntakticko-prozodický rozbor, a to ručne odvodené heuristiky, metódy používajúce gramatiky a korpusovo orientované metódy (modelovanie hraníc fráz sa vykonáva prostredníctvom Markovových reťazcov alebo CART techník. Prozodická informácia prispieva k osobitému charakteru vety, robí reč prirodzenou a jednoduchšou na pochopenie pre poslucháča. Prozódia hrá dôležitú úlohu nielen po umeleckej stránke reči, ale aj v bežnej hovorovej (napríklad v thajskom jazyku zmena dôrazu znamená úplne iné slovo).

### 5.2.2 Fonetická transkripcia

Pri fonetickej transkripcii sa prepisuje písaný text do podoby, v akej sa vyslovuje. Vstupný text na syntézu sa konvertuje na fonémy, ktoré sú zapísané pomocou ASCII znakov. Po tomto prepise môže nasledovať prozodická analýza a vkladanie informácií o dobe trvania, vrcholoch a pod. Tento proces môže byť ručný alebo automatický. Ručný prepis je veľmi zdĺhavý a náročný proces. Zaručená je však menšia chybovosť a možnosť opravy po vypočutí nasyntetizovaného textu. Pri automatickej transkripcii môže byť prepis spravený podľa pravidiel alebo slovníka. Samostatnou možnosťou transkripcie je natrénovanie pomocou neurónových sietí alebo skrytých Markovových modelov. Vstupom trénovania by bola skupina viet spolu s ich fonetickým prepisom. Po natrénovaní by sa dal tento systém použiť aj pre nové dáta. Nevýhodou tejto možnosti je však veľká vstupná množina na trénovanie (niekoľko tisíc až milión viet).

### 5.2.3 Generovanie prozódie

Prozódia (slovo gréckeho pôvodu) sa zaoberá vlastnosťami reči ako sú rytmus, intonácia, výška tónu. Opisuje všetky akustické vlastnosti reči, ktorých doménou nie je jednoduchý fonetický segment, ale väčšie jednotky skladajúce sa z viacerých segmentov, napríklad celé vety. Preto sa prozodické javy nazývajú supra-segmentálne. Tieto javy vnímame ako dôraz, akcent alebo rôzne modifikácie intonácie, rytmu a hlasitosti.

Hovorená reč okrem informácie nesie v sebe aj spôsob, ako má byť informácia prečítaná. Tejto informácii hovoríme prozodická. Prozódia robí syntetizovanú reč viac prirodzenou a počúvateľnou. Najdôležitejšie vlastnosti prozódie sú:

- **Intonácia** (melódia) – zaznamenáva zmenu frekvencie základného hlasivkového tónu ($F_0$).
- **Časovanie** – podáva informácie o členení súvislej vety (o pauzách), o rytme

a tempe reči.
- **Intenzita** (hlasitosť) – určuje energiu reči a teda jej hlasitosť.

#### 5.2.3.1 Intonácia

Intonácia alebo melódia reči súvisí s priebehom základnej hlasivkovej frekvencie $F_0$. Generovanie intonácie prebieha v dvoch krokoch. Prvým krokom je vytvorenie symbolického popisu intonácie ako doplnku k ortografickej transkripcii reči. Medzi najznámejšie symbolické popisy patrí:

- **ToBI** (Tones and Break Indices). Tento transkripčný mechanizmus popisuje prízvuk a rozlišuje hranice medzi vyslovovanými úsekmi (frázami a pod.). Používa rôzne symboly, napr.: *L* na pokles priebehu kontúry $F_0$, *H* na nárast, * na označenie prízvučnej slabiky, - na označenie neukončenej prozodickej frázy, nezvyčajne vysoká výška hlasu na začiatku frázy sa označí %*H*, atď. ToBI rozlišuje 5 typov hraníc medzi prozodickými frázami s hodnotami od 0 po 4 (kde hodnota 4 je najsilnejšia, „úplná“ hranica sprevádzaná pauzou) [81].
- **INTSINT** (International Transcription System for Intonation). Táto metóda sa venuje opisu intonačnej krivky. Na popis sa zaviedli absolútne a relatívne tóny. Absolútne tóny sa vzťahujú na rozsah výšky hlasu: *T* maximálna, *M* stredná, *B* spodná. Relatívne tóny monitorujú výšku z pohľadu susedných tónov, teda *H*, ak je aktuálny tón vyšší, *L* ak je nižší a *S* ak je rovnaký. Symbol *U* hovorí, že tón sa nachádza vo vzostupnej sekvencii tónov, alebo symbol *D*, ak je tón v klesavej sekvencii [81].
- **Tilt** – slúži na popísanie intonácie v parametrickom tvare, takže sada parametrov sa namiesto symbolom pripisuje jednotlivým častiam. Po sčítaní parametrov tohto modelu dostaneme výslednú intonáciu [81].

Ďalším krokom bude podľa výsledku z prvého kroku vygenerovať výslednú intonáciu (kontúru $F_0$). Modely kontúr delíme podľa výslednej oblasti, v ktorej sa s intonáciou pracuje, na:

- **akustické** – vychádzajú z akustickej reprezentácie intonácie pomocou priebehu základného hlasivkového tónu v čase. Medzi najznámejšie patrí Fujisakiho model, Tilt a tzv. modely akustickej štylizácie. Akustické modely sú z menovaných skupín najpoužívanejšie [81].
- **percepčné** – tieto modely sú založené na tom, ako intonáciu vnímajú poslucháči. Môžeme sem zaradiť modely percepčnej štylizácie a model IPO [81].
- **lingvistické** – tieto modely majú základ vo všeobecnej lingvistickej reprezentácii prozódie, a preto sú najzložitejšie. Najznámejšími sú teória kontúr výšky hlasu (angl. pitch contour theory) a teória postupnosti tónov (angl. tone sequence theory) [81].

Z hľadiska generovania hlasivkovej frekvencie delíme intonačné modely do týchto skupín:

- **generovanie podľa pravidiel** – najjednoduchší generátor. Vychádza sa z transkripčného popisu. Na základe týchto údajov sa navrhnú lingvistické pravidlá a vygeneruje sa priebeh základnej hlasivkovej frekvencie. Parametrami môžu byť napr.: rozsah výšky hlasu, typ vety (opytovacia, oznamovacia.), použitá interpunkcia, segmentácia vety na slovnej a fonetickej úrovni, trvanie jednotlivých hlások, atď. V tejto metóde sa detegujú aj javy vo vete ako je prízvuk na jednotlivých slovách, charakteristika priebehu základnej hlasivkovej frekvencie (stúpavá, klesavá.), odhad vetného prízvuku a pod. Zo všetkých zistených údajov a charakteristík sa interpoluje priebeh $F_0$. Tým sa predíde aj skokovým zmenám. Často sa využíva model ToBI [81].
- **parametrické modely intonácie** – pracujú s podobnými informáciami ako modely generujúce podľa pravidiel. Priebeh základnej hlasivkovej frekvencie a iné dôležité charakteristiky sú reprezentované pomocou vhodne zvoleného parametrického modelu. Patrí sem Fujisakiho model a model Tilt. Používajú sa i parametrické modely vychádzajúce zo symbolického popisu ToBI (parametrami bývajú 2 až 3 riadiace priamky udávajúce absolútnu vrchnú ($T$) a spodnú ($B$), prípadne referenčnú ($R$) úroveň výšky hlasu. Model sa dá natrénovať aj z reálnych dát [81].
- korpusovo orientované generovanie $F_0$ – stále častejšie sa na generovanie kontúr využívajú korpusy prirodzených rečových nahrávok (z nich sa automaticky nastavujú parametre generátora $F_0$). Intonačný model potom pracuje s databázou kontúr. Na vytváranie databáz sa používajú CART stromy, štatistické metódy (neurónové siete alebo HMM), alebo sa priamo využije označená reč, napr. pomocou ToBI. Najjednoduchším a efektívnym modelom je transplantovaná prozódia, kedy sa pre každého rečníka priamo vložia všetky typy jednotlivých kontúr. V prípade zložitejších modelov je treba vyberať kontúry automaticky. Na vybratie najbližšej kontúry uloženej v databáze sa používajú rôzne príznaky (slovné druhy, skladba slova, syntaktický kontext, prítomnosť prízvuku.) [81].

### 5.2.3.2 Časovanie

Časovaním môžeme nazvať aj frázovanie, keďže veta sa delí na viac krátkych prozodických úsekov – fráz.

Prozodická fráza je skupina slov, ktoré sú vnímané ako súzvučný intonačný celok na základe zvukovej charakteristiky hranice celku [77]. Jednoducho povedané, je to časť vety, ktorá musí byť povedaná jedným dychom. Jednotlivé prozodické frázy sú od seba navzájom oddelené pauzou, ktorá môže mať rôzne trvanie. Medzi úseky, ktoré majú väčšiu textovú spojitosť, sa vkladá kratšia pauza, ako medzi tie, ktoré ju majú menšiu. Preto je dôležitá spolupráca s textovými analyzátormi, či už so syntaktickým alebo morfologickým, lebo pomocou nich sa dá zistiť, ktoré slová spolu tvoria frázy a akú majú jednotlivé

frázy medzi sebou väzbu. Nie všetky druhy fráz sa dajú určiť pomocou jedného analyzátora. Na odhalenie niektorých fráz je potrebné poznať zmysel a význam vety.

## 5.3 Modifikácia pravidiel výslovnosti pri fonetickej transkripcii

Ľudia vo všeobecnosti čítaný text vyslovujú správne. Toto sa snažíme simulovať aj na systémoch fonetického prepisu. V praxi vieme tento výsledok dosiahnuť viacerými metódami. Prvá metóda sa dá definovať ako vedomostný prepis (knowladge-based) podľa LTS (letter-to-sound) pravidiel, kedy sa produkčné pravidlá vytvoria za pomoci fonetika – experta na stanovenie všeobecne platných LTS pravidiel. Základy takéhoto prístupu pre slovenčinu položil prof. A. Kráľ spolu s S. Daržágínom, kde vytvorené LTS pravidlá reflektujú množinu pravidiel pre ortoepický (fonetický) prepis slovenčiny [64].

Druhá metóda sa nazýva korpusová metóda (data-driven), ktorá automaticky generuje LTS pravidlá podľa kontextuálnych javov nájdených v manuálne popísanom korpuse. Súbor vytvorených pravidiel sa potom aplikuje na vstupný text. Cieľom tohto návrhu je vytvorenie transkripčného modelu, ktorý sa má naučiť pravidlá výskytu foném v trénovacej množine viet. Korpusová metóda má oproti vedomostnej metóde jednu nespornú výhodu v tom, že pri návrhu transkripčného modulu nie je potrebné mať k dispozícii experta fonetika. Nevýhodou je, že existujúce pravidlá treba neustále dopĺňať o nové (napr. priberanie nových slov z iných jazykov). Pri vytváraní systému sa vychádza z predpokladu, že je definovaný vzťah medzi postupnosťou znakov v ortografickom prepise a postupnosťou v ortoepickom prepise. Ide o tri rôzne typy vzťahov, a to 1:1, 1:N a N:1. Príklad takýchto vzťahov pre slovenský jazyk je pre prvý prípad fonéma „a“, ktorej vždy zodpovedá iba fonéma „a“, bez ohľadu na fonémy pred ňou a za ňou. Vzťah 1:N reprezentuje napríklad fonéma „d“, ktorej zodpovedá buď „d“ alebo „t“, v závislosti od predošlých a nasledujúcich foném. Posledná možnosť je reprezentovaná fonémami „i“, „y“, ktorým vždy zodpovedá prepis na „i“. Najzložitejšie je vytváranie pravidiel pre typ vzťahu 1:N. Toto bude rozobraté v nasledujúcej podkapitole.

Ďalšou metódou je transkripcia podľa slovníka. V slovníku sú uložené výnimky výslovnosti pre daný jazyk a pri fonetickej transkripcii sa najskôr pozerá do slovníka. Ak sa tam dané slovo nachádza, použije sa jeho fonetický prepis zo slovníka. Veľkosť slovníka však zvyčajne býva limitujúcim faktorom, preto sa snažíme mať v slovníku čo najmenej slov a radšej robiť prepis podľa pravidiel.

Nevýhodou všetkých vyššie spomenutých metód je, že ani jedna nedokáže pokryť všetky výnimky, ktoré sa v jazyku môžu vyskytnúť. Takéto výnimky vznikajú priberaním cudzích slov do jazyka a ich zaradením do bežnej dennej komunikácie (napríklad slovo internet). Zostáva vyriešiť úlohu, ako spraviť dokonalý prepis a dokonalý rečový synte-

tizátor. V praxi to znamená vytvoriť systém, ktorý môžeme postupne upravovať alebo vylepšovať, resp. inak povedané systém, ktorý je schopný sa učiť. To znamená vytvoriť systém syntézy, ktorý umožňuje prijímať „online“ úpravy a zmeny transkripcie, ktoré si syntetizátor nejakým spôsobom zapamätá.

### 5.3.1 Spodobovanie

Spodobovanie alebo asimilácia je proces, v ktorom sa jedna alebo viac hlások zvukovo prispôsobí druhej. K tomuto javu dochádza vtedy, ak v reči za sebou nasledujú neznelá a znelá spoluhláska (vzniknú dve znelé), alebo naopak znelá a neznelá spoluhláska (vzniknú dve neznelé). V tomto procese hláska nadobúda alebo stráca takú vlastnosť, ktorou sa odlišovala od susednej hlásky (vlastnosť, ktorou sa akusticky a artikulačne podobá susednej hláske) [80]. Spodobovanie sa prejavuje v znelostných pároch (Tab. 5).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **neznelé** | p | t | c | ts | tS | k | x | f | s | S |
| **znelé** | b | d | J\ | dz | dZ | g | G | w | z | Z |

***Tab. 5.*** *Znelostné páry*

Pri spodobovaní existujú aj výnimky, napríklad pri slovách, ktoré nie sú predponové, ako smer, svah, sloh, atď. Rovnako sa nespodobujú ani predložky „s“ a „k“ pred niektorými tvarmi osobných zámen, napríklad s ním, k nám. Ďalej si rozoberieme niektoré špeciálne prípady transkripcie.

Samohláska „ä“ sa číta tromi spôsobmi [{], [e] a [a]. V spisovnej slovenčine sa vyskytuje len po párnych spoluhláskach „p“, „b“, „m“, „v“ (napr. päta, holúbä, mäso, deväť). Taktiež dochádza k diferenciácii a výslovnosť „ä“ [päta, mäso] začína pociťovať ako príznaková – ako znak vyššieho štýlu výslovnosti. Dnes sa pokladajú za základné tvary aj slová s obyčajným „e“ [deveť, peť, meso], [80]. Preferovanie samohlásky „ä“ neodráža živú jazykovú skutočnosť a jej realizáciu si v spisovnej výslovnosti nemožno ani vynucovať, ani predpisovať [78].

Pre spoluhlásku „v“ existujú v slovenskom jazyku päť rôznych prepisov [v, u_^, w, f, f_v]. Ako [v] sa vyslovuje pred samohláskou, dvojhláskou alebo zvučnými hláskami (napr. voda, viať, vrana, vlak). Ako [u_^] sa vyslovuje, keď je v na konci slabiky (napr. krv, dievča). [w] sa vyslovuje pred znelými párovými spoluhláskami (napr. vdova, vďaka). [f] sa vyslovuje pred neznelými párovými spoluhláskami (napr. včela, včas). [f_v] sa vyslovuje na začiatku slabiky, ak bezprostredne za „v“ nenasleduje samohláska alebo „r“, „l“, „ŕ“, „Ĺ“, „ľ“, „j“ (v dome, v Prešove).

Fonetický prepis pre „r“ a „l“ môže byť nasledovný: [r, r=, r=: a l, l=, l=:]. Tieto hlásky patria medzi znelé nepárové a sú aj slovotvorné, podobne ako samohlásky. Slabičné [r=] sa od neslabičného [r] líši počtom kmitov, dlhé od krátkeho sa líši v dĺžke trvania (vrkoč, krv, vŕba, stĺp).

Spoluhlásku „n“ vyslovujeme v slovenčine tromi spôsobmi [n, N, N\]. [N] vyslovujeme vtedy, keď po „n“ nasleduje „k“ alebo „g“ (banka, cengať, mienka). [N\] je nazývané aj úžinové „n“, a to vtedy, keď za ním nasleduje ďasnová úžinová spoluhláska „s“, „z“, „š“, „ž“ (banský, inžinier, Slovensko). Pre tento znak neexistuje ekvivalent v IPA ani SAMPA abecede od Ivaneckého. Preto sa prepisuje ako obyčajné „n“. Tento znak bol pridaný až do abecedy SAMPA vyvinutej na SAV.

Spoluhláska „j“ sa vyslovuje ako [j, i_^]. [i_^] sa odlišuje od „i“ tým, že je oslabená samohlásková rezonancia. Vyslovuje sa aj ako súčasť dvojhlások [i_^a, i_^e, i_^u]. [j] vyslovujeme vtedy, ak sa „j“ nachádza na začiatku slabiky, nemusí to byť absolútny začiatok slova. Takto sa vyslovuje aj vtedy, ak je pred ním spoluhláska a medzi ním a predchádzajúcou spoluhláskou je slabičná hranica, čiže „j“ je prvou hláskou v slabike (jama, jasný, bojisko, benjamín, injekcia). Ak sa nachádza na konci slabiky, vyslovujeme ho ako nešumovú hlásku [i_^] bez ohľadu na to, či nasleduje spoluhláska alebo samohláska (krajší, kraj, vylejte).

Spoluhláska „ch“ sa môže vyslovovať ako [h, G, x]. „Ch“ vyslovíme ako [h], keď sa nachádza pred znelou hláskou. Znelostné spodobovanie „ch“ a „h“ vzniká tým, že oproti neznelému [x] stoja dve znelé hlásky [G] a [h\]. Ak sa na asimilačnom mieste nachádza hláska „h“, vyslovuje sa podľa pravidiel spodobovania v slovenskom jazyku (juh, prah miestnosti). Ak sa „h“ alebo „ch“ nachádzajú na mieste, kde dochádza k spodobovaniu a hneď po nich nasleduje „h“, výslovnosť je zvyčajne [G h\] (lieh horí, duch Helsiniek, Váh hučí).

Spoluhláska „m“ sa môže vysloviť dvoma spôsobmi [m, F]. Ako [F] ho vyslovujeme vtedy, ak vo vnútri slova po „m“ nasleduje „v“ alebo „f“ (komfort, lymfatický, symfónia). Nasledujúce tabuľky (Tab. 6, Tab. 7) zobrazujú prepis grafém na fonémy k nim prislúchajúce:

| G | a | e | i | o | u | ä | | á | é | í | ó | ú | | ia | ie | iu | ô |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | a | e | i | o | u | { | | a: | e: | i: | o: | u: | | i_^a | i_^e | i_^u | u_^o |
| F2 | | | | | | e | | | | | | | | | | | |
| F3 | | | | | | a | | | | | | | | | | | |

***Tab. 6.*** *Prepis grafém samohlások*

| G | p | b | t | d | ť | ď | k | g | c | dz | č | dž | f | v | s | z | š | ž | ch | h | j | r | ŕ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| F1 | p | b | t | t | c | J\ | k | g | ts | dz | tS | dZ | f | v | s | z | S | Z | x | h | J | r | r=: |
| F2 | b | p | d | d | J\ | c | g | k | dz | ts | dZ | tS | f_v | w | z | s | Z | S | G | x | i_^ | r= | |
| F3 | | | | | | | | | | | | | | u_^ | | | | | h | G | | | |
| F4 | | | | | | | | | | | | | | f | | | | | | | | | |
| F5 | | | | | | | | | | | | | | f_v | | | | | | | | | |

| G | l | Í | ľ | m | n | ň | w | x |
|---|---|---|---|---|---|---|---|---|
| F1 | l | l= | L | m | n | J | v | ks |
| F2 | | l=: | | F | N | | | gz |
| F3 | | | | | N\ | | | |
| F4 | | | | | | | | |
| F5 | | | | | | | | |

***Tab. 7.*** *Prepis grafém spoluhlások*

### 5.3.2 Návrh modulu transkripcie

Možnosti na zlepšenie transkripcie v doteraz existujúcom syntetizátore sú nasledovné:

- **Ručný prepis** v SAMPA abecede. Rozhranie syntetizátora umožní používateľovi ručne prepísať chybné slovo. Po prepise si používateľ môže znovu vypočuť syntetizovaný text už s opraveným slovom (alebo viacerými slovami).
- **Výber slova s nesprávnou transkripciou** s ponukou možnosti alternatívnej výslovnosti. Rozhranie syntetizátora ponúkne používateľovi (napr. po kliknutí na nesprávne slovo) všetky možnosti, ako sa dá ešte dané slovo v slovenčine vysloviť. Možnosti sa zostavia na základe pravidiel výslovnosti. Na ústave telekomunikácií sa opierame o pravidlá získané v práci pána Cerňaka. Tieto pravidlá boli určené spomínanou korpusovou metódou, ktorej úspešnosť je približne 97% [65]. Na základe vytvorených pravidiel vieme vytvoriť možné typy výslovnosti jednotlivých foném a ponúknuť používateľovi zoznam možností rôznych výslovností opravovaného slova. Napríklad pri slove „teraz“ syntetizátor na základe pravidiel zistí, že hlásky spôsobujúce problém môžu byť „t“ a „z“. Ponúkne možnosti opravy pre „t“ a „z“, napr. t sa číta ako „t“ alebo „ť“, z sa môže spodobiť na „s“. Používateľ si vyberie z ponúkaných možností správnu a môže si znovu vypočuť syntetizovaný text už s opraveným slovom (alebo viacerými slovami). Ak ani jedna z ponúkaných možností nevyhovuje, je možnosť ručnej opravy textu. Na základe ručnej opravy textu je možné rozšíriť oznam alternatívnej výslovnosti jednotlivých foném.
- **Automatické/inteligentné prehľadávanie možností** alternatívnej výslovnosti.
- **Akustická oprava výslovnosti**
- **Oprava rovnakých slov** (alebo rovnakého kmeňa slov) v texte.

Pre proces fonetickej transkripcie je potrebné nájsť optimálny spôsob učenia, teda naučiť systém, aby dokázal spracovať nové údaje a aplikovať ich v ďalšom procese syntézy. Návrh takého učenia je podrobnejšie popísaný v kapitole 5.5.7.

Na začiatku transkripcie je k dispozícii slovník Ábela Kráľa a súhrn pravidiel. Ak nastane oprava slova, je zrejmé, že dané slovo nie je v slovníku a že jeho prepis podľa pravidiel nebol správny. Takto opravené slovo sa na začiatku uloží do zoznamu výnimiek. Zoznam obsahuje pôvodné slovo, opravené slovo a meno používateľa, ktorý ho opravil. Iba zbierať slová a ukladať do slovníka výnimiek by však neskoršie mohlo proces syntézy výrazne spomaliť. Preto je potrebné pri určitom množstve záznamov prejsť slovník výnimiek. Ak sa postupnosť hlások vyskytuje v slovníku viackrát a vždy s rovnakou výslovnosťou, vytvorí sa z nich nové pravidlo a súbor pravidiel sa rozšíri (napríklad *inter* v slovách „internet“, „interný“ sa vždy číta tvrdo). Veľmi dôležité je nájsť hranicu a určiť, či postupnosť hlások (alebo slovo) zostane v slovníku výnimiek, alebo sa vytvorí pravidlo. Teda ak je viacero slov s rovnakým základom, je vhodnejšie pre takéto slová vytvoriť nové pravidlo. Ak by však nové pravidlo znamenalo, že algoritmus transkripcie musí prejsť celým slovom, je výhodnejšie ponechať slovo v slovníku. Proces učenia transkripcie môžeme zhrnúť do dvoch základných bodov:

- ukladanie výnimiek do slovníka
- štatistické vyhodnotenie zmien výslovnosti z pohľadu času a pretrénovanie pravidiel

### 5.3.3 Implementácia modulu transkripcie s možnosťou učenia

Vstupný blok modelu transkripcie načíta vetu zo vstupného XML súboru. Je dôležité načítať slová v správnom poradí, pretože sa navzájom ovplyvňujú pri výslovnosti. Po načítaní sa vstupný text prepíše podľa pravidiel a výstup sa zobrazí na stránke. Registrovanému používateľovi sa umožní vypočuť syntetizovaný text a označiť si chybne zosyntetizované slová. Počas opravnej fázy používateľ postupne opravuje označené slová. Program vygeneruje všetky možné výslovnosti daného slova a ponúkne ich v rolovacom menu. Možnosti sú vytvárané na základe tabuľky prekladu grafém do foném. Pre tieto potreby bol vytvorený XML súbor nazvaný *transkripcia.xml*, v ktorom sú definované tieto vzťahy [84].

```
<?xml version="1.0" encoding="UTF-8"?>
<transcription xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
   <grapheme val="a">
          <pho>a</pho>
   </grapheme>
   <grapheme val="á">
          <pho>a:</pho>
   </grapheme>
   <grapheme val="ä">
          <pho>{</pho>
          <pho>e</pho>
          <pho>a</pho>
   </grapheme>
   <grapheme val="b">
          <pho>b</pho>
          <pho>p</pho>
   </grapheme>
   <grapheme val="c">
          <pho>ts</pho>
          <pho>dz</pho>
   </grapheme>
   <grapheme val="č">
          <pho>tS</pho>
          <pho>dZ</pho>
```

```
    </grapheme>
.
.
.
```

```
N i e k t o r= a:
N i e k d o r a:
N i e k d o r= a:
N i e g t o r a:
N i e g t o r= a:
N i e g d o r a:
N i e g d o r= a:
N\ i_^e k t o r a:
N\ i_^e k t o r= a:
N\ i_^e k d o r a:
N\ i_^e k d o r= a:
N\ i_^e g t o r a:
N\ i_^e g t o r= a:
N\ i_^e g d o r a:
N\ i_^e g d o r= a:
N\ i e k t o r a:
N\ i e k t o r= a:
N\ i e k d o r a:
N\ i e k d o r= a:
N\ i e g t o r a:
N\ i e g t o r= a:
N\ i e g d o r a:
N\ i e g d o r= a:
```

Používateľ si teda vyberie jednu z ponúkaných možností a opravenú verziu si môže vypočuť. Prehrá sa mu nielen opravované slovo, ale aj slovo pred ním a za ním, pretože, ako bolo spomenuté, výslovnosť chybného slova je závislá v niektorých prípadoch aj od slov, ktoré dané slovo obklopujú. Ak sa používateľovi stále nepáči výslovnosť slova, môže celý proces opravy zopakovať.

### 5.3.4 Databáza opráv

Ako už bolo spomenuté v návrhu tohto modulu, proces opravy slov by nemal zmysel bez toho, aby sa syntetizátor učil nové slová a ich správnu výslovnosť.

Pri vytváraní databázy opravených slov sme sa snažili vyhnúť akejkoľvek nadbytočnosti. Pred zápisom sa vždy skontroluje prítomnosť daného slova v databáze (prvého aj druhého z ukladanej dvojice, v tabuľkách prvé slovo a druhé slovo). Ak slovo neexistuje, zapíše sa do tabuľky. Ak už slovo zapísané je, zistí sa jeho ID z dôvodu vytvárania tabuľky ID slov, ktoré sú vedľa seba vo vete (tabuľka slovo jedna dva). Opravované slovo však môže byť na prvom alebo druhom mieste tejto dvojice, preto sa zapíše iba fonetika opravovaného slova. Ďalšou tabuľkou je tabuľka väzieb (tabuľka fonetika), kde sa nachádza aj

informácia o váhe opravovaného slova. Táto váha závisí od dôveryhodnosti používateľa. Dôveryhodnosť používateľa sa meria jeho skúsenosťami. Ak je táto hodnota slova väčšia, znamená to, že bolo rovnako opravené veľkým počtom používateľov alebo používateľmi s väčšími skúsenosťami. Preto sa po zápise fonetickej výslovnosti zvýši aj dôveryhodnosť používateľa.

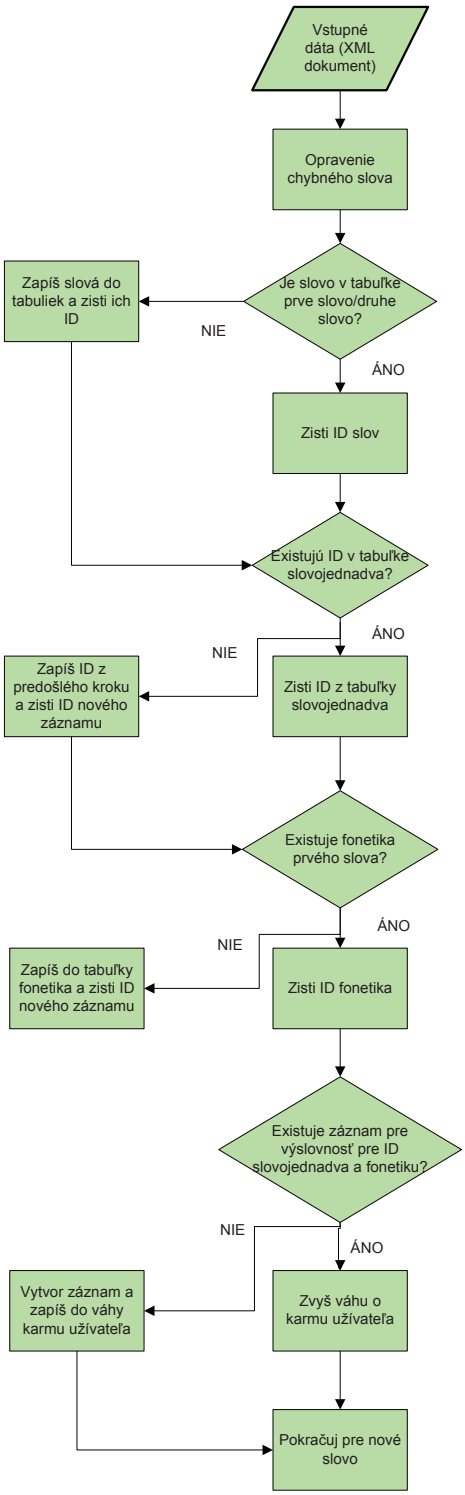
Keďže ide o webovú aplikáciu, je možné jej používanie viacerými používateľmi naraz. Výhodou je získanie veľkého množstva vstupných dát, ktoré ďalej umožnia vytvorenie nových pravidiel pre fonetickú transkripciu [84].

Výstup z modulu pre fonetickú transkripciu je uložený do XML súboru. Príklad takéhoto výstupu je zobrazený nižšie:

```
<?xml version="1.0" encoding="UTF-8"?>
<text>
  <sentence text="Automotokluby v Európe majú najväčšiu sieť asistenčných
služieb" delim=".">
    .
    .
    .
      <word val="najväčšiu">
        <transkript>LTS-RULES</transkript>
        <syl>
          <pho>n</pho>
          <pho>a</pho>
          <pho>I_^</pho>
          <pho>v</pho>
          <pho>E</pho>
          <pho>tS</pho>
          <pho>I_^U</pho>
        </syl>
      </word>
      <word val="sieť">
        <transkript>DICTIONARY</transkript>
        <syl>
          <pho>s</pho>
          <pho>I_^E</pho>
          <pho>J\</pho>
        </syl>
      </word>
      <word val="asistenčných">
        <transkript>LTS-RULES</transkript>
        <syl>
          <pho>a</pho>
```

```
        <pho>s</pho>
        <pho>I</pho>
        <pho>s</pho>
        <pho>c</pho>
        <pho>E</pho>
        <pho>n</pho>
        <pho>tS</pho>
        <pho>n</pho>
        <pho>I~</pho>
        <pho>h</pho>
      </syl>
    </word>
    <word val="služieb">
      <transkript>LTS-RULES</transkript>
      <syl>
        <pho>s</pho>
        <pho>l</pho>
        <pho>U</pho>
        <pho>Z</pho>
        <pho>I_^E</pho>
        <pho>b</pho>
      </syl>
    </word>
  </sentence>
```

Ako vidieť, vstupná veta je uložená v atribúte `segment`, kde je rozčlenená do podtagov `word`. Každý tag `word` obsahuje podtagy `pho`, kde sú prepísané jednotlivé grafémy do foném buď podľa pravidiel, alebo podľa slovníka. Toto je uložené v atribúte `transkript`, pre pravidlá je to hodnota `LTS-RULES` a pre prepis na základe slovníka `DICTIONARY`.




***Obr. 23.*** *Algoritmus opravy fonetickej transkripcie slova a jeho uloženie do databázy*

### 5.3.5 Inteligentné učenie

Z vyššie opísaných metód na zlepšenie transkripcie sa posledné dva body dajú zaradiť aj do kategórie učenia syntetizátora. Možnosti učenia a automatizovanej opravy sú nasledovné:

- Automatické/inteligentné prehľadávanie možnosti alternatívnej výslovnosti. Táto možnosť je podobná druhej možnosti z návrhu tohto modulu (kapitola 5.3.2), je však oveľa inteligentnejšia. Všetky možnosti nie sú zoradené náhodne alebo podľa abecedy, ale podľa pravdepodobnosti výskytu a aj určitej logiky. Rozhranie syntetizátora ponúkne používateľovi (napr. po kliknutí na nesprávne slovo) všetky možnosti s tým, že hlásky, ktoré môžu spôsobiť nesprávne prečítanie, sú nahradené postupne všetkými hláskami z abecedy. Možnosti sa zoradia logicky a podľa najvyššej pravdepodobnosti výskytu. Napríklad pri slove „teraz“ syntetizátor na základe pravidiel zistí, že hlásky spôsobujúce problém sú „t“ a „z“. Ponúkne možnosti opravy pre „t“ a „z“, napr. „t“ sa číta ako „t“ alebo „ť“, z sa môže spodobiť na „s“. Napríklad „t“ nahradené samohláskou je nezmysel, pretože dve samohlásky na začiatku slova sa v slovenskom jazyku nevyskytujú. Preto takéto možnosti patria na koniec. Taktiež spoluhlásky sú zoradené logicky, napríklad najskôr „n“ - vznikne slovo „neraz“ a pod. Ďalšie sú možnosti, kde sa postupne prvý až posledný znak v slove nahradia všetkými možnosťami z abecedy. Vytvorenie takéhoto zoznamu si však môže vyžiadať viac času.
- Akustická oprava výslovnosti. Tento spôsob opravy si vyžaduje dokonalý rečový rozpoznávač. Používateľ po vypočutí zosyntetizovaného textu opraví chybné slovo alebo slová tak, že ich sám prečíta. Môže prečítať aj celú vetu, v ktorej sa chybné slovo alebo slová nachádzajú. Rozpoznávač vetu zapíše a syntetizátor porovná, ktoré slovo alebo slová boli nesprávne a nahradí ich. Používateľ si môže znovu vypočuť syntetizovaný text už s opraveným slovom a v prípade chyby proces zopakuje.
- Oprava rovnakých slov (alebo rovnakého kmeňa slov) v texte. Pri syntéze dlhšieho textu sa dá predpokladať, že sa niektoré slová vyskytujú v texte viacnásobne. Preto v prípade akejkoľvek opravy sa prehľadá syntetizovaný text a ak nájde opravované slovo/slová/časť slova, označí ich (farebne vysvieti alebo podčiarkne). Používateľ prejde označenými slovami a rozhodne, či opraviť niektoré alebo všetky označené slová (implementuje sa funkcia prepíš slovo alebo prepíš všetky slová). Môže nastať prípad, že syntetizátor deteguje dva slovotvorné základy, pričom bude výslovnosť v nich rôzna (teda závislá od susedných foném), napr. internet a netreba. V oboch sa nachádza spojenie net, v prvom slove sa číta tvrdo a v druhom mäkko. Takáto funkcionalita významne urýchli opravu slov, kdeže používateľ nemusí sám prehľadávať všetky slová a vyberať správne možnosti.

Rovnako aj pri týchto spôsoboch opravy implementujeme alebo zdokonalíme už existujúci systém učenia. Teda ukladáme opravené slová do databázy, ako je to opísané

v predošlých kapitolách.

Pri tomto postupe hľadáme optimálnu hranicu pre robenie nových pravidiel a ponechávanie slov v slovníku výnimiek. Pri vytváraní pravidla berieme do úvahy aj váhu opravovaného slova, ktorá sa vytvára na základe počtu opráv alebo dôveryhodnosti používateľa, ktorý slovo opravil. Treba sa taktiež vyvarovať situácii, aby sa do pravidiel dostalo nekorektne opravené slovo. Jednou z možností vytvárania pravidiel a triedenia slov do slovníka je vytvárať ich manuálne. Tento proces je síce veľmi pomalý, ale je presný. Chybovosti sa dá predísť aj tým, ak by pravidlá nezávisle od seba vytvárali viacerí ľudia. Ich čiastkové pravidlá porovnáme a vyberieme výsledné pravidlá a konečný slovník výnimiek. Druhou možnosťou je tento proces zautomatizovať, čo povedie k vyššej efektivite. Je však potrebné vytvoriť algoritmus na prehľadávanie databázy. Pre dostatočne veľkú váhu slova alebo iba časti sekvencie foném opakujúcich sa vo viacerých slovách vytvoríme pravidlo a to sa pridá k už existujúcim pravidlám. Pri slovách, ktoré sú na hranici, teda je ich príliš málo na vytvorenie pravidla a dosť veľa na to, aby zostali v slovníku, počkáme, či v krátkom čase nepribudnú ďalšie. Proces vytvárania nových pravidiel je nutné v určitých časových intervaloch opakovať.

## 5.4 Modul určovania slovných druhov

Dobrý rozbor syntetizovaného textu z morfologického hľadiska a tiež správne určenie slovných druhov a gramatických kategórií je základ pre ďalšie bloky morfologicko-syntaktickej analýzy a z globálneho hľadiska je potrebný pre kvalitnú syntézu. Napríklad správne určený slovný druh pomôže k správnemu prečítaniu skratky alebo číslovky vo vete.

Gramatiku slovenského jazyka delíme na vetnú skladbu (syntax) a tvaroslovie (morfológia). V jazykovede je vetná skladba, alebo inak povedané syntax, časť gramatiky zaoberajúca sa vzťahmi medzi slovami vo vete, správnym tvorením vetných konštrukcií a slovosledom. Tvaroslovie skúma slovné druhy, skloňovanie, časovanie a stupňovanie slovných druhov a pravidelným odvodzovaním slov pomocou predpôn, prípon a vpôn. Dôležité je poznať vzťahy oboch týchto disciplín, ale viac sa budeme zaoberať morfológiou.

Morfologickou anotáciou vieme zapísať všetky vlastnosti slova, ktoré mu vieme určiť. Každá informácia je vyjadrená znakom. Takáto anotácia sa nazýva `POS`-tag. Na vstupe tohto procesu je veta, teda slová z daného jazyka a špecifikovaný tagset (konečný počet morfologických značiek). Výstupom je najlepší `POS`-tag pre každé slovo. Jeden z takýchto tagsetov používa aj Slovenský národný korpus.

Morfologickej anotácii podliehajú všetky textové jednotky – tokeny. Tokeny sú všetky znaky medzi dvoma medzerami, patria sem aj interpunkčné znamienka. Na zápis morfologických značiek sa používajú písmená latinskej abecedy, číslice a matematické

symboly. Súbor jednotlivých znakov tvorí jeden tag k jednému tokenu. Tag vyjadruje hodnoty formálnych kategórií, ktoré sú pre daný token relevantné. V Slovenskom národnom korpuse (ďalej len SNK) sa používajú tagy s variabilným počtom znakov, ich poradie v tagu je však záväzné [83].

Viacero modulov syntetizátora vychádza zo Slovenského národného korpusu. Celý korpus obsahuje 1 207 939 tokenov a z toho asi 982 924 slov písaných slovenskou abecedou. Predpokladáme, že slová písané slovenskou abecedou sú slovami slovenského jazyka. Má význam ich ďalej spracovať. Mnoho cudzojazyčných slov používa rovnako ako slovenský jazyk základnú abecedu. V SNK by však slová z cudzích jazykov mali byť označené tagom %. Toto označenie je platné pre citátový výraz, ale malo by umožniť vynechanie cudzích slov z korpusu. Korpus môže obsahovať aj špeciálne znaky, ako znaky na označenie fyzikálnych veličín, matematické operácie, interpunkčné znamienka, atď. SNK sa skladá z viacerých korpusov, najznámejšie sú tieto dva:

| | **Delenie slovných druhov podľa kritérií** | | |
|---|---|---|---|
| **Slovné druhy** | **Ohybnosť** | **Vecný význam** | **Vetnočlenská platnosť** |
| Podstatné mená | skloňujú sa (ohybné) | plnovýznamové | sú vetné členy |
| Prídavné mená | skloňujú sa (ohybné) | plnovýznamové | sú vetné členy |
| Zámena | skloňujú sa (ohybné) | plnovýznamové | sú vetné členy |
| Číslovky | skloňujú sa (ohybné) | plnovýznamové | sú vetné členy |
| Slovesá | skloňujú sa (ohybné) | plnovýznamové | sú vetné členy |
| Príslovky | neohybné | plnovýznamové | sú vetné členy |
| Predložky | neohybné | neplnovýznamové | nie sú vetné členy |
| Spojky | neohybné | neplnovýznamové | nie sú vetné členy |
| Častice | neohybné | neplnovýznamové | nie sú vetné členy |
| Citoslovcia | neohybné | neplnovýznamové | nie sú vetné členy |

***Tab. 8.*** *Delenie slovných druhov*

Viacero modulov syntetizátora vychádza zo Slovenského národného korpusu. Celý korpus obsahuje 1 207 939 tokenov a z toho asi 982 924 slov písaných slovenskou abecedou. Predpokladáme, že slová písané slovenskou abecedou sú slovami slovenského jazyka. Má význam ich ďalej spracovať. Mnoho cudzojazyčných slov používa rovnako ako slovenský jazyk základnú abecedu. V SNK by však slová z cudzích jazykov mali byť označené tagom %. Toto označenie je platné pre citátový výraz, ale malo by umožniť vynechanie cudzích slov z korpusu. Korpus môže obsahovať aj špeciálne znaky, ako znaky na označenie fyzikálnych veličín, matematické operácie, interpunkčné znamienka, atď. SNK sa skladá z viacerých korpusov, najznámejšie sú tieto dva:

- Jednojazyčný korpus písaných textov – aktuálna verzia je *prim-4.0*, korpus obsahujúci všetky texty. Má 526 082 640 tokenov (65% publicistické, 17% umelecké,

16% odborné a 2% neurčené texty).
- Ručne morfologicky anotovaný korpus – aktuálna verzia je *r-mak-3.0*, obsahuje 1 207 939 tokenov (44,3% umelecké, 36,7% publicistické a 19,0% odborné texty), [76].

Pri jednojazyčnom korpuse sa morfologická anotácia robí pomocou programu a tento nedokáže určiť všetky morfologické vlastnosti úplne presne. Pri použití tohto korpusu je úspešnosť maximálne taká, ako úspešnosť programu na jeho anotovanie. Preto použijeme ručne anotovaný korpus. Tento korpus nie je úplne bezchybný, pretože sa nájdu slová, pri ktorých sa nedá slovný druh určiť jednoznačne. Proces anotácie však vykonávajú dvaja anotári nezávisle, čo vedie k ešte menšej chybe. Pri odstránení všetkých nepotrebných značiek z korpusu zostane 982 924 tokenov, teda slov slovenskej abecedy. Tento slovník považujeme za dostatočne veľký na to, aby sme mohli z neho vychádzať a vytvoriť slovníky pre jednotlivé moduly (určovanie slovných druhov a kategórií, skratiek.). Výhodou oboch slovníkov je, že dokážu určiť aj slová, ktoré sa nenachádzali v korpuse.

„Na prvom mieste vždy stojí informácia o príslušnosti k slovnému druhu (podľa zaužívanej desaťčlennej slovnodruhovej typológie), resp. k slovnej triede (sem patria špecifické textové jednotky vrátane interpunkcie a neslovných elementov vyskytujúcich sa v bežnom texte). Nasledujú značky pre príslušné gramatické kategórie (záväzne), resp. značky pre špeciálne skupiny (nezáväzne – stoja na konci tagu po dvojbodke a označujú vlastné mená a chybné zápisy).“ [75] Prehľad značiek prislúchajúcich k slovným druhom je v tabuľke (Tab. 9) [83].

| Slovný druh | Značka | Slovný druh | Značka | Slovný druh | Značka |
|---|---|---|---|---|---|
| Substantívum | S | Prepozícia | E | Interpunkcia | Z |
| Adjektívum | A | Konjunkcia | O | Neurčiteľný slovný druh | Q |
| Pronominum | P | Partikula | T | Neslovný element | # |
| Numerále | N | Interjekcia | J | Citátový výraz | % |
| Verbum | V | Reflexívum | R | Číslica | 0 |
| Particípium | G | Kondicionálová morféma | Y | Vlastné meno | :r |
| Adverbium | D | Abreviácia, značka | W | Chybný zápis | :q |

***Tab. 9.*** *Prehľad značiek priradených k slovným druhom v SNK*

V tabuľke je zoznam skupiny znakov (hodnota a príklad) špecifikujúcich slovný druh podstatné meno. Do výslednej morfologickej značky sa vyberá iba jeden znak z každej pozície. Platná morfologická značka pre podstatné meno môže vyzerať napr. `SSfs6` alebo `SSmp2:r` [83].

Príklad použitia predošlej tabuľky na morfologickú analýzu vidieť v nasledujúcej vete:

*„Pripravili ohnisko/SSns4 na grilovanie/SSns4 a o chvíľu/SSfs4 sa už po lúke/SSfs6 niesla lahodná vôňa/SSfs1, ktorá prilákala ďalších stravníkov/SSmp4“.*

Slovné druhy sa určujú na základe posledných troch písmen slov. Každá koncovka má priradený slovný druh. Slovenská gramatika má tú vlastnosť, že dáva do skupín slová s rovnakou koncovkou a morfologickou vlastnosťou. Využitím tohto poznatku vznikol slovník obsahujúci trojpísmenkové koncovky slov získaných zo SNK. K slovám v slovníku má priradený len slovný druh. Jeho výsledky dosahújú úspešnosť 90%, ak sa neurčujú interpunkčné znamienka. Úspešnosť na prvý pohľad vyzerá vysoká, ale v preklade to znamená, že jedno z desiatich slov je určené nesprávne. Preto chceme túto presnosť zlepšiť a prípadne pridať ďalšie morfologické informácie určovaným slovám (gramatické kategórie).

| Pozícia | Znak | Hodnota | Príklad |
|---|---|---|---|
| 1. slovný druh | S | substantívum | slovo, ryba, ústav, muž |
| 2. paradigma | S | substantívna | chlap, žena, srdce |
| | A | adjektívna | hlavný, vedúci, Mastný, starká, vstupné |
| | F | zmiešaná | kuli, gazdiná |
| | U | neúplná | kanoe, kupé |
| 3. rod | m | mužský živ. | hrdina, hlavný, Mastný |
| | i | mužský neživ. | strom, rýľ |
| | f | ženský | ulica, pani, vedúca, Slaná, hradská |
| | n | stredný | mesto, vysvedčenie, dievča, mláďa |
| 4. číslo | s | jednotné | slovo, ryba, ústav, muž |
| | p | množné | slová, ryby, ústavy, muži/mužovia |
| 5. pád | 1 | nominatív | pán, matka, Slaná, more, mláďa |
| | 2 | genitív | pána, matky, Slanej, mora, mláďaťa |
| | 3 | datív | pánovi, matke, Slanej, moru, mláďaťu |
| | 4 | akuzatív | pána, matku, Slanú, more, mláďa |
| | 5 | vokatív | pane, mami, Táni, |
| | 6 | lokál | pánovi, mame, Slanej, mori, mláďati |
| | 7 | inštrumentál | pánom, matkou, Slanou, mláďaťom |

***Tab. 10.*** *Tabuľka anotácie pre podstatné meno v SNK*

### 5.4.1 Návrh modulu morfologickej analýzy

Keďže naším cieľom je dokonalý syntetizátor s výstupom nerozoznateľným od hovorenej reči, je potrebné zdokonaliť aj proces morfologickej analýzy. Návrh ideálneho určovania slovných druhov je nasledovný:

- Určovanie podľa koncoviek. Ako bolo spomenuté vyššie, základom na určenie

z koncovky slova sú posledné 3 písmená. V prípade, že slovný druh nie je možné jednoznačne určiť z posledných 3 písmen, treba brať do úvahy aj predošlé znaky a prehľadávať slová smerom k ich začiatku, až kým nie je slovný druh určený. Proces prehľadávania sa zaznamená do cart stromov. Vznikne tak súbor popisujúci pravidlá na určenie slovných druhov.

- Určenie slovného druhu z kontextu. Tento spôsob predpokladá správne určenie minimálne polovice slovných druhov vo vete prvým spôsobom. Na ich základe sa potom určia slovné druhy ostatných slov. Vhodné je obe metódy kombinovať. Teda najskôr podľa zostavených pravidiel určiť slovné druhy pre čo najviac slov a potom z kontextu určiť slová, ktoré sa prvým spôsobom nedali jednoznačne určiť. Samozrejme v ideálnom prípade sa všetky slová určia prvým spôsobom.

| **Veta** | Videl | som | strom, | ktorý | je | zelený. |
|---|---|---|---|---|---|---|
| **Slovné druhy** | sloveso | zvratné zámeno | podstatné meno | spojka/ ~~opytovacie zámeno~~ | sloveso | prídavné meno |
| **Veta** | Ktorý | stom | je | zelený? | | |
| **Slovné druhy** | ~~spojka~~/ opytovacie zámeno | podstatné meno | sloveso | prídavné meno | | |

***Tab. 11.*** *Príklad nejednoznačnosti slovného druhu*

Jednoduchý príklad v Tab. 11 znázorňuje princíp vyššie popísanej metódy. V dvoch rôznych vetách sa nachádza to isté slovo, v našom príklade je to slovo ktorý. V každej vete je však iným slovným druhom. Predpokladom je, že máme jednoznačne určené slovné druhy pred a za takýmto slovom. Pomocou nich potom určíme slovný druh nejednoznačného slova. V prvom prípade treba zvážiť, či sa v strede vety medzi podstatným menom a slovesom vyskytujú opytovacie zámená. Ideálne by bolo, keby syntetizátor vedel aj rozpoznať, či ide o vetu oznamovaciu alebo opytovaciu. Oba tieto poznatky vedú k riešeniu, a to k určeniu, že v tomto prípade ide o spojku. Podobne sme riešili druhý prípad. Na začiatku opytovacej vety spojka v slovenskom jazyku nebýva. Tento príklad je veľmi jednoduchý a pri syntéze reálnych textov sa môžu vyskytnúť aj oveľa zložitejšie prípady. Preto je potrebné vytvoriť všeobecne logické pravidlá platné pre slovenský jazyk, podľa ktorých sa takéto prípady vyriešia.

- Určenie ostatných parametrov ako rod, číslo, čas. Pre správne čítanie čísloviek alebo skratiek je určenie vyššie spomenutých parametrov dôležité. Napríklad máme vetu: *„Na výlete boli 4 ženy a 4 muži.“* Táto veta sa prečíta ako: *„Na výlete boli štyri ženy a štyria muži.“* Vyriešiť takýto problém však v oblasti syntézy reči vôbec nie je triviálna záležitosť. Určenie rodu, čísla a času najlepšie vykonáme z koncovky daného slova, prípadne slova pred ním a za ním. Rozšírením pravidiel

z prvého bodu vzniknú komplexné pravidlá na určenie nielen slovných druhov, ale aj ostatných slovných parametrov.
- Ručná oprava nesprávne určeného slovného druhu.
- Automatické/inteligentné prehľadávanie možnosti alternatívnych slovných druhov.
- Oprava rovnakých slov (alebo rovnakého kmeňa slov) v texte.

### 5.4.2 Vytváranie trojfonémového slovníka

Pri vytváraní slovníka je na začiatku dôležité zodpovedať si otázku, na koľko percent musí byť tag určený, aby sme ho považovali za správny. Nami stanovená hranica je 99%. Do zostávajúceho 1% sú uvažované prípady, keď ide o cudzojazyčné slovo (napr. a ako neurčitý člen v anglickom jazyku) alebo ako jednoslabičné slovo v spojení s iným slovom (oktáva a mol) alebo v prípade skratky.

Vytvoríme slovník, kde je ku všetkým koncovkám v slovenskom jazyku priradený slovný druh a gramatické kategórie. Proces vytvárania slovníka čo najviac zautomatizujeme.

Na začiatku sa z korpusu, uloženého v súbore tak, že v jednom riadku je jedno slovo, vyberajú koncovky týchto slov spolu s ich morfologickou informáciou (slovný druh, v niektorých prípadoch aj gramatické kategórie). V prípade, že slovo je kratšie ako 3 písmená (napr. „a“, „za“, „do“), sa extrahuje celé slovo a podľa počtu písmen sa doplnia medzery na začiatok tak, aby bola vždy dĺžka 3. Takto vybraté koncovky aj tagy sa uložia do súboru, tagy sú oddelené tabulátorom. V nasledujúcej tabuľke (Tab. 12) je príklad vety z korpusu, kde sa hrubo vyznačený text ukladá z celého riadku.

| Slovo z korpusu s kontextom | Koncovka s tagom |
|---|---|
| < Vilko/`SSms1:r`> s Majou doleteli na lúku s koňmi | lko `Sms1` |
| Vilko < s/`Eu7`> Majou doleteli na lúku s koňmi | s `E7` |
| Vilko s < Majou/`SSfs7:r`> doleteli na lúku s koňmi | jou `Sfs7` |
| Vilko s Majou < doleteli/`VLdpco+`> na lúku s koňmi | eli `V` |

***Tab. 12.*** *Vytváranie koncoviek s tagom*

Ďalej sa koncovky zatrieďujú. Celý korpus obsahuje 1 207 939 tokenov, z toho 1 210 sa vytriedilo v predchádzajúcej úprave (napr. znak „/“, ktorý bude program anotovať neskôr sám, bez slovníka). Pri takom vysokom čísle je zrejmé, že obsahuje veľa duplicít, čo nám pomôže zistiť pre každú koncovku percentá výskytu pre daný slovný druh. Každej koncovke sa pridelí reťazec slovných druhov, ktorý obsahuje všetky tagy zo slov s danou koncovkou. Tagy sú znova oddelené tabulátormi. Tak dostaneme zoznam jedinečných koncoviek spolu so zoznamom tagov. Týmto postupom sa vyberie 8 328 unikátnych koncoviek. Výsek zo zoznamu je nižšie.

| Koncovka | Nájdené slovné druhy (tagy) v korpuse |
|---|---|
| ône | Sfs2 Sfs2 Sfp1 Sfp1 Sfp1 Sfs2 Sfp1 |
| 764 | 0 |
| kňu | Sfs4 Sfs4 Sfs4 Sfs4 Sfs4 Sfs4 Sfs4 Sfs4 |
| ěst | % |
| @ | # # # # # # # # # # # # # # # # # # Z Z |
| drž | V Sfs1 Sfs1 V V V V V V Sfs1 Sfs4 Sfs1 |

***Tab. 13.*** *Zoznam koncoviek so slovnými druhmi*

Ďalším krokom je odstránenie všetkých koncoviek, ktoré nie sú zo slovenského jazyka. V niektorých koncovkách sa okrem cudzojazyčných slov (keďže databáza SNK pochádza z rôznych zdrojov) sa vyskytujú interpunkčné znamienka alebo čísla. Tieto znaky vieme nájsť v texte a samostatne otagovať, takže ich nemusíme mať aj v slovníku. Odstránenie nežiaducich koncoviek prebieha tak, že sa každý znak koncovky porovná s písmenami slovenskej abecedy. Ponechajú sa iba také koncovky, ktoré obsahujú všetky znaky iba zo slovenskej abecedy. Výsledný počet koncoviek získaných zo SNK je 6 887 [83].

Posledný krok pri vytváraní slovníka je určenie pravdepodobnosti výskytu. Pre každú koncovku spočítame tagy. Tagy pre rovnaký slovný druh ale rôzne gramatické kategórie berieme samostatne. Ku každému tagu v koncovke vypočítame jeho percentuálne zastúpenie. Hodnota je určená s presnosťou na tri desatinné miesta (v prípade percent na jedno). Všetky tagy danej koncovky zoradeníme od najvyššej po najnižšiu pravdepodobnosť. Takto zoradené koncovky spolu s percentuálnym ohodnotením tagov predstavujú finálnu podobu slovníka (Tab. 14).

| Koncovka s tagmi na výstupe z programu | |
|---|---|
| - upé | Sns6(0.286) Anp1(0.143) Aip1(0.143) Afp4(0.143) Ans1(0.143) Sns2(0.143) |

***Tab. 14.*** *Príklad koncovky s tagmi*

### 5.4.3 Vytváranie štvorfonémového a päťfonémového slovníka

Trojfonémový slovník koncoviek tvorí základ pre určovanie slovných druhov, z ktorého bude modul vychádzať. Pre presnejšie a menej chybové určovanie slovných druhov a gramatických kategórií je lepšie ak použijeme viac slovníkov (štvor- a päťfonémový). Tie použijeme v prípade nedostatočnej percentuálnej úspešnosti trojfonémového slovníka koncoviek.

Tieto dva slovníky vytvoríme analogickým spôsobom ako trojfonémový slovník. Na začiatku z celého korpusu vybereme štvorfonémové (päťfonémové) koncovky. Oproti trojfonémovému slovníku zostal počet koncoviek nezmenený, zmenili sa iba dĺžky. Potom vytvoríme zoznamy unikátnych koncoviek spolu s priradenými slovnými druhmi. Počet unikátnych štvorfonémových koncoviek je 26 279 a päťfonémových až 52 700. Vyčleneníme koncovky, ktoré obsahujú aj iné znaky ako slovenské písmená. Tým klesne počet koncoviek v prípade štvorfonémových koncoviek na 24 242 a v prípade päťfonémových na 50 577.

Ďalej vychádzame z predpokladu, že keď sa všetky koncovky v trojfonémovom slovníku rozšíria (pridaním všetkých možných foném pre koncovku) dostaneme kompletný štvorfonémový slovník. Štvorfonémový slovník obsahuje teda redundantnú informáciu pri tých koncovkách, ktoré sú určované s viac ako 99% pravdepodobnosťou v trojfonémovom slovníku. Tieto nadbytočné skratky odstránime zo štvorfonémového slovníka nasledovne: kontrolujeme všetky trojfonémové koncovky v slovníku, a ak nespĺňajú požadovanú presnosť, tak v štvorfonémovom slovníku zostanú všetky koncovky, ktoré dostaneme rozšírením danej trojfonémovej koncovky. Ako posledný krok vypočítame percentuálne pravdepodobnosti koncoviek, ktoré sú v slovníku. Obdobným spôsobom zredukujeme koncovky z päťfonémového slovníka koncoviek. Tu sa koncovky porovnávajú so štvorfonémovými koncovkami. Výsledný počet v štvorfonémovom slovníku koncoviek je 19 516 a v päťfonémovom slovníku koncoviek 28 862 [83].

### 5.4.4 Prehľadávanie slovníkov

Pomocou troch vytvorených slovníkov vie modul určovať slovné druhy. Na jednoduchom príklade (Tab. 15) diplomová práca bude vysvetlené, ako prebieha určovanie.

Slovo „diplomová“

| 3-fonémový slovník | | |
|---|---|---|
| -ová | Afs1 | (0.438) |
| | Sfs1 | (0.369) |
| | Snp4 | (0.097) |
| | Snp1 | (0.080) |
| | V | (0.011) |
| | % | (0.004) |

| 4-fonémový slovník | | |
|---|---|---|
| -mová | Afs1 | (0.815) |
| | Sfs1 | (0.185) |

| 5-fonémový slovník | | |
|---|---|---|
| -omová | Afs1 | (1.000) |

***Tab. 15.*** *Príklad postupného prehľadávania slovníkov pre slovo „diplomová“*

V prvom slovníku sa nájde 6 rôznych kategórií pre koncovku „-ová“. Ani jedna z možností nespĺňa potrebnú 99% úspešnosť, hľadá sa rozšírená koncovka v druhom slovníku. V ňom nájde len 2 možnosti bez požadovanej presnosti. Tá je splnená až pri

päťpísmenkovom slovníku (koncovka „-omová“). Získaná morfologická informácia *Afs1* nám určuje, že slovo „diplomová“ je prídavné meno (A), ženského rodu (f), jednotného čísla (s), v prvom páde [83].

Slovo „práca“

| 3-fonémový slovník | | |
|---|---|---|
| -áca | Sfs1 | (0.488) |
| | Sis2 | (0.312) |
| | V | (0.140) |
| | Afs1 | (0.028) |
| | Sms4 | (0.023) |
| | Sms2 | (0.009) |

| 4-fonémový slovník | | |
|---|---|---|
| -ráca | Sfs1 | (0.827) |
| | V | (0.165) |
| | Sms4 | (0.008) |

| 5-fonémový slovník | | |
|---|---|---|
| práca | Sfs1 | (1.000) |

***Tab. 16.*** *Príklad postupného prehľadávania slovníkov pre slovo „práca“*

Podobným postupom ako pri predchádzajúcom prípade analyzujeme slovo „práca“ a hľadáme jeho morfologické parametre. Z nich vieme, že slovo „práca“ je podstatné meno (S), ženského rodu (f), jednotného čísla (s), v prvom páde [83]. Na tomto slovnom spojení vidno spojitosť gramatických kategórií medzi slovami. Táto vlastnosť môže byť využitá pri korekcii morfologickej informácie pomocou kontextovej metódy.

### 5.4.5 Implementácia modulu morfológie

Naším cieľom je vytvoriť taký modul morfologickej analýzy, ktorý pre TTS syntetizátor dokáže automaticky robiť morfologický rozbor textu určeného na syntézu. Rozbor má slúžiť ako pomôcka pre ostatné bloky, ktoré pracujú so spomínaným textom. Správne určené morfologické kategórie pomáhajú napríklad bloku skratiek, aby sa skratky správne identifikovali, následne rozvinuli a prečítali. K pôvodnému slovníku trojfonémovýových koncoviek pridáme štvor- a päťfonémové slovníky. Tieto slovníky sa využijú v prípade, že slovný druh nebude určený dostatočne presne.

Proces určovania je implementovaný nasledovne (Obr. 24). Po načítaní slov zo vstupného súboru XML sa načítajú aj slovníky koncoviek. Najskôr sa porovná, či vstupné slovo nie je interpunkčný znak alebo číslo. V prípade zhody sa mu priradí príslušná značka a pokračuje sa analyzovaním ďalšieho slova. V analyzovanom slove sa vyberú posledné tri písmená. Ak slovo obsahuje menej ako 3 písmená, doplnia sa pred slovo medzery. Vznikne trojfonémová koncovka, ktorá reprezentuje určované slovo. Ak koncovka pozostáva z písmen slovenskej abecedy, prehľadáva sa slovník, ale ak nie, volí sa určenie slovného druhu sekundárnym spôsobom, a to pomocou regulárnych výrazov. Na ich základe sa potom určí tag.

Pri zhode koncoviek s trojfonemovým slovníkom sa pridelený tag zapíše do pamäte. Okrem tagu sa zapíše aj percentuálna pravdepodobnosť. Pravdepodobnosť sa vypočítava pri zostavovaní slovníka z výskytu daného tagu pri koncovke. Ak je pravdepodobnosť nižšia ako stanovená hranica, prehľadáva sa štvorfonémový slovník. Ak sa nájde koncovka v slovníku, prepíše sa informácia o tagu. V prípade nedostačujúcej pravdepodobnosti sa vykoná rovnaká procedúra aj pre päťfonémový slovník. Ak ju nenájde ani v slovníku pre päťfonémové koncovky, použije sa informácia z toho slovníka. V prípade, že pravdepodobnosť je kedykoľvek väčšia, zapíše sa tag do výsledkov. Výsledok určovania tvorí skupina všetkých tagov spolu s ich pravdepodobnosťami výskytu v korpuse *r-mak-3.0* pri danej koncovke. Určením morfologickej informácie pre jedno slovo sa jeden cyklus končí a začína ďalší. Po skončení všetkých cyklov sa získaná informácia zapíše do výstupného XML súboru.

Výsledná podoba výstupného súboru XML modulu morfologickej analýzy je nasledovná:

```
<?xml version="1.0" encoding="utf-8"?>
<text>
<sentence val="Pre Veroniku">
<word val="Pre">
<morfology>
<expectation>
<pos>E</pos>
<form>
<case>4</case>
<probabilityF>0.998</probabilityF>
</form>
<form>
<case>7</case>
<probabilityF>0.001</probabilityF>
</form>
<probabilityE>0.999</probabilityE>
</expectation>
<expectation>
<pos>Q</pos>
<probabilityE>0.001</probabilityE>
</expectation>
</morfology>
</word>
<word val="Veroniku">
<morfology>
<expectation>
<pos>S</pos>
```

```
<form>
<gender>f</gender>
<nr>s</nr>
<case>4</case>
<probabilityF>1.000</probabilityF>
</form>
<probabilityE>1.000</probabilityE>
</expectation>
</morfology>
</word>
</sentence>
</text>
```

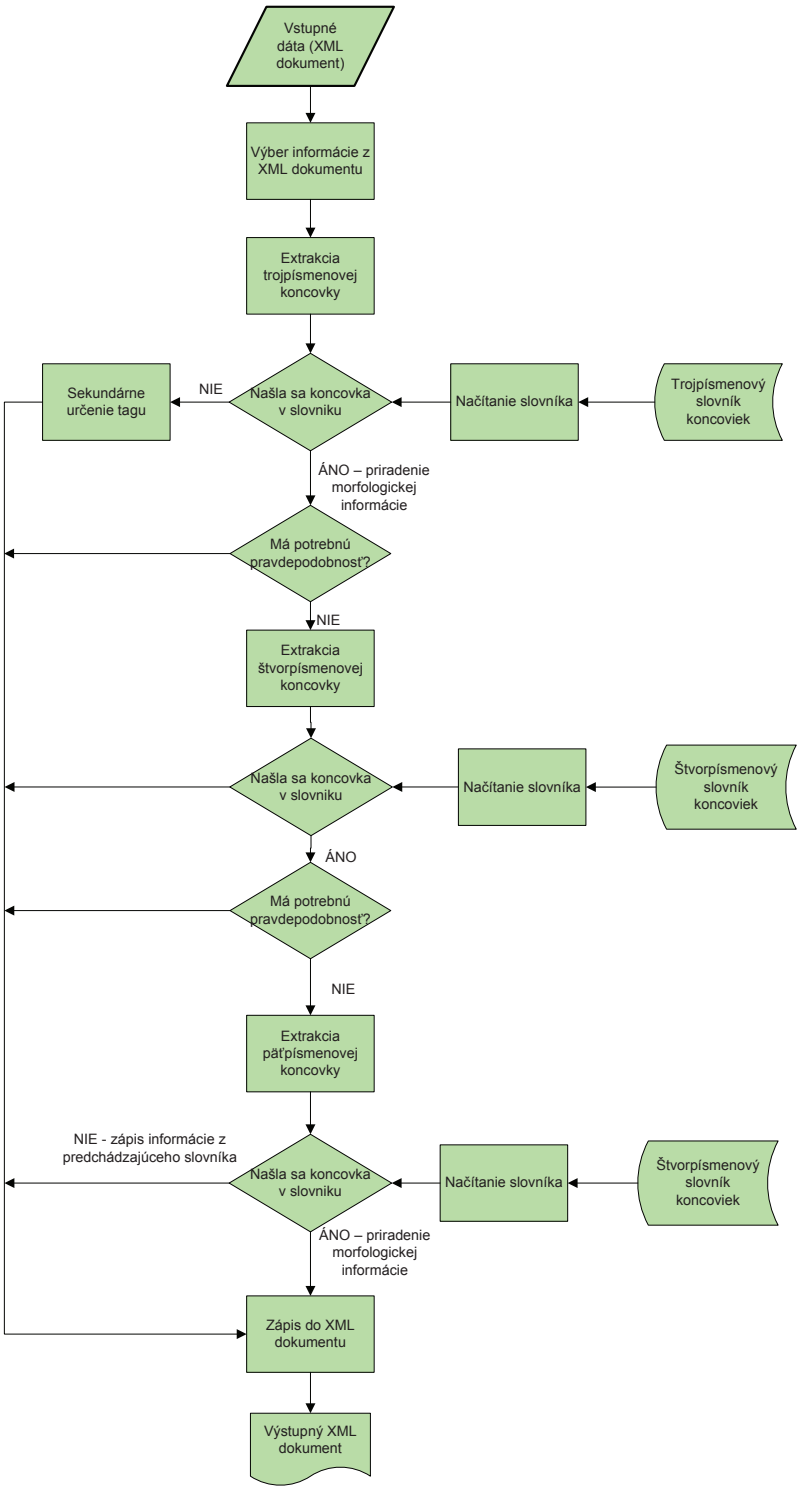
Na začiatku je na vstupe veta rozdelená na segmenty `sentence` a v nich sú jednotlivé slová vety rozdelené do tagov `word`. Výstupné informácie sa ukladajú do tagu `morfology`. Jeho dcérskym tagom je `expectation`, kde je uložená informácia o jednom slovnom druhu, ktorý bol nájdený pre danú koncovku. V tagu `pos` je uložené meno tohto slovného druhu a v tagu `probabilityE` je uložená jeho pravdepodobnosť. Ak algoritmus neurčí slovný druh jednoznačne, tak sa všetky možnosti uložia podľa pravdepodobnosti. Ak sa jedná o ohybný slovný druh (okrem slovesa), bude obsahovať aj gramatické kategórie v tagu `form` (podtagy pre rod (`gender`), číslo (`nr`), pád (`case`)). Ak nie je koncovka špecifická iba pre jedny gramatické kategórie, tak slovný druh obsahuje viacero tagov `form` zoradených tiež podľa pravdepodobnosti. Pravdepodobnosť form je uložená v tagu `probabilityF`.

### 5.4.6 Frázovanie

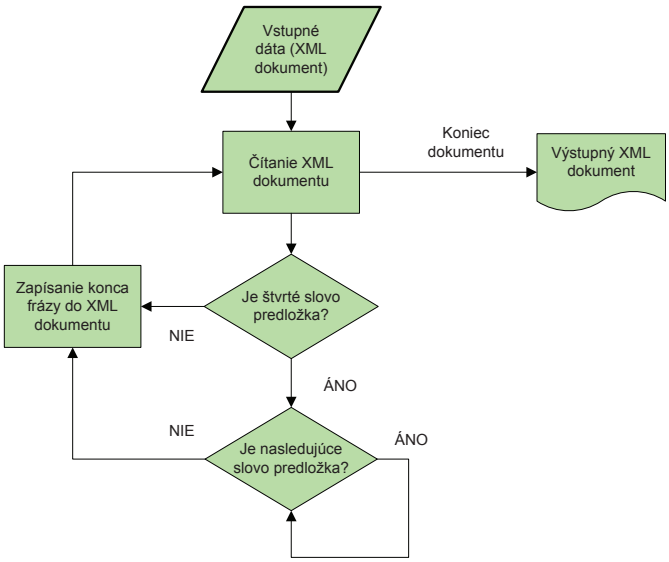
Frázovanie je tiež úlohou tohto modulu. Frázovanie, ako bolo spomenuté v kapitole 5.2.3.2, člení vetu na celky, ktoré sú tvorené vnútornými väzbami medzi slovami. Morfologickou anotáciou sa môžu niektoré tieto väzby určiť (napríklad predložka a slovo nasledujúce za ňou).

Vzhľadom na to, že ešte nie je vypracovaná komplexná štúdia týchto väzieb, určíme sme si na začiatok jednoduché pravidlo. Za frázu pokladáme každé 4 slová, ale v prípade, že na štvrtom mieste sa nachádza predložka, predĺži sa fráza o jedno slovo.

Na zápis frázy do XML dokumentu slúži tag `phrase`. Tento tag je uložený na úrovni tagu `word`. Vzájomnou dohodou medzi jednotlivými blokmi syntetizátora sa zvolí prázdna hodnota pre tento tag [83].




***Obr. 24.*** *Algoritmus určovania slovných druhov*




***Obr. 25.*** *Algoritmus frázovania*

Príklad takto rozčlenenej vety zapísanej v XML dokumente (tretia fráza obsahuje 6 slov, lebo na štvrtom a piatom mieste sú predložky *s* a *o*):

```
<?xml version="1.0" encoding="utf-8"?>
<text>
<sentence val="V čase keď chodil na základnú školu sedával v jednej
<phrase/>
<word val="V"><morfology><expectation><pos>E</pos><form><case>6</
<word val="čase"><morfology><expectation><pos>S</pos><form><gende
<word val="keď"><morfology><expectation><pos>O</pos><probabilityE
<word val="chodil"><morfology><expectation><pos>V</pos><probabili
<phrase/>
<word val="na"><morfology><expectation><pos>E</pos><form><case>4<
<word val="základnú"><morfology><expectation><pos>A</pos><form><g
<word val="školu"><morfology><expectation><pos>S</pos><form><gend
<word val="sedával"><morfology><expectation><pos>V</pos><probabil
<phrase/>
<word val="v"><morfology><expectation><pos>E</pos><form><case>6</
<word val="jednej"><morfology><expectation><pos>N</pos><form><gen
<word val="lavici"><morfology><expectation><pos>S</pos><form><gen
<word val="s"><morfology><expectation><pos>E</pos><form><case>7</
<word val="o"><morfology><expectation><pos>E</pos><form><case>6</
<word val="tri"><morfology><expectation><pos>N</pos><form><gender
<phrase/>
<word val="roky"><morfology><expectation><pos>S</pos><form><gende
```

```
<word val="starším"><morfology><expectation><pos>A</pos><form><ge
<word val="spolužiakom"><morfology><expectation><pos>S</pos><form
<phrase/>
</sentence>
</text>
```

### 5.4.7 Inteligentné učenie

Pre zdokonalenie vytvoríme podobný princíp učenia ako pre fonetickú transkripciu.

- Ručná oprava nesprávne určeného slovného druhu. Používateľské rozhranie syntetizátora umožní opraviť nesprávne určené slovo (napr. prepísaním do správneho tvaru buď ako text alebo v SAMPA abecede). Syntetizátor si zapamätá opravné slova uložením do súboru výnimiek. Na základe tohto slovníka sa modifikujú slovníky na určovanie slovných druhov.
- Automatické/inteligentné prehľadávanie možnosti alternatívnych slovných druhov. Ak syntetizátor neurčí slovný druh jednoznačne, ponúkne používateľovi na výber všetky možnosti výrazov pre dané slovo. Tieto možnosti syntetizátor určí na základe postupov na určovanie slovných druhov spomenutých vyššie. Teda zoznam tvoria všetky možnosti pre danú koncovku, ktoré sa nachádzajú v slovníkoch, aj keď ich pravdepodobnosť je nižšia ako nami určená hodnota 99%. Tieto možnosti sú zoradené od najpravdepodobnejšej až po koncovku s najnižšou pravdepodobnosťou.
- Oprava rovnakých slov (alebo rovnakého kmeňa slov) v texte. Pri syntéze dlhšieho textu sa dá predpokladať, že sa niektoré slová vyskytujú v texte viacnásobne. Preto v prípade akejkoľvek opravy sa prehľadá syntetizovaný text a ak sa nájde opravované slovo/slová/časť slova, označia sa (farebne vysvieti alebo podčiarkne). Používateľ bude môcť prejsť označenými slovami a rozhodnúť, či opraviť niektoré alebo všetky označené slová (implementuje sa funkcia *prepíš slovo* alebo *prepíš všetky slová*). Takáto funkcionalita významne urýchli opravu slov, keďže používateľ nemusí sám prehľadávať všetky slová a vyberať správne možnosti.

Podobne ako pri fonetickej transkripcii, najťažšou úlohou zostáva určiť, ako ukladať opravené slová do slovníka výnimiek a kedy z nich vytvoriť nové pravidlo. Do slovníka výnimiek ukladáme základný tvar a k nemu prislúchajúce tvary spolu s dôveryhodnosťou používateľa. Ak máme skupiny slov, pre ktoré vieme určiť pravidlo na určenie druhu na základe koncovky, tak rozšírime súbor pravidiel o nové pravidlo. Podobne je tu viac možností, ako vytvoriť pravidlá, manuálne alebo automaticky pomocou algoritmu. Obe tieto možnosti boli už podrobne rozobraté, takže sa im nebudeme viac venovať.

## 5.5 Voľba fonetickej transkripcie skratiek podľa kontextu

Čítanie skratiek v texte je ďalším samostatným problémom, ktorého vyriešenie prispeje k vytvoreniu dokonalej syntézy reči. Čítanie skratiek je individuálne aj u každého človeka, čo výrazne zvyšuje náročnosť pri učení syntetizátora. Skratka sa môže nahradiť slovom, môže byt vyhláskovaná alebo vyslovená ako slovo, pokiaľ je to možné. Pri prvej detekcii skratky zistíme, či sa skratka sama v texte nedefinuje (napr. Slovenská národná rada(SNR)). Ak sa nájde skratka, skontroluje sa jej výskyt v slovníku. Ak sa v slovníku nachádza, porovná sa s textom pred skratkou. V prípade zhody sa skratka iba vyhláskuje alebo prečíta (ak spĺňa podmienky vysloviteľnosti (5.5.1)). Ak sa v slovníku nenachádza, iba sa vyhláskuje alebo prečíta. V takomto prípade môže používateľ skratku alebo akronym zadefinovať (definovaniu a opravovaniu skratiek sa podrobnejšie venujú kapitoly 5.5.1 a 5.5.7). To, ako sa skratka prečíta ďalej, záleží aj od jej výskytu v texte. Ak sa nachádza v texte niekoľkokrát za sebou, netreba ju vždy nahradiť slovom, v takomto prípade môže byť prečítaný jej plný význam na začiatku, a potom ju stačí iba hláskovať. V prípade, že sa syntetizuje text dlhší ako iba jedna veta, je vhodné na základe kontextu určiť, k akej oblasti sa bude skratka vzťahovať (napr. ak sa bude čítať text o hokeji, bude zrejmé, že HC Košice nebude znamenať okres Hlohovec).

V úvode kapitoly boli popísané kroky predspracovania textu. Dôležitým krokom pri transkripcii skratiek je normalizácia. Aby proces normalizácie vstupného textu dosiahol čo najlepšie výsledky, mali by sa vykonať nasledujúce procesy:

- lexikálna analýza
- syntaktická analýza pre každú vetu, možnosť získania niekoľkých analýz
- syntaktická dvojzmyselnosť s cieľom vytvoriť čo najlepšiu analýzu
- sémantická dvojzmyselnosť s cieľom vybrať najlepší možný význam
- syntéza skloňovaných výstupov

V prvom kroku sa vytvára lexikálny model syntetizovanej vety. Postupne sa načítava text a symboly vo vete – tokeny sa zaznamenávajú. Na základe týchto dát sa určí lexikálna stavba vety. V kritických prípadoch sa tieto dáta určia pomocou špeciálnych postupov a algoritmov. Lexikálna analýza dokáže určiť slovné druhy, načíta do pamäte skratky a akronymy spolu s ich lexikálnymi informáciami. Potom, na základe slov pred a za skratkou, dokáže skratku správne interpretovať. Medzi zložitejšie problémy lexikálnej analýzy patrí napríklad určenie dvoch skratiek, ktoré nasledujú za sebou a sú od seba závislé (*pozn. red.* a *akad. mal.* by mali byť expandované do formy poznámka redaktora a akademický maliar, avšak samotné *red.* a *mal.* by mohli mať výsledný základný význam ako redukcia a maliarsky) [82].

Ak syntaktický parser neurčí skratku jednoznačne, na výstup pošle všetky možnosti. Najvhodnejší výsledok určí algoritmus, v ktorom sa počas syntaktickej analýzy bodujú

varianty „parse subtrees“. Na základe tohto hodnotenia sa vyberie najlepší výsledok. Syntaktická dvojzmyselnosť môže byť vhodná pri zisťovaní priority hľadanej formy.

V prípade sémantickej dvojznačnosti sa môže v strojovom preklade určiť viac významov daného tokenu. Význam sa určí na základe kontextu celej vety alebo aspoň časti vety. Napríklad výraz 5.p môže znamenať pri športovej hre päť bodov (z angl. five points). Ak analýza nezistí žiadne výrazy súvisiace so športom, tak sa bude výraz interpretovať ako piate poschodie.

Posledným krokom pri normalizácii textu je skloňovanie. Výsledná forma slova sa formuje na základe morfologickej a kontextovej analýzy. Tejto problematike sa ešte budeme venovať ďalej v kapitole 5.5.6.

Pri dokonalom systéme by sa dal do posledného kroku zapojiť aj samotný používateľ. V prípade, že sa skratka nepreloží správne v predošlých krokoch, tak ju poslucháč upraví sám a táto úprava sa ďalej spracuje (uloží sa do slovníka, vytvorí sa nové pravidlo.). Podrobnejšie bude systém učenia a spätnej väzby navrhnutý v kapitole 5.5.7.

### 5.5.1 Návrh modulu transkripcie skratiek

Úlohou tohto modulu je normalizovať zadaný text tak, aby všetky skratky boli expandované do správneho tvaru. Aby sa dosiahli čo najlepšie výsledky, sú navrhnuté tieto kroky:

- Proces spracovania textu pred syntézou nájde skratku. Ak je k dispozícii dlhší text, prejde sa celý a na základe počtu výskytov jednej danej skratky sa ďalej rozhodne, ako skratku čítať. Prehľadá sa slovník skratiek. Ak sa v ňom daná skratka nachádza, nahradí sa plným textom zo slovníka. Ak sa skratka vyskytuje v texte viackrát, nasledujúci výskyt sa vyhláskuje alebo prečíta – ak je vysloviteľný.
- Ak sa skratka v texte nachádza viackrát, môže sa v niektorých prípadoch dať aj vysloviť (napr. UNESCO). Je potrebné zistiť, či je skratka vysloviteľná. Slovo je v slovenskom jazyku vysloviteľné, ak obsahuje samohlásky (a, e, i, o, u) alebo slabikotvornú spoluhlásku (l, Ĺ, r, ŕ). V takomto prípade sa dá skratka prečítať a môže sa tým nahradiť hláskovanie alebo pri viacnásobnom výskyte sa môžu tieto možnosti obmieňať.
- V prípade, že sa skratka v slovníku nenachádza, overí sa jej vysloviteľnosť. Ak je vysloviteľná, postupuje sa ako v druhom bode. V prípade, že nie je ani vysloviteľná, tak sa iba hláskuje.
- Pre inteligentné čítanie skratiek nestačí mať iba úplnú databázu skratiek. Je potrebné aj poznať, k akému slovu sa skratka viaže a jeho parametre: slovný druh, rod, číslo a pád. Napr. majme vetu: „Kvalitné vzdelanie sa poskytuje na STU.“ Výraz „na STU“ treba čítať ako „na Slovenskej technickej univerzite“. Teda dokonalé čítanie skratiek je závislé od správneho určovania slovných druhov a ostatných parametrov.

- V používateľskom rozhraní syntetizátora umožniť používateľovi manuálny prepis skratky do SAMPA abecedy.
- Výber správneho tvaru skratky zo všetkých možností
- Akustická oprava výslovnosti.
- Oprava rovnakých skratiek alebo akronymov v texte.

Podrobný popis implementácie (súčasný stav) jednotlivých oblastí z predošlých krokov je v nasledujúcich kapitolách. Návrh implementácie bodov inteligentného učenia je v kapitole 5.5.7.

### 5.5.2 Lexikón skratiek a akronymov

Prepis skratiek a akronymov je založený na použití slovníka. V slovníku sú uložené v ideálnom prípade všetky skratky, v reálnom prípade čo najviac skratiek používaných v slovenskom jazyku.

Informácie v slovníku sú zaznamenané v správnom formáte a kódovaní. Najvhodnejšie sa ukázalo byť kódovanie Unicode. Toto kódovanie zahŕňa v sebe znaky všetkých svetových jazykov. Zaručuje správnu reprezentáciu pre každý znak ľubovoľného jazyka zobrazovaného v akomkoľvek operačnom systéme. Najefektívnejší spôsob predstavuje kódovanie UTF-8, ktoré pre všetky znaky z ASCII sady (písmená bez diakritiky, číslice, interpunkčné znamienka.) vyžaduje iba 1 bajt, pre znaky slovenského jazyka s diakritikou vyžaduje 2 bajty. Veľkosť súboru v slovenskom jazyku tak bude iba nepatrne väčšia, ale zachovajú sa tak všetky výhody Unicode kódovania [73].

Logická štruktúra slovníka je nasledovná: na začiatku sa nachádza informácia o hláskovateľnosti skratky, nasleduje skratka alebo akronym, na konci je uvedená základná forma expandovanej skratky (nominatív jednotného čísla pre podstatné mená, neurčitok pre slovesá).

Zásadnou požiadavkou na hlavný slovník je, aby obsahoval čo najviac skratiek používaných v slovenskom jazyku. Bolo potrebné rozhodnúť, akým spôsobom tento slovník vytvoriť. Ručné vyhľadávanie a prepis je veľmi zdĺhavý proces, a aj veľmi neefektívny. Ďalšou možnosťou je použitie softvéru. Pre účely tejto publikácie uvádzame príklad s použitím softvéru Bonito. Bonito je grafické rozhranie korpusového manažéra Manatee vyvíjaného na SAV. Program vyhľadáva a ukladá výrazy vyhovujúce vyhľadávaciemu pravidlu tvoreného regulárnymi výrazmi. Výsledky vyhľadávania sa pritom prehľadne zobrazia v okne manažéra Bonito. Program dokáže prehľadávať viacero rozličných korpusov (textových databáz). Rozdiely môžu byť vo veľkosti databázy, štýle a anotácii. Pri tvorbe slovníka využijeme ručne morfologicky anotovaný korpus *r-mak-3.0*, ktorý obsahuje 1 207 939 tokenov. V tomto korpuse je možné vyhľadávať slová podľa obsahu, podľa tagu alebo podľa základného tvaru slova – „lemy“. V nasledujúcej tabuľke (Tab. 17) sú uvedené príklady vyhľadávania v korpuse.

| Popis | Vyhľadávajúci výraz | Príklady |
|---|---|---|
| Troj-znakové slová | [A-Z][A-Z][A-Z] | LEV, SME, OSN |
| Všetky skratky | [tag="W"] | m, s, tzv, MV, SR |
| Troj-znakové slová zároveň skratky | [(word="[A-Z][A-Z][A-Z]") & (tag="W")] | KDH, SAS, OSN, USA |

***Tab. 17.*** *Príklady výrazov na vyhľadávanie v korpuse r-mark-3.0*

| r-mark 3.0 | | prim-4.0_ public-ALL | | r-mark 3.0 3-znakove | | r-mark 3.0 2-znakove | | r-mark 3.0 1-znakove | |
|---|---|---|---|---|---|---|---|---|---|
| **skratka** | **počet** | **skratka** | **počet** | **skratka** | **počet** | **skratka** | **počet** | **skratka** | **počet** |
| Sk | 304 | SR | 297760 | USA | 191 | SR | 136 | s | 136 |
| tzv | 255 | Sk | 259930 | STV | 68 | CZ | 44 | r | 119 |
| napr | 217 | s | 205445 | DVD | 47 | CD | 32 | t | 108 |
| USA | 191 | USA | 198122 | SMK | 42 | PP | 25 | j | 103 |
| J | 172 | M | 181040 | OSN | 41 | TV | 24 | m | 100 |

***Tab. 18.*** *Štatistická početnosť najpoužívanejších skratiek*

Výstup z programu Bonito zodpovedá vyhľadávaniu všetkých trojfonémových skratiek z korpusu Vyhovujúci tvar je označený červenou farbou a značkou KWIC. Skratky sa môžu vo výstupe programu opakovať. Z tohto dôvodu vytvoríme program, ktorý vyexportuje unikátne skratky a štatisticky vyhodnotí najpoužívanejšie skratky (Tab. 18).

Výstup zahŕňa okrem KWIC aj kontextové slová nachádzajúce sa po ľavej aj pravej strane od KWIC. Kontextový rozsah slov sa dá pri exportovaní do výstupného súboru nastaviť. keďže sa tvorila len databáza skratiek, zvolíme nulový rozsah [82].

Tabuľka (Tab. 18) obsahuje päť najpoužívanejších skratiek a ich početnosť v jednotlivých korpusoch: skratky v ručne anotovanom korpuse, skratky v hlavnom automaticky anotovanom korpuse, 3-znakové, 2-znakové a 1-znakové skratky v ručne anotovanom korpuse [82].

### 5.5.3 Hláskovanie skratiek a akronymov

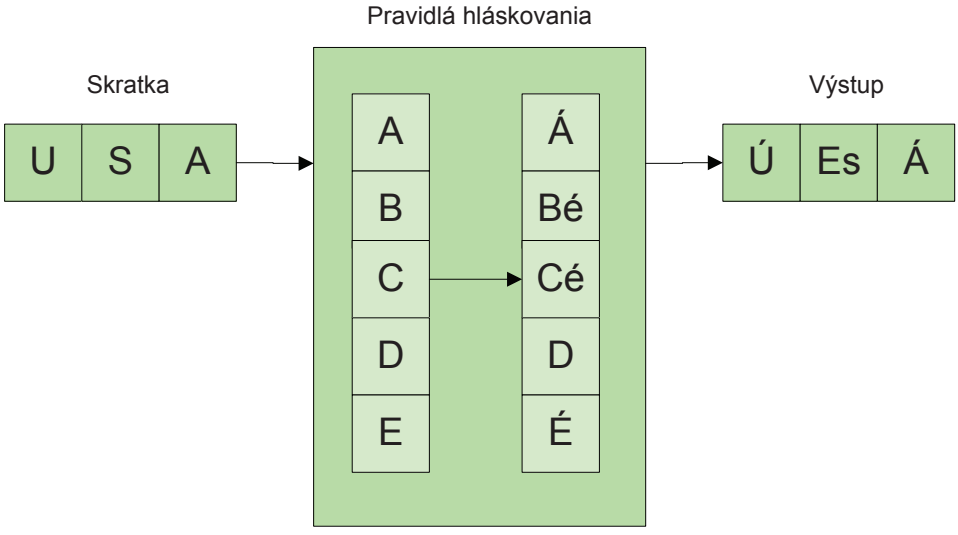
Čítanie skratiek má byť čo najprirodzenejšie. Ich hláskovanie k tomu výrazne prispieva a zvyšuje kvalitu prirodzenosti syntetizovanej reči. Človek pri čítaní dlhších textov s výskytom skratiek používa určitú variabilitu. Niektoré z nich prečíta, vyhláskuje alebo vysloví (ak je to možné). Takúto variabilitu sa snažíme vniesť aj do syntetizovanej reči.

Nie každý výraz je vhodné hláskovať. Je potrebné navrhnúť proces na určenie tejto možnosti. Napríklad skratka BMW sa môže počas normalizácie textu expandovať do tvaru „bé em vé". Ale FEI sa expanduje iba do plného tvaru „Fakulta elektrotechniky

a informatiky“, nikdy nie ako „ef é í“. Spôsobov, ako môže syntetizátor rozlišovať medzi takýmito prípadmi, je niekoľko. Jednou z možností je vytvoriť pravidlá, ale pri veľkom množstve skratiek by bola veľká chybovosť. Praktickejším riešením je uložiť túto informáciu priamo do slovníka. Ukážka časti slovníka je zobrazená nižšie:

```
…
„zost.“ „zostavil“
„zried.“ „zriedkavý“
„zried.“ „zriedkavo“
!“ZŠ“ „Základná škola“
!“UK“ „Univerzita Komenského“
!“MS“ „Majstrovstvá sveta“
```

Môžeme si všimnúť základnú stavbu slovníka. Na prvom mieste je nepovinná informácia, znak „!“. V prípade, ak sa na začiatku riadku znak „!“ nachádza, skratku je možné hláskovať, inak nie. Za týmto znakom nasleduje skratka alebo akronym vložená medzi znaky „“. Na treťom mieste je medzi znakmi „“ expandovaný výraz. Treba si ale uvedomiť, že hláskovanie skratky nie je vždy najlepšou voľbou. Hláskovanie skratky sa aplikovalo do modulu s jednoduchým pravidlom: hláskovať sa môže len skratka alebo akronym, ktorý sa v spracovávanom texte nachádza aspoň druhýkrát. Skratka sa nehláskuje, ak sa v syntetizovanom texte nachádza prvýkrát. Vtedy je prečítaný jej plný tvar. Je to z toho dôvodu, aby sme poslucháča oboznámili s celým významom skratky, ale pri viacnásobnom výskyte nemusí vždy počúvať dlhé významy skratiek [82].




***Obr. 26.*** *Príklad hláskovania akronymu USA*

Samotné hláskovanie prebieha tak, že každé písmeno skratky alebo akronymu nahradíme výrazom za jeho hláskované pravidlo. Tieto pravidlá sú zaznamenané v slovníku

nazvanom Pravidlá hláskovania slovenských písmen. Ilustračný príklad na hláskovanie akronymu USA je uvedený na obrázku [82].

### 5.5.4 Vysloviteľnosť skratiek a akronymov

Vyslovenie skratky v texte je ďalšou možnosťou, ako zvýšiť prirodzenosť čítania skratiek a akronymov v texte. Avšak nie každá skratka alebo akronym sa dá prečítať tak, ako je napísaná. Niektoré výrazy, ako napríklad FIFA, UNESCO, sa čítajú prirodzene tak, ako sú napísané, a ich hláskovanie by pôsobilo rušivo. Iné je to pri skratkách alebo akronymoch ako OSN, SDKÚ, HC a podobne. Tieto je vhodné vyhláskovať alebo expandovať do ich plného tvaru.

Dôležité je naučiť samotný syntetizátor, ako má rozlíšiť, kedy skratku ponechá v jej základnom tvare, teda ju vysloví, a kedy je lepšie ju expandovať alebo hláskovať. Vysloviť sa dajú výrazy, ktoré obsahujú slabikotvorné hlásky, medzi ne patria samohlásky, dvojhlásky alebo spoluhlásky r, ŕ, l, Ĺ, môžu byť slabikotvorné, ale len vtedy, keď sa nachádzajú medzi dvoma spoluhláskami (prst, hŕba). Do modulu sa implementovalo pravidlo, že výraz sa dá vysloviť vtedy, ak po každej spoluhláske s výnimkou slabikotvorných nasleduje samohláska, dvojhláska alebo slabikotvorná spoluhláska [82].

### 5.5.5 Expanzia skratiek a akronymov

Samotná expanzia skratiek pozostáva z troch základných bodov: správne určenie skratky zo vstupu, určenie vhodnej metódy na expanziu a uloženie výstupu (typ výstupu záleží od konkrétneho syntetizátora).

V prvom bode ide o samotné určenie, či sa jedná o skratku alebo akronym. Pri nesprávnom určení sa môže stať, že sa určí za skratku nesprávne slovo, a to spôsobí pri syntéze chybu. Preto je nutné spolupracovať s morfologickou analýzou textu. Morfologická analýza textu dokáže určiť jednotlivým slovám slovné druhy s určitou pravdepodobnosťou. V tomto module sa za skratku určí slovo, ktoré spĺňa jedno z nasledujúcich kritérií:

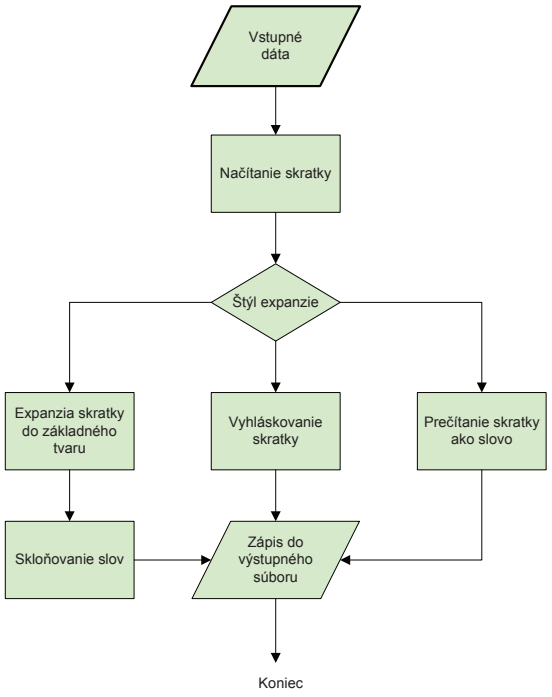
- morfologická analýza ho označí s viac ako 50% pravdepodobnosťou za skratku alebo akronym,
- morfologická analýza určí neisto slovný druh (napr. 29% podstatné meno, 28% sloveso, 20% prídavné meno a 23% iný),
- morfologická analýza nevie určiť slovný druh daného slova [82].

Na základe metód popísaných v predošlých kapitolách môžeme zhrnúť transkripciu skratiek do nasledujúcich tvarov:

- rozvoj do základného tvaru skratky (napr. OSN – Organizácia spojených národov)
- rozvoj do hláskovateľnej formy (napr. OSN – ó es en)
- ponechanie skratky v nezmenenej forme, ak je čitateľná (napr. UNESCO

– „unesko“)

Algoritmus transkripcie skratiek sa dá stručne popísať nasledovne (jeho znázornenie je na obrázku Obr. 27). Po načítaní skratky zo vstupného textu sa určuje spôsob expanzie skratky alebo akronymu. V tomto bloku rozhodovania sa určí najvhodnejší spôsob. Každá z možností transkripcie musí spĺňať vyššie opísané kritériá. Taktiež sa snažíme dodržať variabilitu pri čítaní, takže v prípade viacnásobného výskytu skratiek sa spôsoby čítania obmieňajú na základe splnených kritérií (rozvoj do plného tvaru kombinovaný s hláskovaním alebo s čítaním skratky). Štatistika výskytu skratky sa vykoná hneď pri načítaní vstupného textu. Pri prvotnom prečítaní sa skratka vždy rozvinie do základného tvaru. Je to dobré pre prvotné oboznámenie poslucháča so skratkou. V prípade, že sa nenájde skratka v slovníku, alebo sa v texte vyskytuje viackrát, tak sa môže vysloviť, alebo vyhláskovať.




***Obr. 27.*** *Algoritmus transkripcie skratiek*

Výsledný tvar skratky sa ukladá vo výstupnom súbore XML vo vopred dohodnutom tvare. Pôvodná skratka sa odstráni z elementu, v ktorom bola uložená, a nahradí sa výstupom z modulu do atribútu val elementu `word`. V prípade viacslovného výstupu skratky sa táto skupina slov zaobaľuje do nadradeného elementu `token`, ako zobrazuje nasledujúci príklad. V tomto prípade sa zaznamenáva v atribúte `val` elementu `token` aj pôvodný tvar skráteného slova alebo akronymu [82].

```
<token val="SNS">
<word val="Slovenská">
</word>
<word val="národná">
</word>
<word val="strana">
</word>
</token>
```

### 5.5.6 Skloňovanie skratiek a akronymov

Úprava tvaru ohybných slovných druhov sa nazýva deklinácia (skloňovanie). V slovenskom jazyku sa skloňuje tak, že k slovnému základu sa pridávajú rôzne prípony podľa pádu. Tak vieme meniť gramatický alebo slovný význam slova. V niektorých prípadoch skloňovania dochádza k zmene v kmeni slova (napr. genitív od žena je žien) [70]. Z tohto vidíme, že slová v slovenskom jazyku majú rôzne tvary, ktoré opisujeme gramatickými kategóriami.

Na uľahčenie určenia slovných druhov a kategórií, ako aj na určenie správneho tvaru slova pri skloňovaní, používame vzory skloňovania. Vzor skloňovania je slovo, ktoré reprezentuje určitú skupinu slov. Tvary slov pre jednotlivé pády sú tvorené množinou koncoviek. Spoločnou vlastnosťou je, že tvoria paradigmicky (paradigma je súbor tvarov ohybného slova vyjadrujúci systém jeho gramatických kategórií) odvoditeľné tvary podľa príslušného slovného druhu. Taktiež sa rovnako mení finálna podoba slovného kmeňa [71].

Proces skloňovania sa dá jednoducho opísať tak, že ku koreňu slova sa pridá prípona príslušného pádu a dostaneme výsledný tvar. Samozrejme, toto pravidlo neplatí úplne všeobecne. Existujú výnimky, pri ktorých treba uplatniť individuálne postupy skloňovania. Týchto prípadov je však málo, preto sa na začiatok budeme zaoberať iba väčšinovými prípadmi.

Pre správne skloňovanie musíme poznať koreň slova a príponu. Existuje viacero spôsobov na detekciu koreňa slova. Pri skloňovaní skratiek a akronymov sa využiva slovníková metóda, ktorá je náročnejšia na indexáciu, no umožňuje editáciu chýb.

Proces, pri ktorom sa slová redukujú na ich základný tvar, sa nazýva streaming. Základný tvar alebo koreň slova je taký tvar, ktorý je rovnaký pre všetky morfologické tvary. Na určovanie koreňa existuje viacero spôsobov (algoritmov), napríklad brute-force algoritmy využívajú vyhľadávacie tabuľky, algoritmy orezávajúce prípony alebo stochastické algoritmy, ktoré vytvárajú pravdepodobnostné modely [82].

V našom prípade v moduli na prepis skratiek implementujeme spôsob orezávania prípon. Pracuje tak, že sa načíta vstupný text a pre každé slovo sa overí, či obsahuje príponu z definovanej množiny prípon (napr. „-ovi“, „-och“, „-a“, „-u“). Ak modul príponu

nájde, odstráni ju zo slova a na výstup sa pošle koreň slova. Navrhnuté riešenie umožní určenie slovotvorného základu slov zo vstupného súboru. Týmto spôsobom získame slovník slovotvorných základov [82].

```
…
predseda  predsed
podpredseda      podpredsed
rozhodca  rozhodc
obranca   obranc
futbalista futbalist
```

Na znázornenom výstupe sa v prvom stĺpci nachádza pôvodné slovo, ktorého koreň slova zisťujeme. Všetky slová sú podstatné mená mužského rodu (v module je zatiaľ implementované iba určovanie pre mužský rod). V druhom stĺpci je uložený slovotvorný základ jednotlivých slov.

Ku korektnému skloňovaniu slov potrebujeme poznať okrem jeho slovotvorného základu aj príponu skloňovania (sufix). Rôzne slová majú rôzne prípony skloňovania. Jazykovedci definovali vzory skloňovania, ktoré nám pomáhajú určiť, aké prípony ktorému slovu prislúchajú. Návrh riešenia je podmienený tým, že každému slovu prislúcha práve jeden vzor skloňovania s náležitými príponami.

Vytvorenie slovníka spočíva v zaznamenaní čo najviac podstatných mien mužského rodu. Pre všetky tieto slová je určený vzor skloňovania. Na vyriešenie tohto problemu použijeme Slovenský národný korpus. Morfologickej anotácii podliehajú všetky jeho textové jednotky. Textové jednotky alebo tokeny sú reťazce znakov medzi dvoma medzerami ako aj znaky interpunkcie, pred ktoré sa pri spracúvaní textov v korpuse (pri segmentácii) medzery umelo pridávajú. Morfologická anotácia znázornená v tabuľke (Tab. 9) je aktuálna zo SNK, [82].

Riešenie na automatické určovanie vzorov podstatných mien využíva mohutnú databázu SNK a poznatok, že slová v rôznych pádoch majú v slovenčine inú príponu (sufix). Vhodnými pravidlami vieme určiť vzor skloňovania (samozrejme s určitou chybovosťou identifikovať vzor skloňovania skúmaného slova). Logika celého algoritmu sa dá popísať nasledovnými krokmi:

- Vyhľadanie a exportovanie čo najviac slov zo SNK v určitom páde do jednej databázy. V tomto kroku sa vytvorí 7 databáz slov (prvá v nominatíve, druhá v genitíve atď.). Pod pojmom databáza sa myslí určitá štruktúra dát uložená v pamäti, ktorá je prispôsobená na rýchle vyhľadávanie a prácu s textom.
- Implementácia kritérií a pravidiel.
- Načítanie všetkých 7 databáz a určenie výsledného vzoru konkrétneho slova.
- Zápis výslednej databázy slov a vzorov skloňovania.

Postupujeme metódou jednoznačného určenia vzoru na základe koncovky. Napr. nominatív jednotného čísla pre slovo „hrdina“ má ako jediný vzor sufix „-a“. Na základe tohto kritéria sa každému podstatnému menu v mužskom rode končiacemu sufixom „-a“

priradí vzor „hrdina“. Napr.: predseda, vodca, hokejista, kolega atď. Všetky navrhnuté kritériá a pravidlá triedenia pre mužský rod sú zobrazené v tabuľke (Tab. 19) [82].

| Vzor | Pád | Koncovka | Príklad |
|---|---|---|---|
| Chlap | 4. akuzatív | -a | chlap**a**, otc**a** |
| Hrdina | 1. nominatív<br>4. akuzatív | -a<br>-u | predsed**a**, vodc**a**<br>predsed**u**, vodc**u** |
| Dub | 6. lokál | -e<br>-u | chlad**e**, kokpit**e**<br>zemiak**u**, nádych**u** |
| Stroj | 6. lokál | -i | vankúš**i**, revolver**i** |

***Tab. 19.*** *Navrhované kritéria pre určovanie vzorov skloňovania*

„Z tabuľky teda vidíme, že na jednoznačné určenie vzoru chlap musí podstatné meno mužského rodu v 4. páde (akuzatíve) mať koncovku „–a“. Pre jednoznačné určenie vzoru hrdina máme k dispozícii dva prípady určenia: v 1. páde (nominatíve) s koncovkou „–a“ a v 4. páde (akuzatíve) s koncovkou „–u“ atď “ [82].

Znovu využijeme program Bonito, a to na vytvorenie pádových databáz. Text v SNK je anotovaný rôznymi meta-znakmi a každej textovej jednotke priraďujeme rôzne atribúty. V tagoch sú uložené morfologické informácie textu. Pojem anotácia v oblasti lingvistiky znamená pridávanie informácií k textom tvoriacim korpus. Nižšie je uvedený príklad exportovanej vety z Bonita, kde sú určené tagy pre podstatné mená. Bonito umožňuje vyhľadávanie v korpuse podľa viacerých kritérií, ktoré sa líšia v závislosti od korpusu. Nami použitý *r-mark-3.0* dokáže vyhľadávať slová podľa troch kritérií, a to `word`, `lemma` a `tag`. Vyhľadávanie prebieha pomocou tagu. Na získanie všetkých podstatných mien mužského rodu (životné aj neživotné) jednotného čísla v prvom páde (nominatíve) použijeme vyhľadávacie kritérium:

```
S.[mi]s1 – substantívum, ľubovoľnej paradigmy, mužský rod (životné alebo
neživotné), jednotné číslo, nominatív
```

Výstup z programu exportujeme do súboru s požadovaným kódovaním. Pomocou triediacich kritérií vygenerujeme výsledný zoznam podstatných mien a ich vzorov skloňovania.

Všetky navrhnuté slovníky (slovník slovotvorných základov a slovník určovania vzorov) zlúčime do jedného slovníka. Dôvodom je vyššia prehľadnosť a tiež fakt, že jedna informácia bez druhej nemá zmysel. Výsledný slovník obsahuje na prvom mieste podstatné meno v základnom tvare, slovotvorný základ a vzor skloňovania pre dané slovo. Výsledný formát slovníka je uložený v súbore CSV (comma-separated values) – údaje sú oddelené bodkočiarkou [82]. Príklad slovníka vidieť v nasledujúcich riadkoch:

```
…
antiglobalista;antiglobalist;hrdina
škodca;škodc;hrdina
občan;občan;chlap
zamestnávateľ;zamestnávateľ;chlap
…
```

### 5.5.7 Inteligentné učenie

Samostatnou možnosťou, ako zlepšiť čítanie skratiek, je implementovať spôsob učenia ako pri fonetickej transkripcii. Táto implementácia spočíva v nasledovných krokoch:

- Na používateľskom rozhraní syntetizátora umožníme používateľovi manuálny prepis skratky do SAMPA abecedy. Po prepise si používateľ znovu vypočuje syntetizovaný text už s opravenou skratkou (alebo viacerými skratkami). V prípade, že ide o úplne novú skratku, pridá sa skratka do slovníka skratiek. Ak ide iba o prepis skratky do správneho tvaru vo vete, tak sa uloží skratka do slovníka výnimiek pre určovanie slovných druhov, z ktorého sa po určitom čase doplnia nové pravidlá (pozri kap. 5.4.7).
- Výber správneho tvaru skratky zo všetkých možností. Rozhranie syntetizátora ponúkne používateľovi (napr. po kliknutí na skratku) všetky možnosti, čo sa pod danou skratkou skrýva (napr. PPP môže znamenať point-to-point protocol, purchasing power parity alebo public private partnership), prípadne aj rôzne tvary výslovnosti, napr. pre množné číslo, iný pád alebo rod. Možnosti sa zostavia na základe slovníka skratiek a slovníka výnimiek. Pri zoraďovaní možností sa skratka v základnom tvare kladie na posledné miesto, pretože zvyčajne výber možnosti nastáva v iných tvaroch ako základnom. Pri skratke, ktorá má viac významov (viď príklad vyššie), je vhodné označiť tie, ktoré sa nehodia. Stačí označiť jeden tvar a syntetizátor automaticky odstráni všetky tvary skratky zo zoznamu. Podľa obsahu textu treba do slovníka výnimiek uložiť poznámku, kedy je ktorý tvar skratky vhodný (napr. pri textoch z ekonomiky sa PPP rozvinie do tvaru projekty verejno-súkromného partnerstva). Používateľ si vyberie z ponúkaných možností správnu a môže si znovu vypočuť syntetizovaný text. Ak ani jedna z ponúkaných možností nevyhovuje, nastane možnosť ručnej opravy textu. V takom prípade sa nový tvar uloží znova do slovníka výnimiek.
- Akustická oprava výslovnosti. Používateľ si vypočuje zosyntetizovaný text a opraví skratku tak, že ju sám prečíta. Môže prečítať aj celú vetu, v ktorej sa skratka nachádza. Rozpoznávač vetu zapíše a syntetizátor deteguje opravenú skratku. Používateľ si môže znovu vypočuť syntetizovaný text už s opravenou skratkou a v prípade chyby sa proces zopakuje. Syntetizátor porovnaním slovníka skratiek a slovníka výnimiek zistí, či sa daná skratka už v záznamoch nachádzala. Ak nie,

doplní text do slovníka výnimiek.

- Oprava rovnakých skratiek alebo akronymov v texte. V prípade akejkoľvek opravy sa prehľadá syntetizovaný text a ak nájde opravovanú skratku/akronym, označí ich (farebne vysvieti alebo podčiarkne). Používateľ rozhodne či opraviť niektoré alebo všetky označené skratky/akronymy (implementuje sa funkcia prepíš slovo alebo prepíš všetky slová).

Znova je potrebné nájsť optimálny spôsob učenia a zvoliť, čo všetko ukladať do slovníka výnimiek. Je potrebné zvážiť, či ukladať iba rôzne tvary skratiek, alebo ukladať aj slovo pred skratkou a po nej, či vytvoriť nový slovník výnimiek pre skratky v základnom tvare a nový slovník pre iné tvary, ktorý sa potom spojí s výnimkami pre určovanie slovných druhov. Hľadáme aj potrebnú hranicu a určíme, kedy sa vytvorí nové pravidlo a kedy zostane skratka v slovníku výnimiek. Všetky tieto princípy a pravidlá určíme na začiatku učenia a v prípade potreby ich modifikujeme, alebo dopĺňame novými údajmi.

## 5.6 Modifikácia prozódie syntetizovaného textu

Vo všeobecnosti je syntetizovaná reč umelá a neprirodzená. Preto sa okrem samotnej syntézy snažíme nastaviť aj prozódiu (výšku, energiu, tempo) tak, aby sme sa čo najviac priblížili prirodzenej reči.

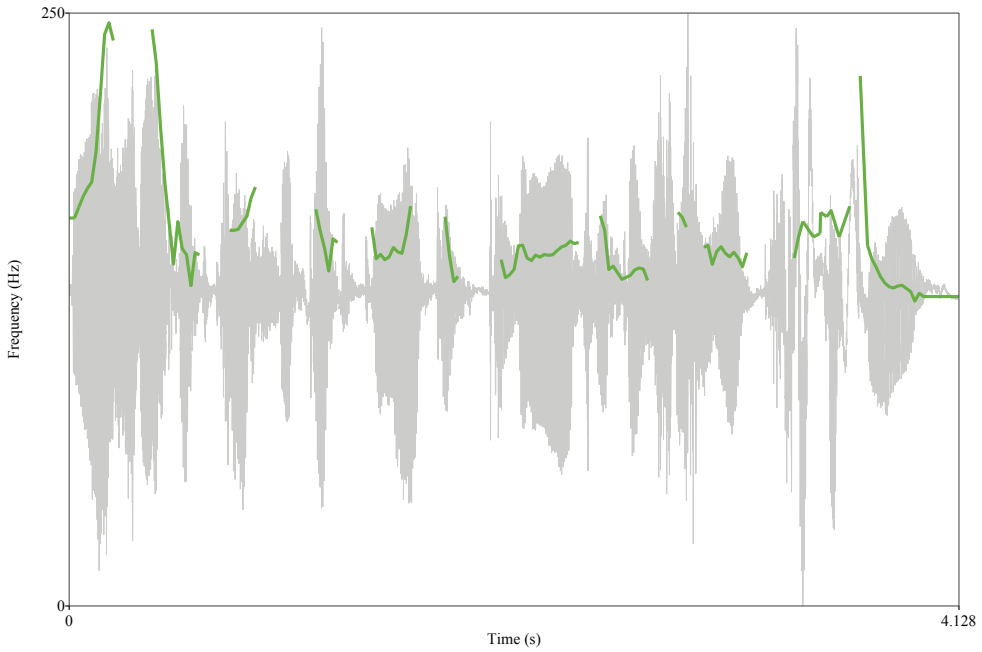
Pod zmenou alebo pohybom tónovej zložky reči rozumieme melódiu reči. Môžeme ju definovať aj ako zmenu výšky hlasu v časovej oblasti. Z fyziologického pohľadu je hlas generovaný vo vokálnom trakte, preto je jeho sila závislá od objemu vzduchu v pľúcach. Preto je logické, že ku koncu vety bude mať hlas klesavú tendenciu. Slovenský jazyk rozlišuje tri typy melódií, a to uspokojivo končiacu, neuspokojivo končiacu a neuspokojivo nekončiacu. Melódie vety teda môžeme deliť nasledovne [78]:

- Podľa ukončenia melódie
  - › končiaca – signalizuje koniec vety, prípadne možnosť pokračovať ďalšou vetou
  - › nekončiaca
- Podľa ukončenia myšlienky
  - › uspokojivá – uspokojivo uzatvára myšlienku, necítiť potrebu pokračovať ďalšou vetou
    - uspokojivo končiaca – má klesavú melódiu (kadenciu)
  - › neuspokojivá – neuzatvára myšlienku a cítiť potrebu ďalšieho doplnenia
    - neuspokojivo končiaca – ukončenie vety, ale naznačuje nutnosť pokračovať ďalej (antikadencia). Najčastejšie sa vyskytuje pri zisťovacích otázkach – odpovedá sa na áno/nie. Od oznamovacích viet sa odlišujú iba zvukovo, nie v gramatickej forme. V písanej forme sa odlišujú otáznikom.

- Neuspokojivo nekončiaca – myšlienka je otvorená, ukončenie sa signalizuje následným pokračovaním vyhovorenia (semikadencia).

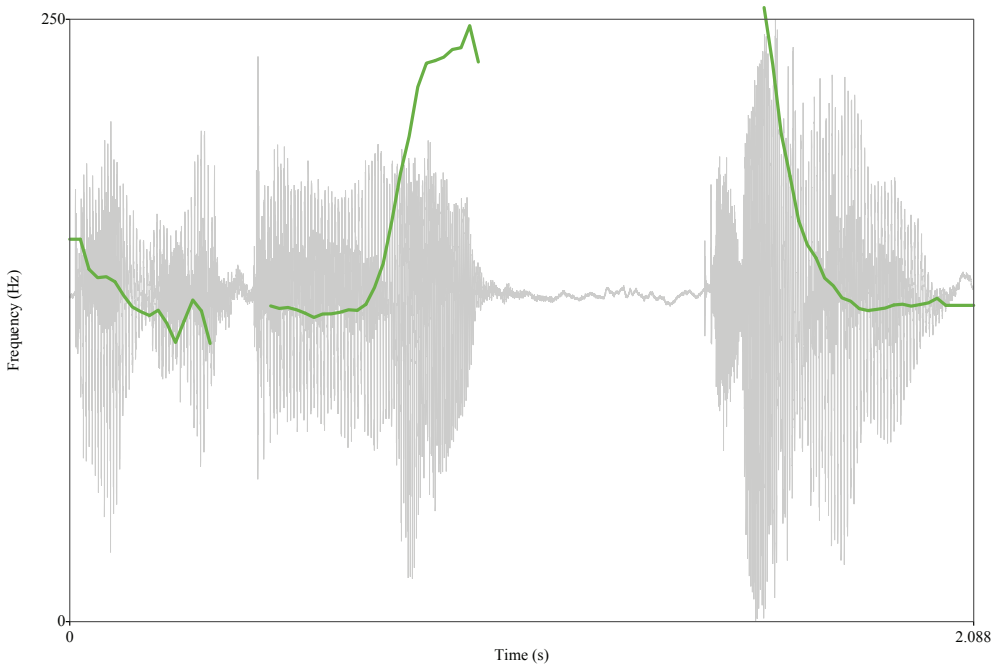
### 5.6.1 Typy viet v slovenskom jazyku

Pri štúdiu prozódie vychádzame zo Slovenského národného korpusu (SNK). Zo SNK extrahujeme dostatočný počet otázok na to, aby sa typy otázok dali rozdeliť na podskupiny. Kritériom pri rozdeľovaní je priebeh fundamentálnej frekvencie v čase. Výsledkom práce sú tri skupiny otázok. Otázka typu **Y** (áno/nie z angl. yes/no) je taká otázka, na ktorú sa odpovie áno alebo nie (Obr. 28). Ďalší druh otázky je typu **R** (Obr. 29), čo znamená rozlučovaciu otázku. Takáto otázka sa skladá z viacerých viet pospájaných spojkami či, alebo. Posledným typom je **Z** (Obr. 30). Sem patria všetky ostatné otázky, najčastejšie začínajú opytovacím zámenom a zisťujú určitú informáciu [79].

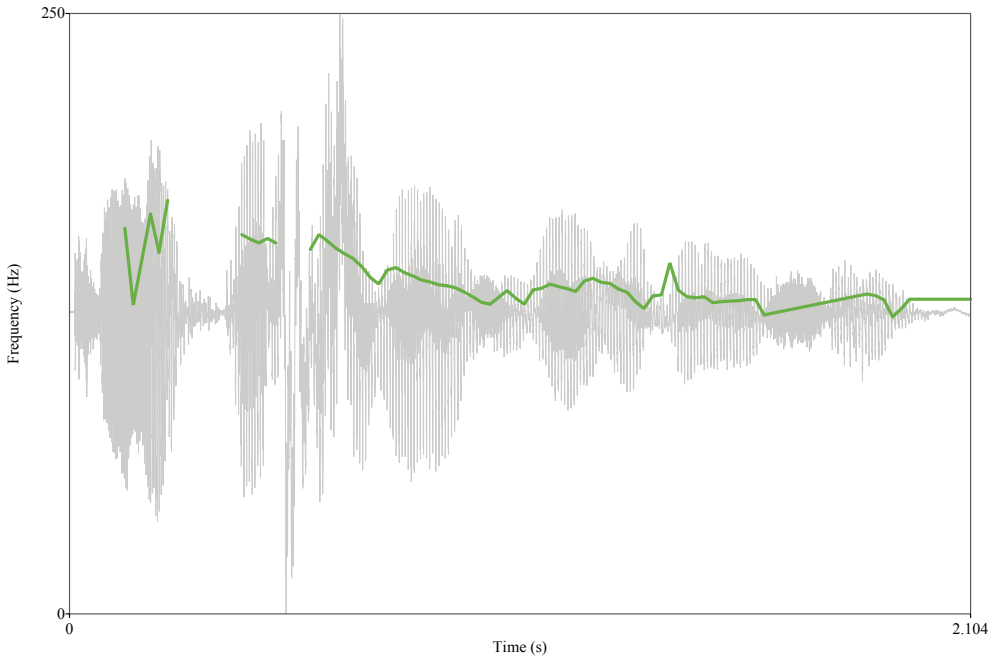



***Obr. 28.*** *Intonačný vrchol leží na začiatku a konci vety. (Veta: „Musí byť architekt na to, aby niečo dosiahol, politikom?“ (typ **Y**))*

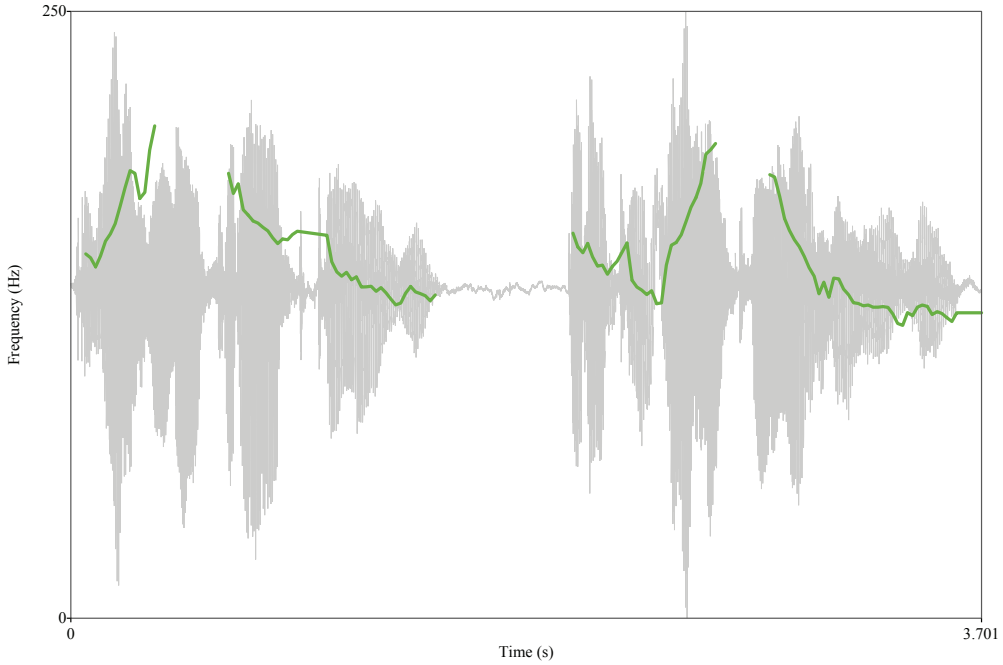
Pri oznamovacích vetách sa prozódia určuje rovno pre súvetia. Súvetia sa delia nasledovne: dôsledkové (spájame vety spojkou preto) (Obr. 31), dôvodové (spájame spojkou veď, totiž) (Obr. 32), odporovacie (spájame spojkou ale) (Obr. 33), stupňovacie (spájame spojkou ba) (Obr. 34), vylučovacie (spájame spojkou alebo) (Obr. 35) a zlučovacie (spájame spojkou a) (Obr. 36).



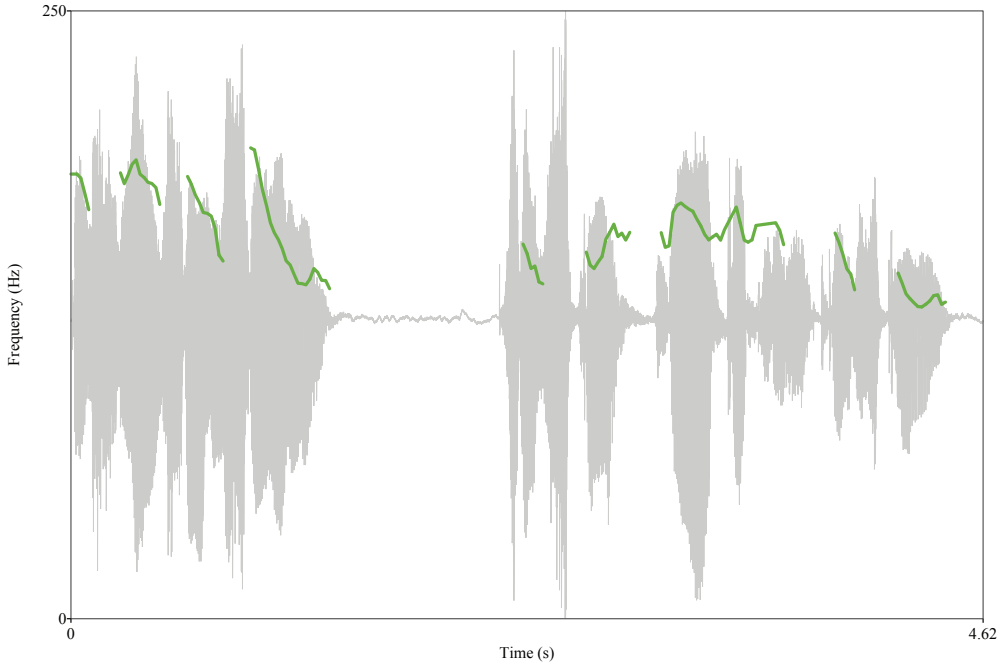

***Obr. 29.*** *Intonačný vrchol leží na slove „či“. (Veta: „Takže áno či nie?“ (typ **R**))*



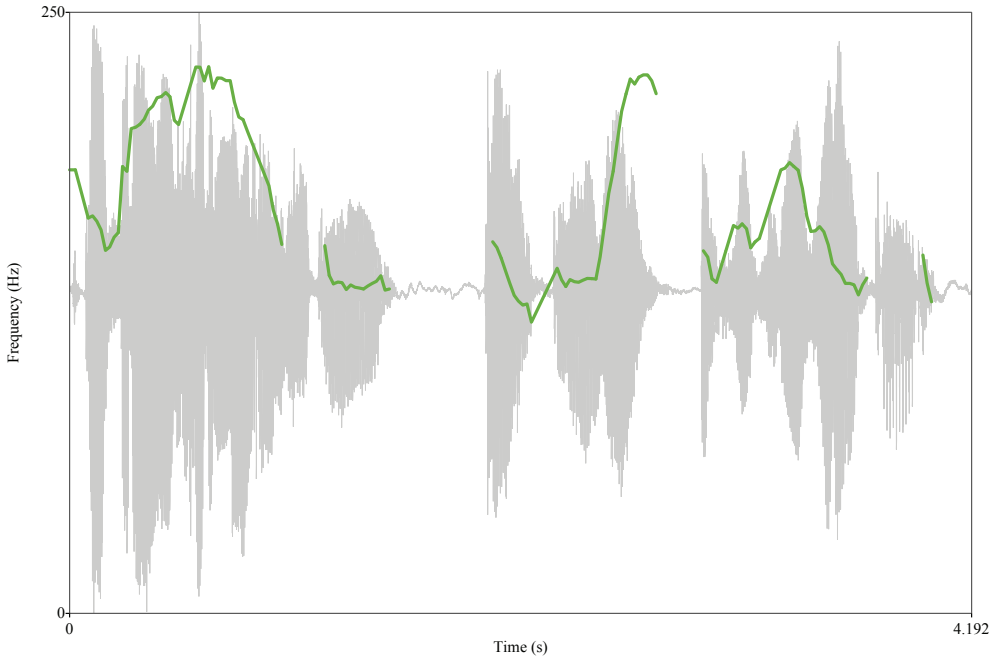

***Obr. 30.*** *Intonačný vrchol leží na začiatku opytovacej vety.(Veta: „Čo je podľa vás výhodnejšie?“ (typ **Z**))*



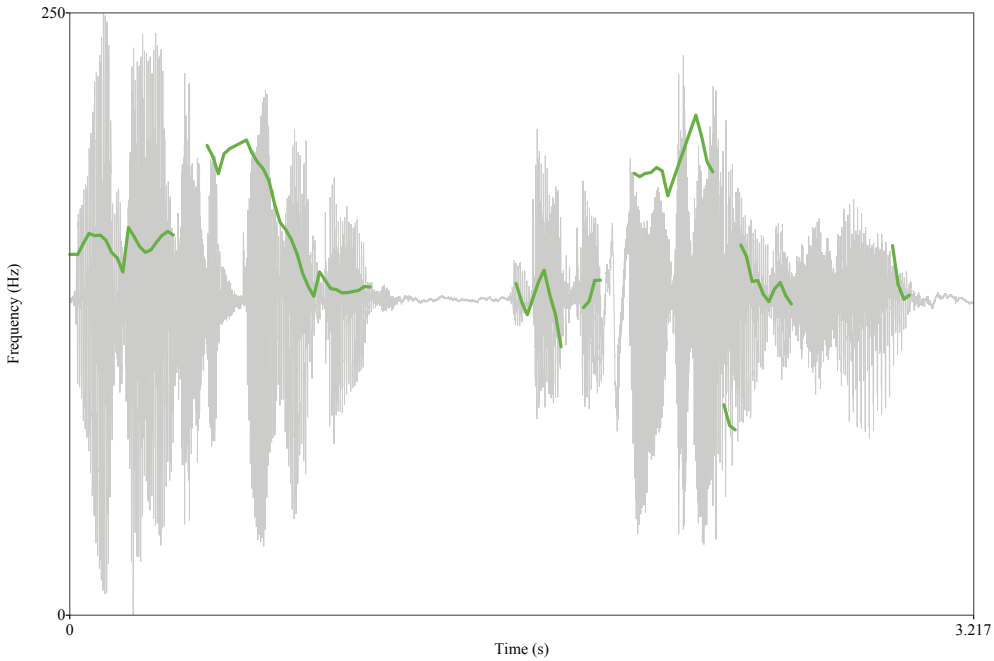

***Obr. 31.*** *Priebeh fundamentálnej frekvencie dôsledkového súvetia. („Lož má krátke nohy, preto ďaleko nezájde.“)*



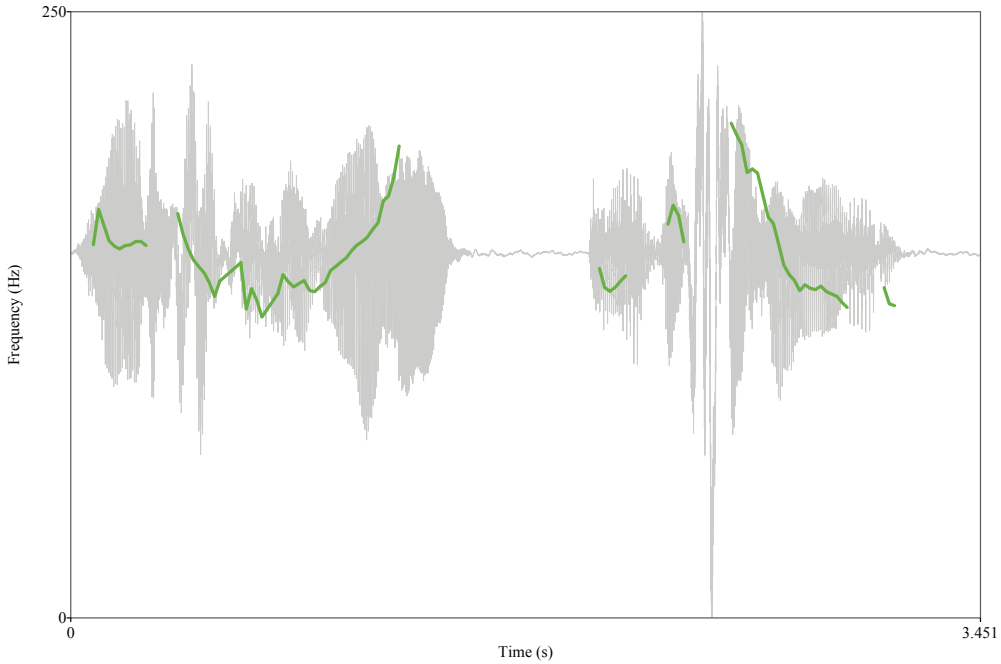

***Obr. 32.*** *Priebeh fundamentálnej frekvencie dôvodového súvetia. („Išiel som s nimi, cesta bola totiž klzká.“)*



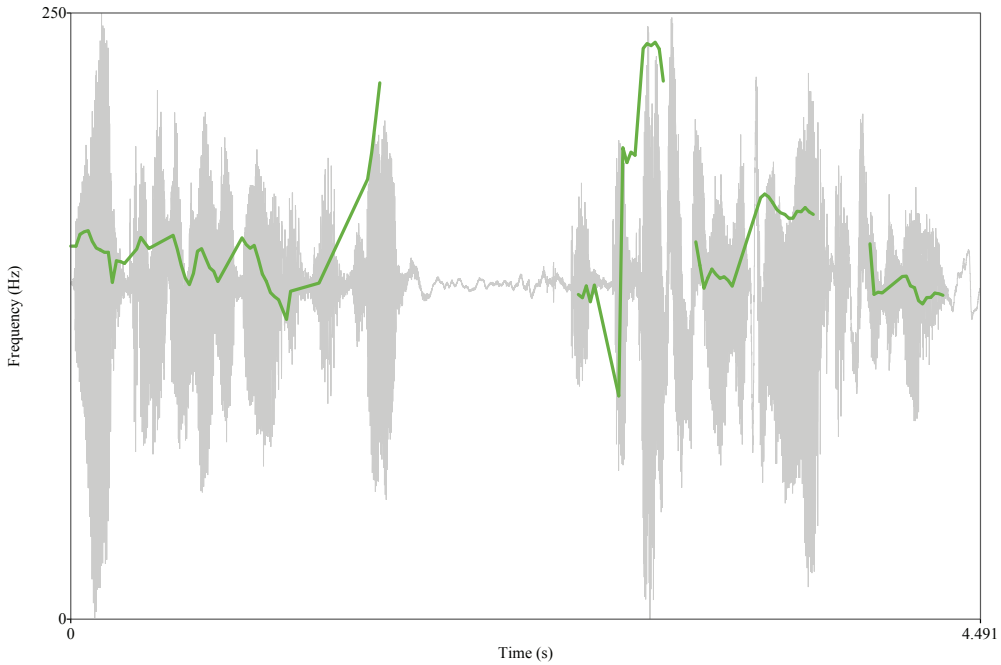

***Obr. 33.*** *Priebeh fundamentálnej frekvencie odporovacieho súvetia. („Dobré zrno ostane, ale plevy uchytí vietor.“)*




***Obr. 34.*** *Priebeh fundamentálnej frekvencie stupňovacieho súvetia. („Nedal si povedať, ba dokonca odišiel.“)*



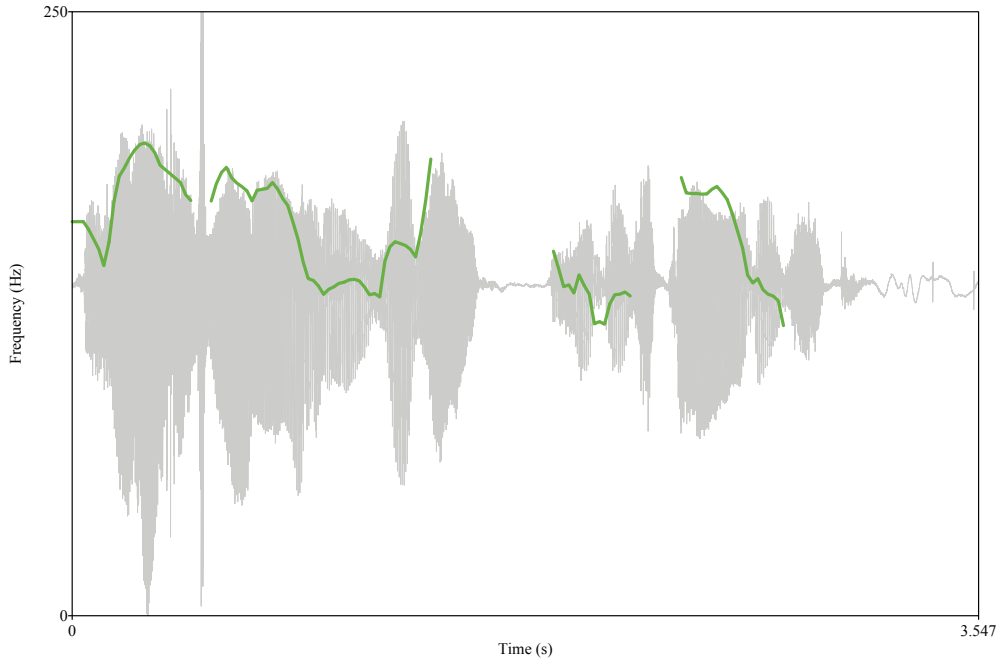

***Obr. 35.*** *Priebeh fundamentálnej frekvencie vylučovacieho súvetia. („Hlasujete za vojnu alebo ste za mier.“)*




***Obr. 36.*** *Priebeh fundamentálnej frekvencie zlučovacieho súvetia. („Najprv sa zasmiali deti a potom sa zasmial aj učiteľ.“)*

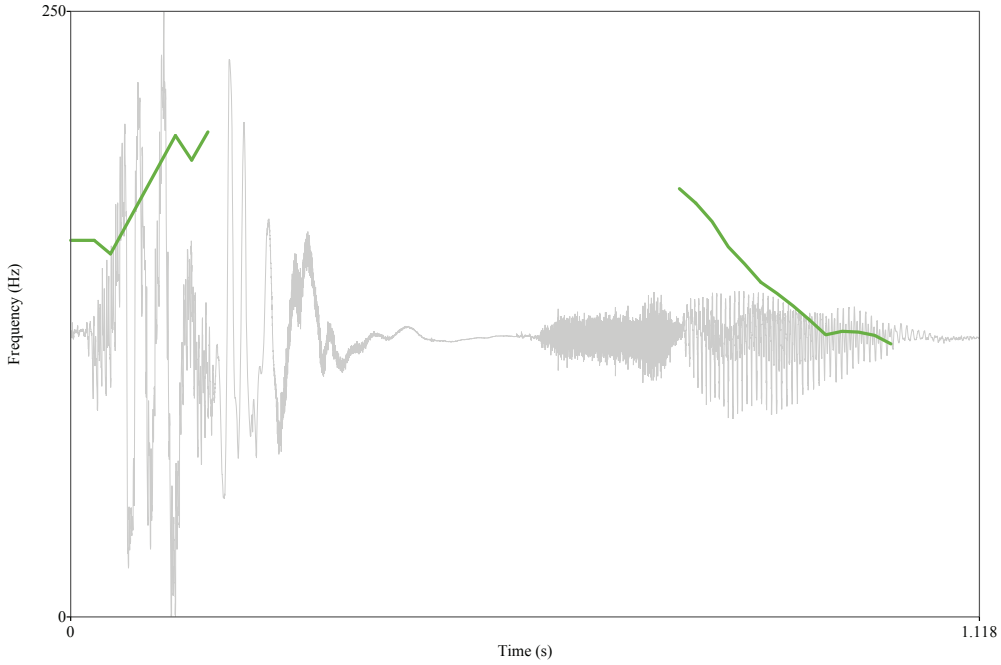
Na všetkých priebehoch vidíme klesavú tendenciu základnej hlasivkovej frekvencie. Taktiež si treba všimnúť, že vety sú oddelené prozodickou hranicou (čiarka alebo krátka pauza). Pri všetkých typoch súvetí, okrem vylučovacieho, sú obidva priebehy klesajúce. Počiatok nového stúpania znamená začiatok novej vety. Pri vylučovacom súvetí však nenastane pokles priebehu $F_0$, ale $F_0$ začne stúpať. Vyplýva to z podstaty súvetia, kde jedna veta má vylučovať druhú, preto výpoveď nemôže byť uspokojivo ukončená. Výsledkom tejto analýzy je, že pri oznamovacích vetách, resp. súvetiach, máme dva typy priebehov, a to **O** a **V**. **V** označuje vylučovacie súvetie a **O** všetky ostatné.

V prípade podraďovacích súvetí sa spájajú dve nerovnocenné vety. Vedľajších viet poznáme niekoľko druhov, presne toľko, koľko je vetných členov (podmetová, prísudková, prívlastková, predmetová, príslovková času, príslovková miesta, príslovková spôsobu, príslovková príčiny). Priebeh hlasivkovej frekvencie sa nemení v závislosti od typu vedľajšej vety, má vždy klesajúci charakter. Môžeme teda tvrdiť, že vedľajšia veta vždy iba dopĺňa hlavnú vetu.

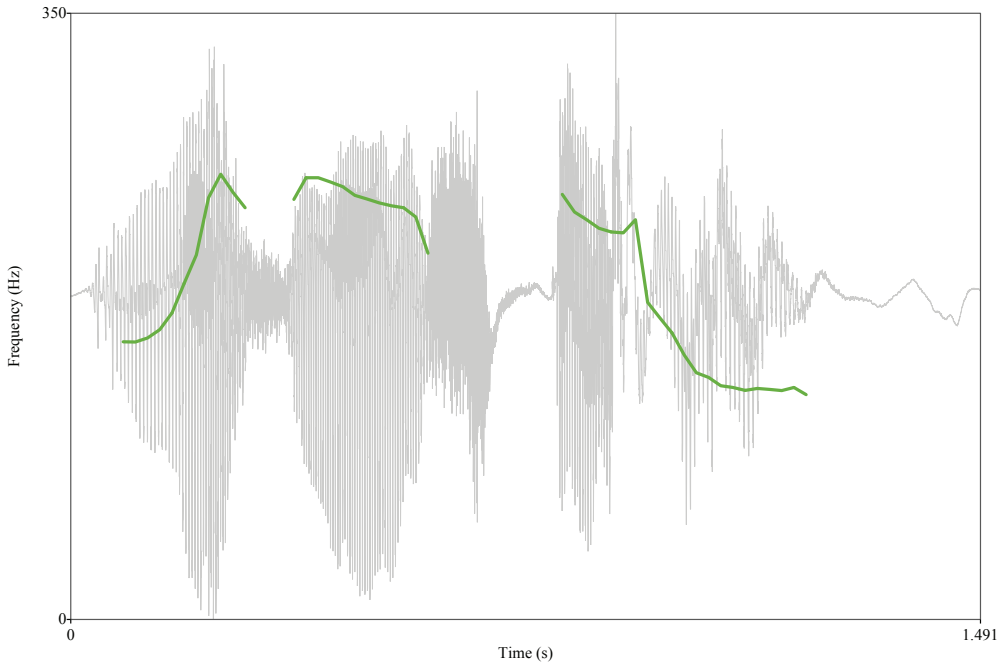



***Obr. 37.*** *Priebeh podraďovacieho súvetia s vedľajšou vetou predmetovou. („Žena oznámila mužovi, že dostal list.“)*

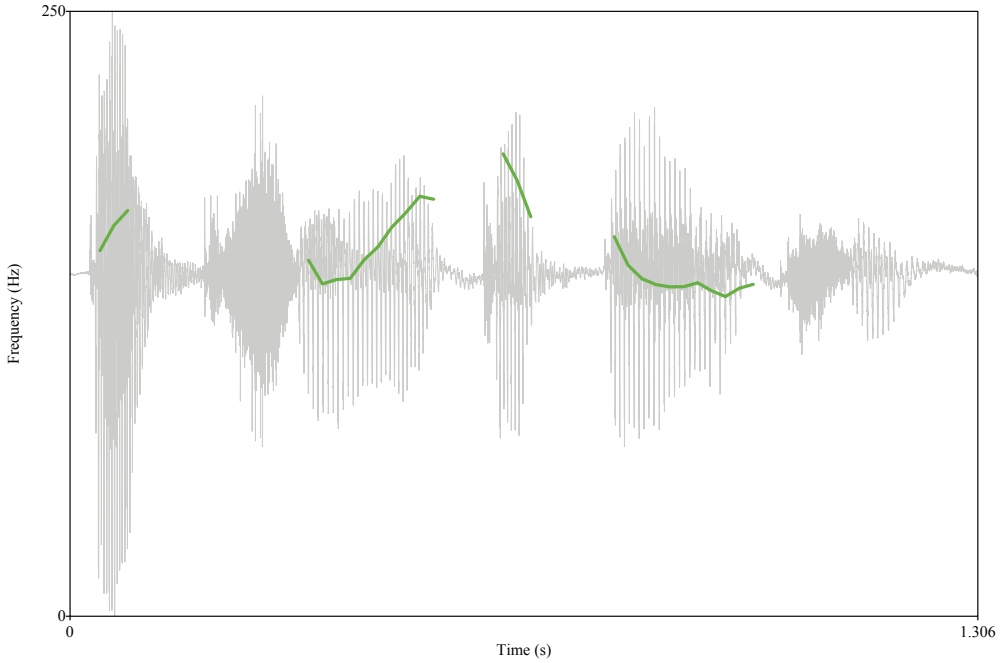
Pri skúmaní priebehov ostatných typov viet (rozkazovacia, želacia, zvolacia) sa ukázalo, že všetky majú klesavý charakter priebehu fundamentálnej frekvencie. Rozdiel sa prejaví až v prízvuku. Pri rozkazovacích vetách je veľmi dôležitý dôraz.




***Obr. 38.*** *Priebeh fundamentálnej frekvencie v rozkazovacej vete. („Poď sem!“)*




***Obr. 39.*** *Priebeh fundamentálnej frekvencie v želacej vete. („Nech len nespadne!“)*




***Obr. 40.*** *Priebeh fundamentálnej frekvencie vo zvolacej vete. („Tak sa mi to páči!“)*

Výsledkom predošlých analýz je delenie na tieto základné modely:

- 3 modely pre opytovaciu vetu:
  - typ **Y**
  - typ **R**
  - typ **Z**
- 1 model pre zvyšné typy (oznamovaciu, rozkazovaciu, želaciu a zvolaciu)

Pri súvetiach môžu byť tiež rôzne prípady priebehu. Pri detekcii čiarky sa budú rozlišovať tieto typy modelov [81]:

- priraďovacie súvetie
  - typ **V** – prvá veta má stúpavú melódiu a druhá klasickú melódiu oznamovacej vety
  - typ **O** – všetky vety v súvetí majú klesavú melódiu
- podraďovacie súvetie – celé súvetie má postupnú klesavú tendenciu kontúry základnej hlasivkovej frekvencie

## 5.6.2 Návrh zmeny prozódie

Pri prozódii a určovaní modelov je dôležité uvedomiť si úlohu prízvuku vo vete. Intonácia a jej priebeh je rôzna nielen pre rôzne typy viet, ale závisí aj od výšky hlasu, hlasivkovej frekvencie, tempa a intenzity. Najdôležitejšie prozodické vlastnosti ako rytmus,

tempo a melódia sa viažu na väčšie jednotky ako hlásky, a to na slová a vety. Tieto vlastnosti nazývame intonácia.

V slovenskom jazyku je prízvuk pevne určený na prvej slabike slova. Prízvuk je výraznosť jednej slabiky nad ostatnými slabikami a vytvára hranicu medzi slovami. Pri bežnej hovorenej reči nevznikajú medzery medzi slovami. Prestávkami sa oddeľujú väčšie celky ako frázy. Poloha prízvuku poskytuje tiež informácie o hovoriacom. Prízvuk dávame zvyčajne na slová, ktoré pokladáme za dôležitejšie než tie ostatné.

Úprava prozódie si vyžaduje:

- Zmenu základnej hlasivkovej frekvencie. Základná hlasivková frekvencia je iná pre mužov, ženy a deti. V snahe čo najvernejšie napodobniť vypovedania je nutné frekvenciu upraviť podľa originálu. Vplyv frekvencie je signifikantný najmä pre znelé časti reči, ktoré môžeme vďaka periodickému charakteru považovať za nosiča frekvencie vypovedania. Samotná úprava frekvencie reči je možná len na znelých fonémach. Keďže neznelé časti nemajú periodický charakter, nie je možné ich frekvenciu meniť. Základnú hlasivkovú frekvenciu reči meníme priblížením alebo oddialením jednotlivých periód znelej fonémy. Priblížením periód fonémy stúpa frekvencia v dôsledku skrátenia výslednej periódy, oddialením naopak klesá, ako to popisuje algoritmus PSOLA (kapitola 2.7). Zmena vzdialeností periód spôsobuje neželanú zmenu dĺžky fonémy. Preto je nutné adekvátne k zmene frekvencie fonémy pridať alebo odstrániť periódy, na ktoré aplikujeme rovnaký postup. Dĺžka nového signálu je nepriamo úmerná zmene základnej hlasivkovej frekvencii. Čím viac sa zvýši frekvencia, tým bude signál kratší z dôvodu skrátenia periód. Periódy pre skrátenie alebo predĺženie fonémy vyberáme zo stredu fonémy, pretože ten najmenej podlieha prechodovým javom prejavujúcim sa na hranici dvoch foném. V rámci úpravy prozódie reči vystupujú tempo reči a základná hlasivková frekvencia ako úzko späté faktory a nemôžu sa vnímať separátne. Modifikácia dĺžky by nastavila správne časovanie fonémy, ale následné prekladanie periód v záujme úpravy frekvencie by túto dĺžku mohlo výrazne zmeniť. Pri opačnom postupe, teda najprv vykonať prekladanie a následne doplniť alebo odstrániť potrebný počet periód z dôvodu úpravy dĺžky vyhovorenia, by sa zmenil počet periód vo výslednom signáli a tým by sa znehodnotilo vykonané prekladanie. Tento postup sa môže riešiť viacerými spôsobmi, medzi základné patrí postupná iterácia, pri ktorej sa najprv upraví dĺžka fonémy a potom sa vykoná prekladanie pre nastavenie požadovanej frekvencie. Podľa odchýlky od požadovaných hodnôt sa buď proces úpravy ukončí, alebo sa znova zopakuje. Toto riešenie je časovo náročné a nie je ideálne. Iné známe metódy sú napríklad lineárna a kvadratická interpolácia, ktoré daný problém riešia vhodnejším spôsobom pri zachovaní primeranej časovej náročnosti výpočtu [40].
- Zmenu tempa. Nezhoda dĺžok foném spôsobuje počuteľný a predovšetkým merateľný rozdiel medzi vypovedaniami. Napriek tomu, že každý človek má iné tempo reči a každé vyhovorenie aj tej istej osoby môže mať rôzne tempo, modifikácia na

konkrétne tempo je pomerne presná. Keďže znelé a neznelé fonémy majú odlišný charakter, musíme k modifikácii ich tempa pristupovať inými spôsobmi. Neznelá fonéma má šumový charakter, teda jej priebeh je takmer stochastický. Neznelé fonémy nemajú nijaké charakteristické vlastnosti, ktoré by sa mali brať do úvahy pri úprave tempa. Predĺženie trvania neznelej fonémy je teda pomerne jednoduché a znamená len zopakovanie, prípadne odstránenie časti fonémy prislúchajúcej danému predĺženiu alebo skráteniu. Modifikácia dĺžky vypovedania znelých foném však nie je tak triviálny problém ako modifikácia dĺžky neznelej fonémy. Znelá fonéma má periodický charakter, čo predznamenáva charakteristické znaky, ktoré musíme pri zmene dĺžky vypovedania brať do úvahy. Znelá časť reči nemôže byť predĺžená alebo skrátená jednoduchým vybratím vhodného počtu vzoriek a ich zopakovaním alebo odstránením. Úpravy znelej reči musia byť vykonané v súlade s periodickým charakterom signálu, akékoľvek úpravy teda musia byť vykonávané s celými periódami a ich násobkami. Predĺženie signálu v tomto prípade znamená pridanie počtu periód prislúchajúceho danému predĺženiu. Podobne skrátenie signálu musí byť vykonané odstránením prislúchajúceho počtu periód. Výber periód a operácie s nimi podliehajú mnohým pravidlám.

- Zmenu energie. Energiou v reči rozumieme napríklad prízvuk alebo intenzitu, s akou slová vyslovujeme. Prízvuk je komplexná prozodická kvalita reči odlišujúca niektoré slabiky prúdu reči od slabík iných. V slovenskom jazyku je prízvuk na prvej slabike slova. Osobitným typom prízvuku je vetný prízvuk, ktorý sa nazýva dôraz. Vetným prízvukom sa zvýrazňuje informačné jadro výpovede. Jej súčasťou je zosilnenie hlasu, zmena tónu a predĺženie slabiky. Podľa prevládajúcej zložky je prízvuk charakterizovaný ako dynamický (silový, výdychový), napr. v češtine, alebo ako melodický (tónový), napr. v starej gréčtine, čínštine, atď. V zložených slovách existuje okrem hlavného prízvuku aj vedľajší prízvuk. Podľa umiestnenia prízvuku sa rozlišuje prízvuk stály – viazaný na pevnú slabiku, napr. v slovenčine a v češtine na prvú, voľný – neviazaný na určitú slabiku, pohyblivý, presúvajúci sa v rôznych tvaroch toho istého slova z jednej slabiky na druhú, napr. v ruštine. Podľa charakteru úseku, v ktorom prízvuk diferencuje jednotlivé slabiky, sa rozoznáva prízvuk slovný, prízvuk taktový a prízvuk vetný. Ovplyvnenie tohto parametra v syntetizovanej reči nie je založené, ako vidieť vyššie, iba na logike alebo pravidlách, ale ide aj o subjektívny pohľad. Niekedy treba slová vo vete zvýrazniť, a vtedy ich chceme počuť aj zosyntetizované s väčším dôrazom. Jednou z možností na zmenu energie je vykresliť na rozhraní syntetizátora priebeh energie pre syntetizovanú vetu. Úpravou krivky (zníženie alebo zvýšenie) sa upraví hodnota energie a text sa znovu zosyntetizuje a prehrá s novými hodnotami.

### 5.6.3 Implementácia modulu na zmenu prozódie

Zmena prozódie pozostáva z dvoch samostatných krokov. V prvom kroku alebo module na základe správne určeného typu vety určíme hlasivkovú frekvenciu pre jednotlivé fonémy. Po správnom určení pripravíme zmena frekvencie. Model obrysu konkrétneho typu sa načíta do XML súboru a hodnota zmeny sa zapíše do elementu (podľa internej dohody nazývaného `PitchModel`). Táto hodnota hovorí, o koľko sa zvýši hodnota aktuálnej fundamentálnej frekvencie pre danú fonému.

Tento modul je zaradený ako prvý v reťazci modulov. Po ňom nasleduje modul výberu jednotiek z databázy. K dispozícii sú tri rôzne databázy a vyberá sa na základe toho, o koľko má byť fundamentálna frekvencia vyššia alebo nižšia (podľa hodnoty v `PitchModel`). Z databázy s najvyššou frekvenciou vyberáme pri náraste nad 150, z databázy s najnižšími hodnotami vyberáme pri hodnotách -110, pri zmene v intervale medzi spomenutými hodnotami vyberáme z databázy so strednými hodnotami frekvencií.

V druhom moduli, ktorý je zaradený na konci všetkých modulov, sa vykonáva úprava prozódie a následne syntéza. Táto zmena sa robí na úrovni viet. Zmena prozódie vyžaduje na vstupe tri informácie. V prvom vstupe sú popísané časové údaje (začiatok a koniec) pre jednotlivé fonémy, ktoré budú vo výslednom súbore wav. Tieto časové údaje sa načítavajú z elementov XML dokumentu, a to `timeStart` a `timeEnd`, ktoré sú uložené pod tagom bounds. Druhým vstupom je zosyntetizovaný súbor wav so všetkými parametrami rozvinutými ostatnými modulmi. Posledným vstupom je textový súbor s hodnotami základnej hlasivkovej frekvencie z elementu PitchModel. Pre nájdenie správnych hodnôt zvážime viacero faktorov, a to melodickú kontúru zosyntetizovaného súboru wav pred zmenou prozódie a melodickú kontúru daného typu vety. Ďalej zvážime optimálnu počiatočnú hodnotu základnej hlasivkovej frekvencie, aby bol prekryv čo najmenší. To preto, lebo zmena sy vykonáva pomocou algoritmu PSOLA a pri veľkých zmenách môže zvuk znieť neprirodzene. Z týchto dôvodov sa aj v prvom module vyberá z troch databáz. Takto sa už výberom z databázy priblížime k čo najbližšej hodnote $F_0$. Aby bol prienik modelu a dovtedy zosyntetizovanej vety maximálny, hľadáme tzv. optimálnu frekvenciu fopt. Nasledovné vzťahy platia v prípade, že dovtedy zosyntetizovaný signál má frekvencie s indexmi $f_1, f_2, f_3,., f_n$ a hodnoty modelu $\Delta f_1$ ', $\Delta f_2$ ', $\Delta f_3$ '… $\Delta f_n$ '. Výsledné hodnoty po korektnom umiestnení modelu sú hodnoty $f_{opt}+\Delta f_n$ ' pre jednotlivé fonémy.

$$f_1-\left(f_{opt}+\Delta f_1^{'}\right)+f_2-\left(f_{opt}+\Delta f_2^{'}\right)+\ldots+f_n-\left(f_{opt}+\Delta f_1^{'}\right)=0 \qquad (33)$$

pre $f_{opt}$ platí:

$$f_{opt}=\frac{1}{n}\left(\sum_{i=1}^{n} f_i-\sum_{i=1}^{n}\Delta f_i^{'}\right) \qquad (34)$$

Finálna podoba pre výstupný XML dokument tohto modulu vyzerá nasledovne:

```
<word val="monitor">
<syl val="mo">
<pho val="m">
<bounds>
<byte>
<byteStart></byteStart>
<byteMiddle></byteMiddle>
<byteEnd></byteEnd>
</byte>
<time>
<timeStart>0.49</timeStart>
<timeMiddle></timeMiddle>
<timeEnd>0.50</timeEnd>
</time>
</bounds>

<pitch>
<pitchModel>+10</pitchModel>
<pitchStart>123</pitchStart>
<pitchMiddle>126</pitchMiddle>
<pitchEnd>126</pitchEnd>
</pitch>
<energy>
</energy>
<duration>
</duration>
<pitchmarks>
</pitchmarks>
</pho>
```

Zo štruktúry vidieť, že nasledujúce úrovne ako slovo (`word`), slabika (`syl`), fonéma (`pho`), hranice (`bounds`) môžu byť bytové a časové – iba tie sa využívajú v tomto module, výška tónu (`pitch`), energia (`energy`), trvanie (`duration`) a značka vo výške tónu (`pitchmark`). Posledné tri údaje nie sú využívané modulom na zmenu prozódie. Dôležité sú elementy definované pre výšky tónu, a to:

- `pitchModel` je hodnota, ktorú má veta podľa určeného modelu nadobudnúť,
- `pitchStart` je hodnota základnej hlasivkovej frekvencie na začiatku fonémy,
- `pitchMiddle` je hodnota základnej hlasivkovej frekvencie v strede fonémy,
- `pitchEnd` je hodnota základnej hlasivkovej frekvencie na konci fonémy.

Hodnoty fundamentálnych frekvencií určených podľa vyššie uvedeného vzťahu sa uložia do súboru *F0values.txt*. Ukážka tohto zápisu do *F0values.txt* pre vetu obsahujúcu

18 foném:

```
0  115
1  125
2  135
3  115
4  125
5  125
6  125
7  120
8  118
9  120
10 125
11 122
12 125
13 165
14 185
15 205
16 205
17 215
18 210
```

### 5.6.4 Inteligentné učenie

Úpravu prozódie je tiež možné robiť pomocou spätnej väzby používateľa, a to nasledujúcimi spôsobmi:

- Zmenou cez rozhranie syntetizátora. Syntetizátor vykreslí časový priebeh syntetizovaného textu. Na grafe (alebo viacerých grafoch v prípade dlhého textu) bude možné meniť tempo, frekvenciu a energiu. V prípade opravy intonácie napr. na otázku si syntetizátor uloží túto zmenu do súboru. Pri viacnásobnom výskyte rovnakého typu sa vytvorí nové pravidlo (nový typ vety) na správne čítanie.
- Inteligentnou zmenou cez rozhranie syntetizátora. Syntetizátor ponúkne používateľovi možnosti na zmenu prozódie. Možnosti obsahujú modely podľa vyššie opísaných typov viet (všetky typy pre opytovacie, oznamovacie vety a pre súvetia). Ak napríklad syntetizátor určí nesprávny typ opytovacej vety, používateľ z danej ponuky vyberie správny typ, ktorý sa aplikuje na danú vetu. Ideálnym prípadom je zoradenie možnosti podľa pravdepodobnosti. Teda ak syntetizátor určí, že ide o opytovaciu vetu, ako prvé sú modely opytovacích viet. Priebeh pôvodnej fundamentálnej frekvencie vety sa zapamätá v databáze opráv. Pri viacerých záznamoch sa podobným priebehom s rovnakou opravou priradí pravidlo na správne určenie typu vety.

- Zmena cez akustickú spätnú väzbu. Vyhovorený text je zaznamenaný a syntetizátor spätne zosyntetizuje text už s parametrami rečníka. Na začiatku celého procesu je potrebné zistiť základnú hlasivkovú frekvenciu hovoriaceho, tempo reči, energiu. Tieto parametre sa potom porovnajú s parametrami databázy. Parametre databázy sa upravia na parametre originálneho vyhovorenia a prebehne syntéza. Takto zosyntetizovaný text má všetky charakteristické znaky hovoriaceho. Na tieto úpravy sa vhodným spôsobom ukazujú byť sínusoidálne modely.

## 5.7 Vzťah medzi parametrami sínusoíd pre rôzne tóny

Modifikácia reči podľa originálu pomocou sínusoidálneho modelu so šumom predstavuje úpravu parametrických hodnôt analýzy syntetizovaného signálu podľa analýzy originálneho signálu vypovedania. Kladie sa dôraz na požiadavku, že informácia prenášaná cez kanál musí byť čo najmenšia. Prenos celého výstupného súboru by znamenal najlepšiu dosiahnuteľnú spätnú interpretáciu signálu a súčasne by znamenal prenos nadbytočne veľkého objemu dát. Namiesto toho je možné na strane vysielača analyzovať syntetizovaný aj originálny signál, porovnať ich výstupné dáta a prenášať kanálom len ich rozdiel. Na strane prijímača sa použitím rovnakého syntetizátora vytvorí signál, vykoná sa jeho sínusoidálna analýza, ku ktorej sa pripočíta prenesený rozdiel a syntetizuje sa výsledný signál podobný originálu na strane vysielača.

Pre dosiahnutie najlepšieho pomeru kvality prenosu a veľkosti prenášaných dát je možné zanedbať, alebo naopak zohľadniť pri prenose niektorý z parametrov analýzy (tak ako to bolo popísané na začiatku kapitoly 5.6.2).

Na to, aby sme mohli meniť reč pomocou sínusoíd, musíme nájsť vzťahy medzi parametrami jednotlivých tónov. Pre jednoduchosť na začiatku analyzujeme hudobné tóny, pretože sú časovo menej premenlivé. Testy robíme na nahrávkach rôznych hudobných nástrojov ako flauty, klarinetu, saxofónu, trúbky a violy s fundamentálnou frekvenciou 220 Hz a 440 Hz pre tón *a*.

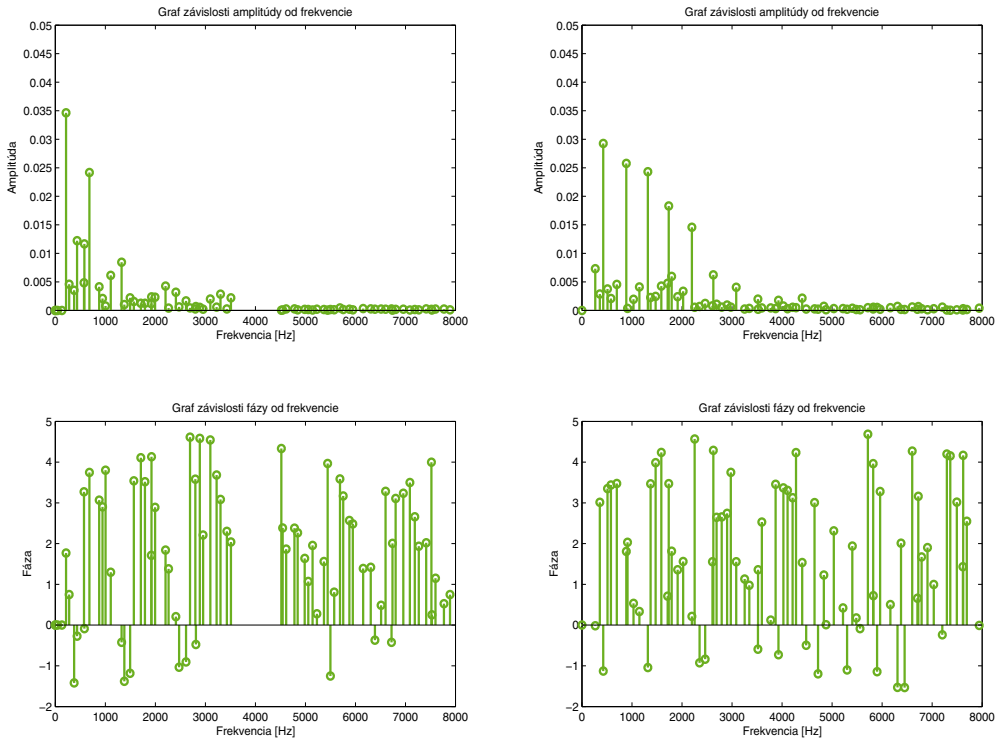
Prvá analýza sa robí pomocou SNM modelu, opísaného v kapitole 3.1. Výstupom analýzy sú parametrické súbory obsahujúce potrebné informácie o sínusoidách zvukového signálu. Príklad štruktúry parametrického súboru vidíme nižšie:

```
SAMPLERATE 16000
FRAME 0
FRAME_LENGTH 111
OCTAVE 0
SUBFRAME 0
Sine 220.147 0.042705 4.39523 1
```

```
Sine 309.884 0.00586416 0.905824 1
Sine 327.693 0.00266752 0.881216 1
(ďalšie dáta)
Sine 7609.66 8.2841E-05 1.40494 1
ENDSUBFRAME 0
ENDOCTAVE 0
ENDFRAME 0
FRAME 1
FRAME_LENGTH 148
OCTAVE 0
SUBFRAME 0
Sine 185.959 0.032042 2.2771 1
Sine 259.407 0.0243673 0.0664685 1
Sine 400.93 0.0110529 -0.743613 1
(ďalšie dáta)
Sine 7733.17 0.000312926 1.33608 1
Sine 7888.59 0.000227412 1.83114 1
ENDSUBFRAME 0
ENDOCTAVE 0
ENDFRAME 1
FRAME 2
FRAME_LENGTH 148
OCTAVE 0
SUBFRAME 0
Sine 186.862 0.032165 2.36601 1
Sine 259.838 0.0237913 0.166868 1
Sine 403.455 0.01116 -0.581405 1
(ďalšie dáta)
```

`SAMPLERATE` označuje vzorkovaciu frekvenciu analyzovanej nahrávky, `FRAME` a `ENDFRAME` označujú začiatok a koniec rámca, `FRAME_LENGTH` je dĺžka rámca, `SINE` zahŕňa parametre sínusoíd v poradí zľava doprava: frekvencia, amplitúda, fáza a znelosť. Postupnou analýzou všetkých vyššie uvedených hudobných nástrojov získame parametrické súbory, ktoré ďalej analyzujeme a spracovávame. V prvom kroku zostrojíme grafy závislosti amplitúdy a fázy sínusoíd od frekvencie (Obr. 41).

Analýzou všetkých priebehov zistíme, že v rámcoch je vždy viacero nosných sínusoíd. V týchto nosných sínusoidách je najväčší podiel energie, a preto nesú aj najviac informácie o zvuku alebo reči. Priebeh fázy sa mení náhodne, nedá sa odvodiť žiadna závislosť, má stochastický charakter. Amplitúda sa mení aj v závislosti od toho, ktorý rámec analyzujeme. Snažíme sa o štatistické vyhodnotenie všetkých tónov. Výpočtom určíme priemerné hodnoty jednotlivých amplitúd pre každú sínusoidu.

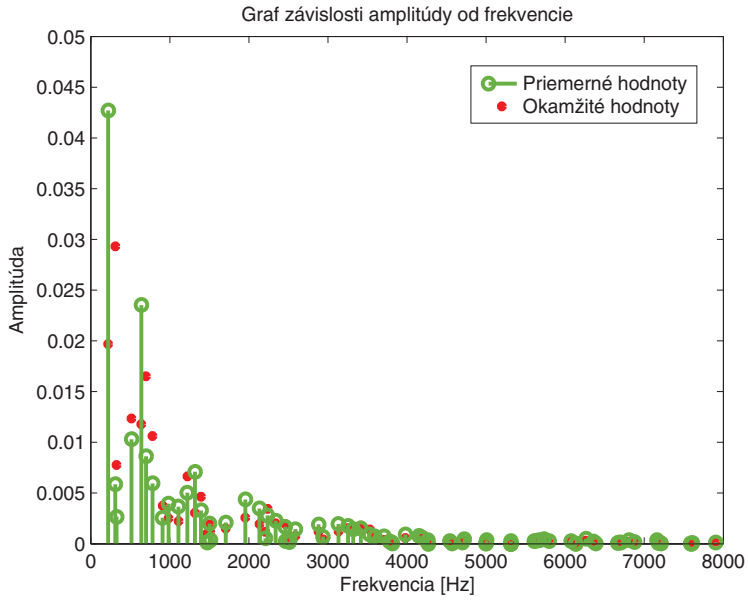



***Obr. 41.*** *Priebehy závislosti amplitúdy a fázy od frekvencie pre saxofón(220 Hz a 440 Hz)*

Analýzou všetkých priebehov zistíme, že v rámcoch je vždy viacero nosných sínusoíd. V týchto nosných sínusoidách je najväčší podiel energie, a preto nesú aj najviac informácie o zvuku alebo reči. Priebeh fázy sa mení náhodne, nedá sa odvodiť žiadna závislosť, má stochastický charakter. Amplitúda sa mení aj v závislosti od toho, ktorý rámec analyzujeme. Snažíme sa o štatistické vyhodnotenie všetkých tónov. Výpočtom určíme priemerné hodnoty jednotlivých amplitúd pre každú sínusoidu.

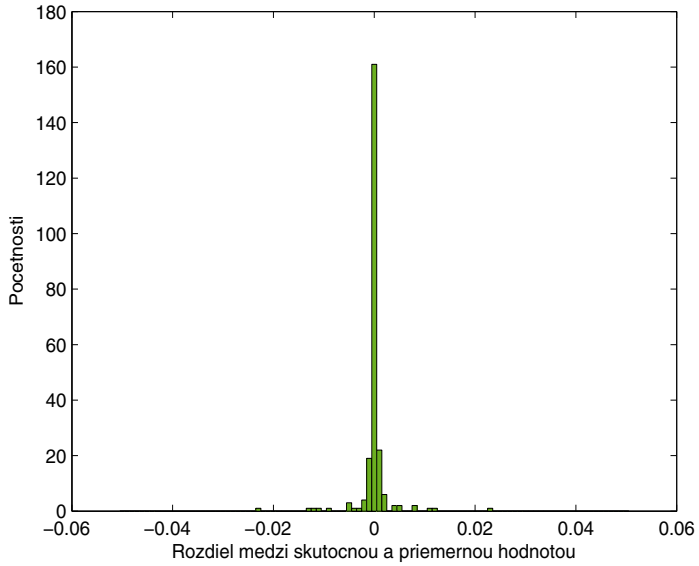
$$A_{avg_i} = \sum_{i=1}^{n} \left( \frac{1}{m} \sum_{j=1}^{m} A_{ij} \right) \tag{35}$$

kde $A_{avg_i}$ je priemerná hodnota amplitúdy *i*-tej sinusoidy, $A_{ij}$ je amplitúda *i*-tej sínusoidy v *j*-tom rámci, *n* je počet sínosíd a *m* je počet rámcov. Priemer počítame tak, že prvá sínusoida z prvého rámca prislúchala prvej sínusoide z ďalších rámcov atď. V ojedinelých prípadoch nastáava situácia, že sínusoidy nemajú rovnaký počet rámcov. Je to z dôvodu chybnej detekcie fundamentálnej frekvencie. Tieto chybné rámce preto pri výpočte premenných hodnôt nebereme do úvahy. Vypočítané aritmetické priemery priradíme k harmonickým frekvenciám analyzovaného zvuku (Obr. 42).



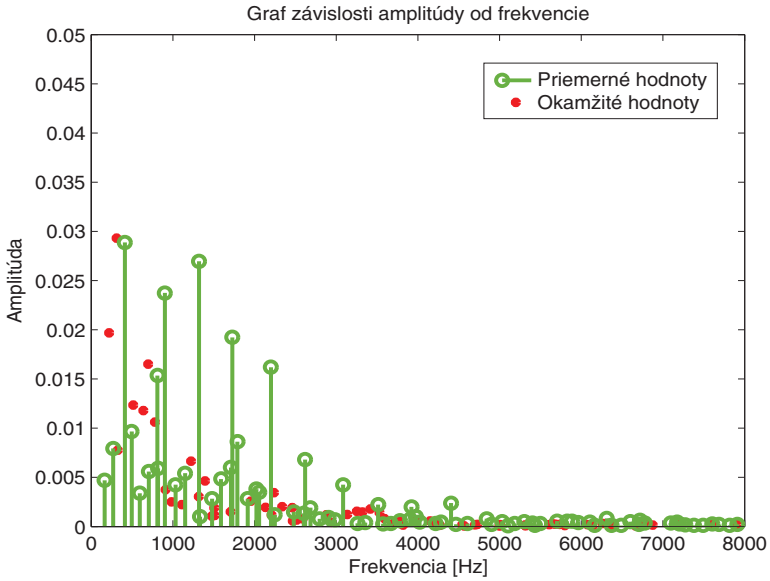

***Obr. 42.*** *Graf okamžitých a priemerných hodnôt saxofónu pre 220 Hz*

Po určení histogramov potvrdíme naše úvahy, teda hodnota rozdielov medzi priemernými a skutočnými hodnotami sa pohybuje okolo nuly (Obr. 43). Môžeme teda tvrdiť, že priemerné hodnoty sa veľmi neodlišovali od okamžitých hodnôt. Chybné rámce sa dajú potom nahradiť rámcom s priemernými hodnotami.

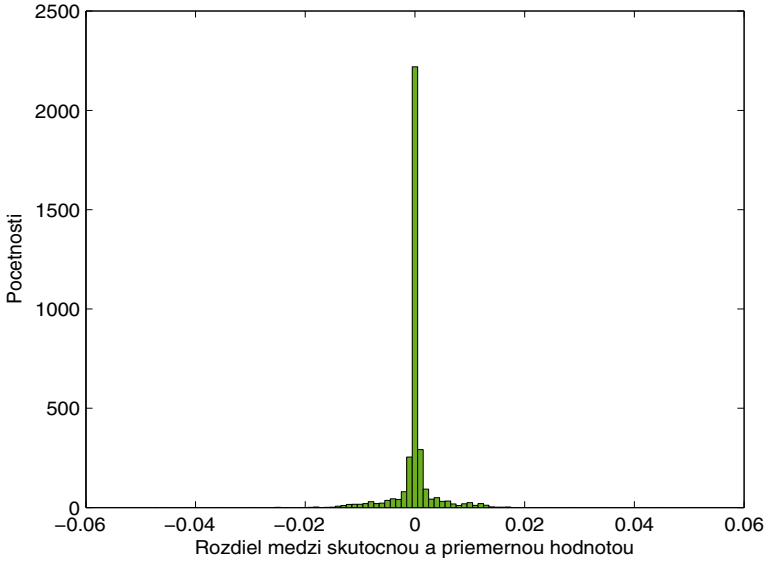



***Obr. 43.*** *Histogram pre saxofón 440 Hz*

Na potvrdenie našich záverov celú analýzu aplikujeme aj na tóny umelo vytvorené pomocou programu Guitar Pro 5. Z parametrických súborov určíme priemerné hodnoty a zobrazíme do grafov okamžitých a priemerných hodnôt a histogramov (Obr. 44, Obr. 45). Tým sa potvrdia naše predpoklady, že rozdiely medzi priemernými a okamžitými hodnotami nie sú veľké a chybné rámce môžeme nahradiť priemernými.




***Obr. 44.*** *Graf okamžitých a priemerných hodnôt amplitúd pre saxofón 440 Hz*



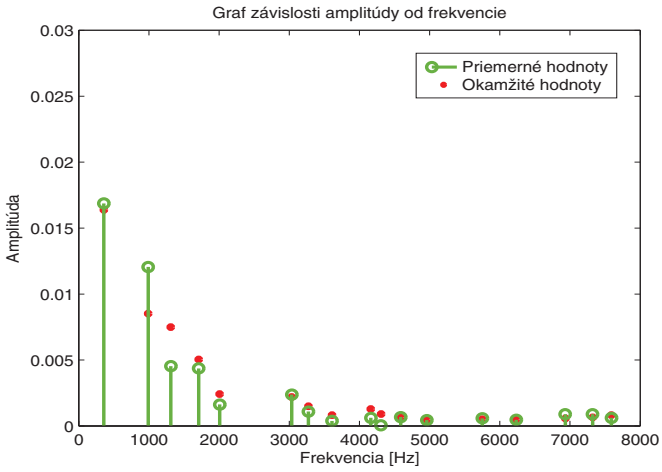

***Obr. 45.*** *Histogram pre saxofón 440 Hz*

Priebeh ľudského hlasu je premenlivejší ako priebeh hudobných tónov, preto aj analýza je zložitejšia. Na analýzu ľudského hlasu máme k dispozícii naspievané samohlásky a, e, i, o v stupnici A dur (Tab. 20).

| Označenie tónu | Frekvencia[Hz] |
|---|---|
| A | 220,000 |
| H | 246,942 |
| cis’ | 277,183 |
| d’ | 293,665 |
| e’ | 329,628 |
| fis’ | 369,994 |
| gis’ | 415,305 |
| a’ | 440,000 |

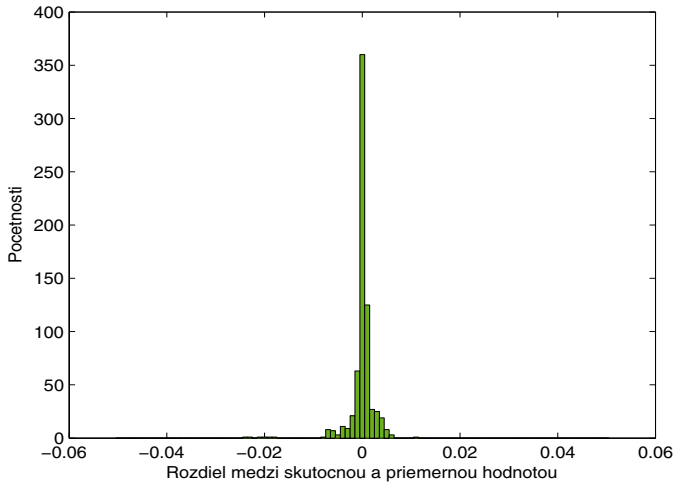
***Tab. 20.*** *Základné frekvencie analyzovaných nahrávok*

Rovnako ako pri hudobných tónoch, aj tu sme si zobrazíme závislosti priebehu amplitúdy od frekvencie. Taktiež vypočítame priemerné hodnoty a zobrazíme histogramy. Znova sa potvrdí naša doterajšia teória, a to že poškodený alebo chybný rámec môže byť nahradený priemerným (Obr. 46).

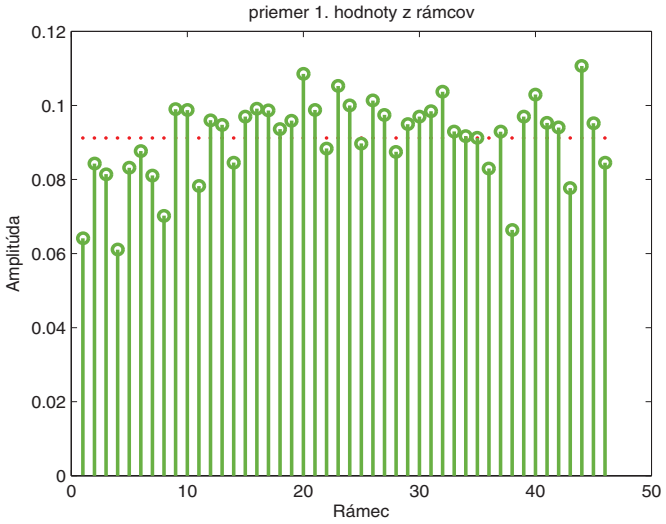



***Obr. 46.*** *Graf okamžitých a priemerných hodnôt amplitúd pre zaspievané „a“ 220 Hz*

Pre dominantné sínusoidy nás zaujíma aj to, ako sa menia jednotlivé amplitúdy v závislosti od frekvencie. Priebehy týchto sínusoíd vykreslíme (Obr. 48). Ako si môžeme všimnúť, hodnoty amplitúd sa pohybujú okolo priemernej hodnoty.

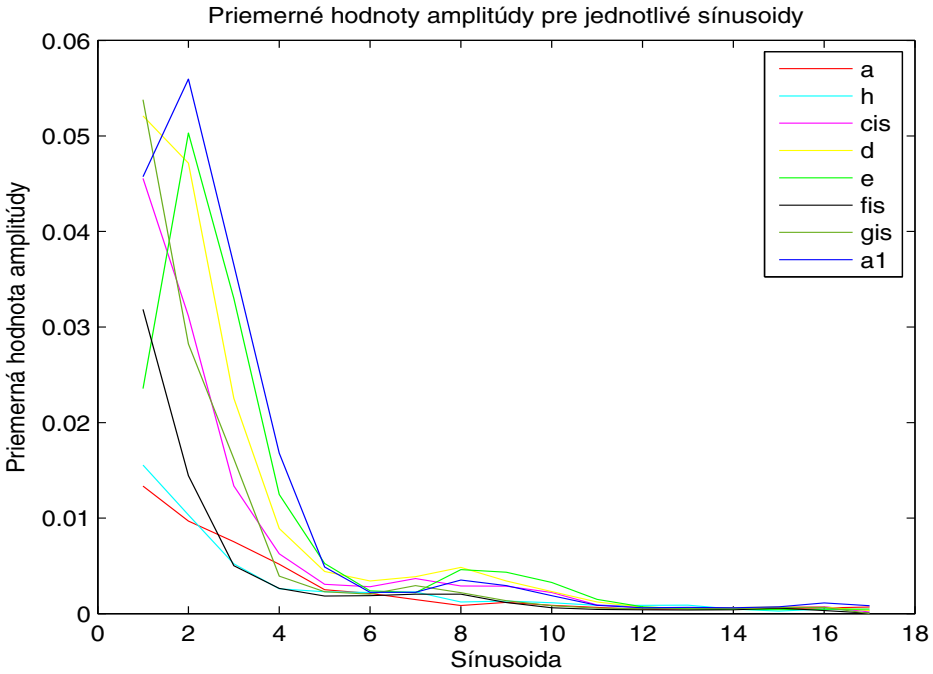



***Obr. 47.*** *Histogram pre zaspievané „a“ 220 Hz*

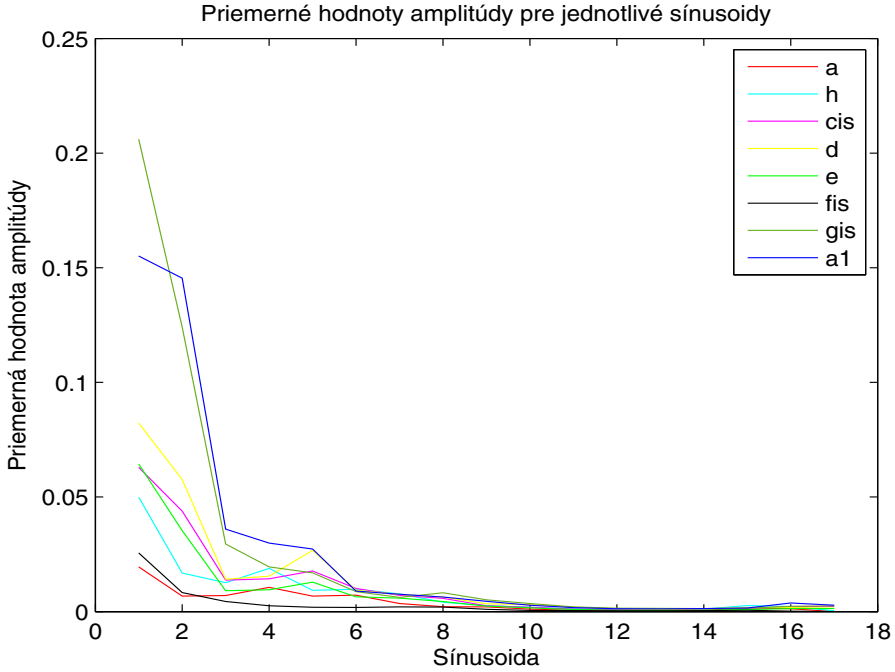



***Obr. 48.*** *Zobrazenie dominantnej sínusoidy pre zaspievané „i“ pri 220 Hz*

V ďalšom kroku zobrazíme, ako sa menia jednotlivé priemery sínusoíd pre analyzované tóny. Všetky priebehy necháme vykresliť do jedného spoločného grafu (Obr. 49, Obr. 50). Z daných priebehov je vidieť určitú závislosť medzi jednotlivými tónmi. Zmeniť jeden tón na druhý nie je triviálne, pretože ako je vidieť, závislosť nie je lineárna a nestačí iba prenásobiť jednotlivé frekvencie konštantou. Rovnako je závislosť iná pre rôzne samohlásky a rovnako môžeme predpokladať, že bude rôzna aj pre ostatné znelé spoluhlásky.



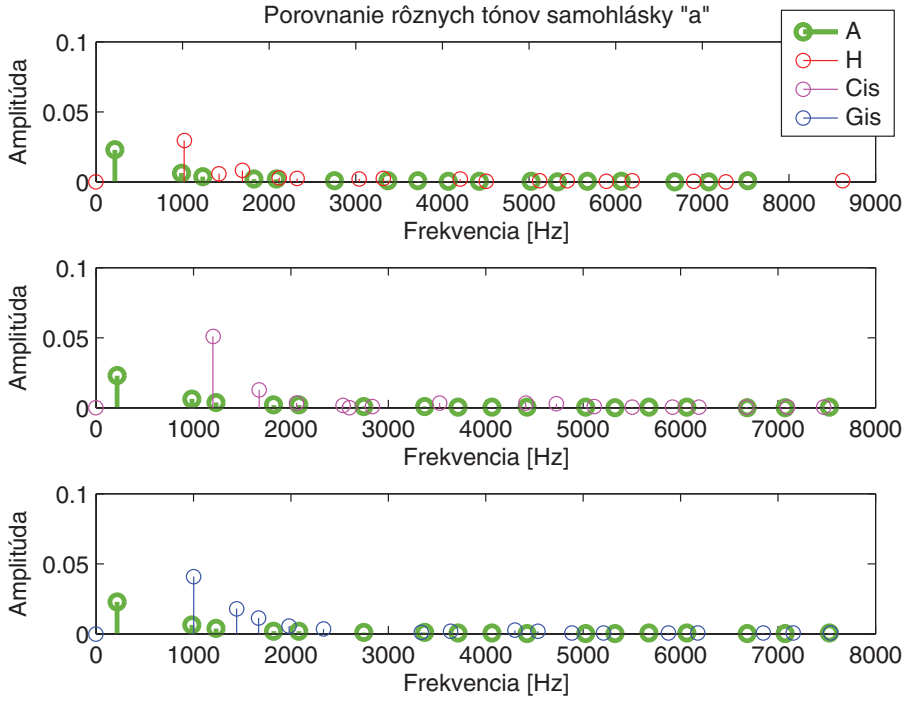

***Obr. 49.*** *Priemerné hodnoty pre zaspievané „a“*




***Obr. 50.*** *Priemerné hodnoty pre zaspievané „e“*

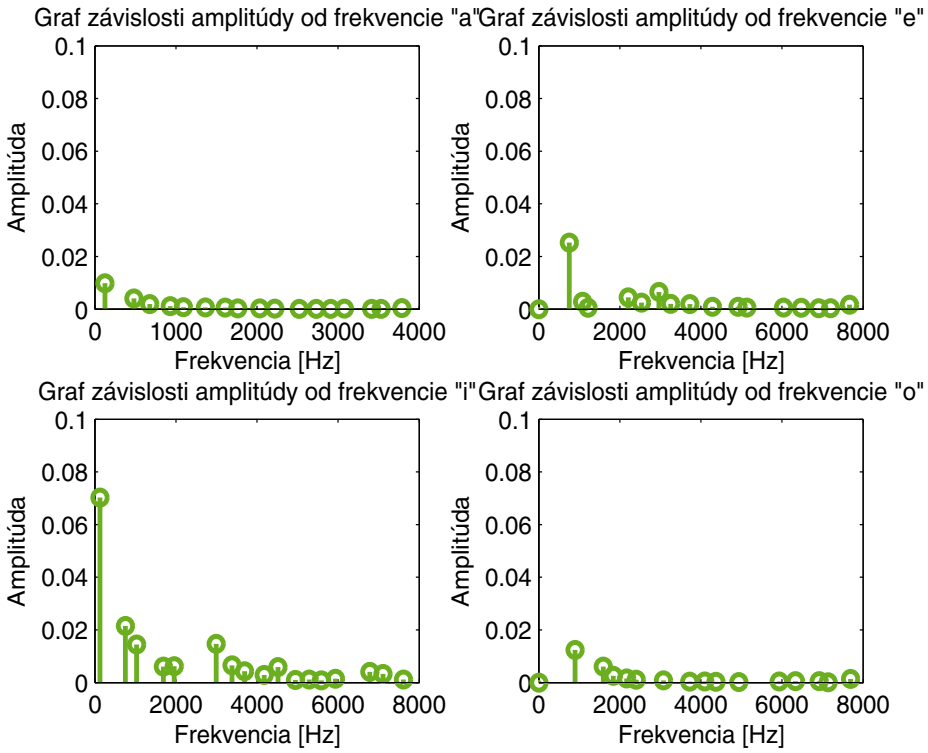
Na analýzu ľudského hlasu použijeme aj iný prístup, a to HNM analýzu popísanú v kapitole 3.4. Jej výhoda oproti použitiu SNM je v tom, že sa stále deteguje hlasivková frekvencia a aj vyššie formanty. Pri SNM sa zdeteguje fundamentálna frekvencia a ostatné sínusoidy sa hľadajú v jej násobkoch. Takže ak sa fundamentálna frekvencia zdeteguje chybne alebo nie veľmi presne, táto chyba sa ďalej iba zväčšuje. Použitím HNM sa tejto akumulácii chyby vyhneme. Z tohto dôvodu však dochádza k veľkým rozdielom počtností sínusoíd v rámcoch, preto je počítanie priemerných hodnôt pôvodnou metódou komplikované a nevedie k relevantným výsledkom.

Priebehy sínusoíd teda analyzujeme pre jednotlivé rámce zvlášť. Nasledujúci obrázok (Obr. 51) znázorňuje priebeh niektorých rámcov tónov stupnice A dur pre samohlásku „a“. Porovnáme aj priebeh tónov pri 220 Hz pre rôzne samohlásky (a, e, i, o, u) (Obr. 52).

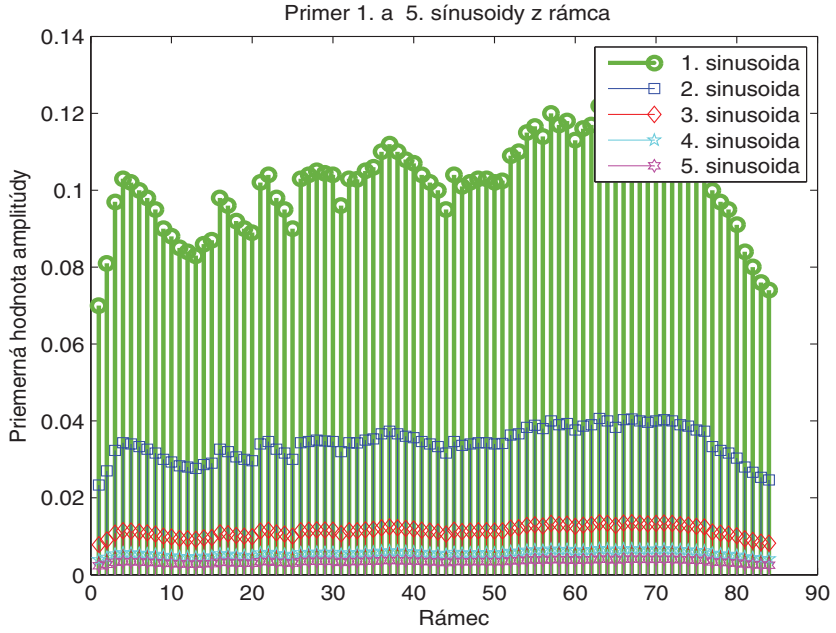



***Obr. 51.*** *Porovnanie rôznych tónov samohlásky „a“*

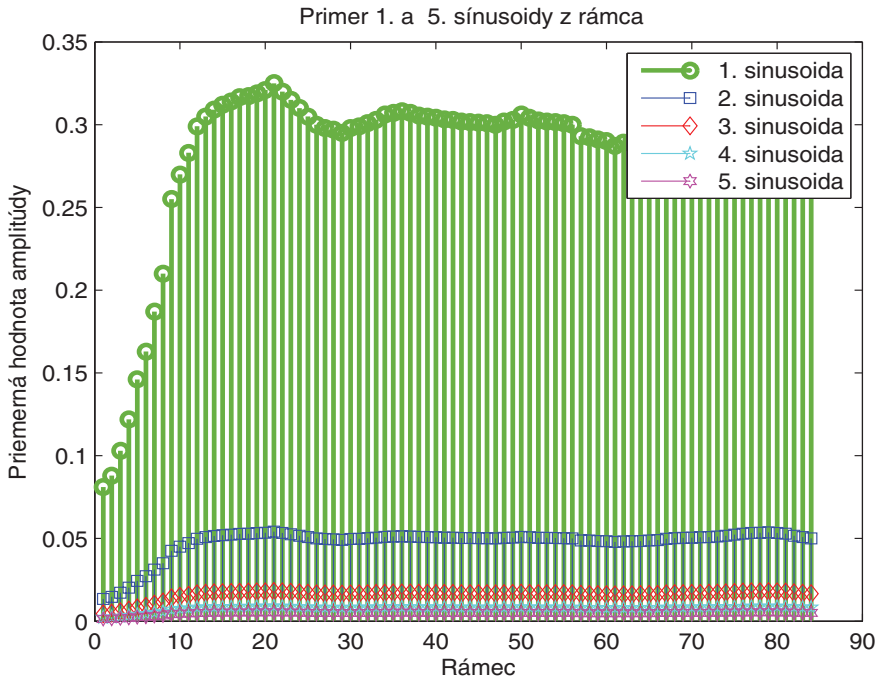
Z priebehov na obrázkoch vidíme, že sa znova treba zamerať na dominantné sínusoidy. Každá samohláska má dominantnú inú sínusoidu. Pre „a“ je to väčšinou prvá a siedma, pre „e“ je to prvá a tretia, pre „i“ je to prvá a pre „o“ je to prvá a siedma. Pre porovnanie analyzujeme priebehy týchto sínusoíd. Porovnáme priebehy pre 220 Hz (Obr. 53) a 440 Hz (Obr. 54). Hodnoty pre jednotlivé sínusoidy sú ustálené a mení sa iba amplitúda. Z toho usudzujeme, že na dosiahnutie zmeny o oktávu vyššie stačí prenásobiť frekvencie jednotlivých sínusoíd konštantou.




***Obr. 52.*** *Rámce tónov stupnice a-dur pre samohlásky „a, e, i, o“ 220 Hz*




***Obr. 53.*** *Priebeh prvej až piatej sínusoidy pre „a“ 220 Hz*




***Obr. 54.*** *Priebeh prvej až piatej sínusoidy pre „a“ 440 Hz*

Pri pokusoch zmeniť prozódiu (v našom prípade išlo o zmenu oznamovacej vety na opytovaciu) bola úspešnosť len 20 – 30%. Z toho usudzujeme, že pri zmene prozódie nestačí iba meniť frekvenciu sínusoíd a dĺžku rámca, ale aj iné parametre ako napríklad fázu. Tiež musíme preskúmať presnú závislosť zmien jednotlivých parametrov, pretože nie je lineárna. Je potrebné vykonať dostatočný počet meraní a testov, aby sme vedeli odvodiť presný vzťah na zmenu prozódie.

## 5.8 Využitie sínusoíd ako korpus pre rečový syntetizátor

Využitie sínusoíd našlo svoje uplatnenie aj ako istý druh kompresie zvuku. Praktické využitie je napríklad pri aplikácii difónovej syntézy na mobilnom telefóne. V prvotnej implementácii difónovej syntézy je databáza vo wav formáte. Takáto databáza má veľkosť 6 MB, a preto sa nie vždy nahrá celá do pamäte telefónu. Dnešné telefóny majú síce aj prídavné pamäte, ale pre správne fungovanie aplikácie je dôležitá veľkosť takzvanej heap memory, čo je vlastne operačná pamäť telefónu. Jedným riešením je zmenšenie databázy iba na časť difónov a ostatné chýbajúce difóny sa poskladali z foném. To má však na výslednú kvalitu syntézy nežiadúci vplyv. Možnosťou komprimácie databázy a jej použitím

v parametrickom tvare sa tento problém vyriešil.

Databázu parametrizujeme pomocou HNM modelu [88]. Pre reálne signály a spracovanie v reálnom čase je HNM model stratovým modelom. Pre bloky postprocessingu, ktoré zvyčajne pracujú v reálnom čase a sú menej komplexné, sa dosahuje nižšia kvalita zvuku. Pri viac komplexných blokoch, ktoré nemusia pracovať v reálnom čase, sa dosahuje kvalita zvuku dosť vysoká na to, aby ľudské ucho nedokázalo rozlíšiť, či ide o originálny signál alebo resyntetizovaný z komprimovanej databázy [88].

HNM analýza je oveľa náročnejšia na čas ako samotná syntéza. Takmer 80% celkového času trvá analýza. Preto pri našej aplikácii proces analýzy a vytvorenie databázy nebeží v reálnom čase. Aplikácia pristupuje k už dopredu vytvorenej komprimovanej databáze. Parametrizovanie databázy sa vykoná na štandardnom PC, čo nám umožňuje použiť aj zložitejšie spôsoby výpočtov na presnejšie určenie parametrov sínusoíd, čím sa zlepší aj výsledok syntézy. V reálnom čase beží len spätná syntéza, ktorá nie je tak časovo a výpočtovo náročná. Syntéza sa už vykonáva v mobilnom telefóne.

Princíp zakódovania databázy v parametrickom tvare (štruktúra parametrického súboru bola opísaná v kapitole 5.7) je nasledovný [88]. Prvý bajt označuje začiatok rámca s hlavičkou F. Nasledujúcich 5 bitov obsahuje údaje ako dĺžka rámca, počet sínusoíd v aktuálnom rámci, základnú hlasivkovú frekvenciu, informáciu o znelosti, násobok amplitúdy.

Dĺžka rámca je binárne zakódovaná do 11 bitov. Treba poznamenať, že prirodzené čísla sú kódované v malom endiane. Algoritmus používa rámce s rovnakým počtom vzoriek, takže reálne zakódovaná hodnota je polovica dĺžka rámca. Pri dekódovaní sa táto hodnota jednoducho vynásobí dvoma. To umožňuje používať rámce s maximom vzoriek 4 095, čo je veľmi užitočné najmä pri analýze s vyššou vzorkovacou frekvenciou. Ďalších 9 bitov obsahuje informáciu o počte sínusoíd v aktuálnom rámci. Maximálna možná hodnota je 511. Informácia o základnej hlasivkovej frekvencii je zakódovaná v nasledujúcich 16 bitoch. Vychádza sa z predpokladu, že je z rozsahu 0 až 600 Hz. Kvantizačný krok je potom $600/2^{16}$. Pre neznelé časti stačí zaznamenať informáciu o obálke signálu. Neznelé časti nemajú fundamentálnu frekvenciu, preto je hodnota nastavená na 100 Hz. Ďalší bit hovorí o tom, či je rámec znelý alebo neznelý. Posledné 3 bity sú rezervované pre násobky amplitúdy, čo uľahčuje binárne kódovanie desatinných hodnôt amplitúdy. V každom rámci sú desiatky sínusoíd. Nájde sa maximálna amplitúda. Násobič amplitúdy je číslo $10^n$, kde $n$ závisí od počtu desatinných miest v amplitúde (ak maximálna amplitúda je 0,08; $n$ = 1 a násobič amplitúdy je 10, potom všetky amplitúdy v rámci sú delené touto hodnotou a potom zakódované).

Táto štruktúra 6 bitov sa objaví vždy iba raz na začiatku rámca. Potom sa už kóduje amplitúda a fáza v 2 bajtoch. Prvé dva bity slúžia na zakódovanie desatinnej hodnoty amplitúdy. Ďalších 8 bitov kóduje amplitúdu. Posledných 6 bitov obsahuje zakódovanú fázu. Dôvodom na kódovanie informácie o fáze je lepšia kvalita výslednej syntézy.

Veľkosť komprimovanej databázy ovplyvňuje výslednú kvalitu zvuku. Pri kódovaní menším počtov bitov nedosahovala výsledná syntéza ani priemernú kvalitu. Popísaným

spôsobom sa dosiahne kompresia 1:4. Pri tomto pomere je aj kvalita syntézy na dobrej úrovni. Efektivita kompresie závisí aj od rečového signálu, pretože počet sínusoíd závisí od komplexnosti a typu rečového signálu.

Pri syntéze z takto komprimovanej databázy je potrebné najskôr získať späť všetky údaje. Veľkou výhodou pri parametrizovanej databáze je možnosť pred samotnou syntézou meniť jednoduchým elegantným spôsobom parametre, a tak meniť prozódiu výslednej syntetizovanej reči. Po získaní vzťahov a závislostí medzi originálnym signálom a želaným signálom bude možné vytvoriť pre syntézu rôzne hlasy. Získanie predpisov na zmenu jednotlivých parametrov však nie je triviálnou záležitosťou a zatiaľ sú k dispozícii iba výsledky z predošlej kapitoly (5.7).

# 6 Zhrnutie

Táto publikácia je venovaná syntéze reči, konkrétne metódam učenia rečového syntetizátora. Napriek pomerne všeobecnému názvu sa monografia veľmi konkrétne zameriava na návrh spôsobu zdokonalenia syntetizovanej reči v inteligentnom učiacom sa systéme. Výsledkom práce je metodika učenia a architektúra inteligentného rečového syntetizátora pre slovenský jazyk, ktorý sa dá prispôsobiť a aplikovať aj pre syntézu iných jazykov. Návrh je nezávislý aj od druhu syntézy (difónová, korpusová, HMM,...).

Publikácia je štruktúrovaná do šiestich kapitol, popisuje návrh inteligentného rečového syntetizátora a využitie sínusoidálnych modelov pri modifikácii prozódie rečového signálu. V prvých štyroch kapitolách je podaný prehľad o problematike syntézy reči, kompresie signálov a sínusoidálneho modelovania. V piatej, najrozsiahlejšej kapitole, je popísaný návrh inteligentného systému pre syntézu reči a pre zmenu prozodických vlastností reči. Táto kapitola predkladá celkový návrh inteligentného systému, architektúru modulárneho rečového syntetizátora a aplikovanie učenia do jednotlivých modulov. Implementácia učenia v moduloch je na rôznej úrovni, najviac rozvinutá funkcionalita je v module fonetickej transkripcie. Moduly si vymieňajú informácie pomocou XML súborov. Architektúra je flexibilná a umožňuje zapojiť do procesu syntézy všetky alebo iba vybrané moduly. Proces učenia môže prebiehať na rôznych stupňoch. Najjednoduchším spôsobom je manuálna oprava chýb. Vyššia forma je, ak syntetizátor ponúkne možnosti opravy chyby pri označení nesprávneho slova alebo vety, v inteligentnejšom prípade sú možnosti zoradené podľa určitého pravidla (napríklad pravdepodobnosť výskytu danej možnosti v jazyku). Najdokonalejšou formou učenia je integrácia syntetizátora s rečovým rozpoznávačom. Používateľ iba číta správne formy slov a viet, všetky úpravy prebiehajú automaticky. Celý proces učenia si však vyžaduje dlhší čas, počas ktorého by niekoľko používateľov systém testovalo, a tým zdokonaľovalo. Zmene prozódie sa venuje samostatný modul. Jednou z možností, ako prozódiu modifikovať, je využitie sínusoidálnych modelov. Vďaka parametrickému modelu, ktorý flexibilne reprezentuje signál, sa môže jednoducho meniť frekvenčný priebeh alebo dĺžka signálu v časovej oblasti. Toto umožňuje pomerne jednoducho signál modifikovať na základe požiadaviek prozódie. Jednou z možností využitia je modifikácia hlasu na hlas inej osoby, čo však už nie je triviálna aplikácia a je potrebné robiť analýzu signálu. Pri tejto analýze vzájomných vzťahov parametrov sínusoíd a vlastností rečníka sa využíval vylepšený HNM model, ktorý poskytoval lepšie možnosti analýzy a následnej modifikácie signálu. Publikácia obsahuje aj popis, ako sa sínusoidálne modely dajú veľmi dobre využiť na kódovanie a kompresiu reči.

# Zoznam použitej literatúry

[1] SPANIAS, A.S., Speech Coding: A Tutorial Review, Proc. IEEE, vol. 82, no. 10, pp. 1541-1582, Oct. 1994

[2] TURI-NAGY M., Využitie sinusoidálneho modelu pre spracovanie audiosignálov, Dizertačná práca, Katedra telekomunikácií, FEI STU, 2006, Bratislava

[3] Black, A. W., Lenzo, K.A., Building Synthetic Voices, 2007, [Online], http://festvox.org/bsv/

[4] Ďuriš, J., Wolfgang von Kempelen a jeho Mechanizmus ľudskej reči, 1996, [Online], http://www.radioart.sk/avr/visuopage.php?id=148

[5] CAMPBELL, W.N.: „CHATR: A high-definition speech re-sequencing system", in Proc. 3rd ASA/ASJ Joint Meeting, pp.1223-1228, 1996

[6] BEUTNAGEL, M. et al.: „The AT&T next-gen TTS system", in Proc. 137th Meeting Acoustical Society America, [Online], 1999, http://www.research.att.com/projects/tts

[7] MURRAY, I. R., ARNOTT, J. L.: „Implementation and testing of a system for producing emotion-by-rule in a synthetic speech", Speech Communication 16, pp. 359-368, 1995

[8] RANK, E., PIRKER, H.: "Generating Emotional Speech with a Concatenative Synthesiser", Proc. ICLSP'98, vol. III., pp. 371-674, 1998

[9] MONTERO, J. M. et al.: "Development of an Emotional Speech Synthetiser in Spanish", in Eurospeech'99, 6th European Conference on Speech Communication and Technology. September 5-9, 1999, Budapest, Hungary. pp. 2099-2102

[10] STYLIANOU, Y., LAROCHE, J., MOULINES, E.: „High-Quality Speech Modification based on Harmonic + Noise Model", Proc. EUROSPEEECH, pp. 451-454, 1995

[11] ABE, M., NAKAMURA, S., SHIKANO, K., KUWABARA, H.: "Voice Conversion Through Vector Quantization", J. Acoust. Soc. Jpn., vol. E-11, pp. 71-77, March

1990

[12] MIZUNO, H., ABE, M.: "Voice Conversion Based on Piecewise linear Conversion Rule of Formant Frequency and Spectrum Tilt", Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 1994

[13] STYLIANOU, Y., CAPPÉ, O., MOULINES, E.: "Continuous Probabilistic Transform for Voice Conversion", IEEE Trans. On Speech and Audio Processing, vol. 6, no. 2, March 1998

[14] WILSON, S, Wave PCM soundfile format, 2003, [Online], http://ccrma.stanford.edu/courses/422/projects/WaveFormat/

[15] Coding Technologies SBR explained, 2005, [Online], http://www.codingtechnologies.com/products/sbr.htm

[16] Coding Technologies aacPlus, 2005, [Online], http://www.codingtechnologies.com/products/aacPlus.htm

[17] Coding Technologies mpSPRO, 2005, [Online], http://www.codingtechnologies.com/products/MP3pro.htm

[18] MIT Media Laboratory MPEG-4 - mp4 Structured Audio homepage, 2000, [Online], http://sound.media.mit.edu/mpeg4/

[19] BOSSE, L., Modified Discrete Cosine Transform (MDCT), 1998, [Online], http://ccrma-www.stanford.edu/ bosse/proj/node27.html

[20] NIKOLAJEVIC, V., FETTWEIS, G.,: New Recursive Algorithms for the Unified Forward and Inverse MDCT/MDST, 2003, [Online], http://www.ifn.et.tu-dresden.de/MNS/veroeffentlichungen/2003/Nikolajevic_V_Kluwer_03.pdf

[21] HERRE, J., MPEG – 4 General Audio Coding, Fraunhofer Institute for Integrated Circuits (IIS), 2006, Germany, [Online], http://sound.media.mit.edu/mpeg4/audio/general/aes106_1-GeneralAudio.pdf

[22] General Audio Coding (AAC plus), 2006, [Online] http://www.chiariglione.org/MPEG/tutorials/papers/icj-mpeg4-si/09-natural_audio_paper/gacoding.html

[23] HERRE, J., Tempotal Noise Shaping, Quantization and Coding Methods In Perceptual Audio Coding: A Tutorial Introduction, Fraunhofer Institute for Integrated

Circuits (IIS), Erlangen, Germany, [Online], http://www.MP3-tech.org/programmer/docs/Temporal%20Noise%20Shaping%20Quantization%20and%20Coding%20Methods%20in%20Perceptual%20Audio%20Coding%20-%20A%20Tutorial%20Introduction.pdf

[24] http://en.wikipedia.org/wiki/MP3, 2008, [Online]

[25] http://wiki.hydrogenaudio.org/index.php?title=Ogg_Vorbis, 2008, [Online]

[26] BLACK, A.W., Perfect synthesis for all of the people all of the time. Proceedings of 2002 IEEE Workshop on Speech Synthesis, pp. 167-170, 2002, California, USA

[27] MOBIUS, B., Corpus-based speech synthesis: methods and challenges. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung, 2000, Univ. Stuttgart, AIMS 6 (4), pp. 87-116

[28] TANAKA, K., MIZUNO, H., ABE, M., NAKAJIMA, S. A, Japanese text-to-speech system based on multi-form units with consideration of frequency distribution in Japanese. Proceedings of the European Conference on Speech Communication and Technology, vol. 2, 1999, Budapest, Hungary, pp. 839–842

[29] MOBIUS, B., Rare events and closed domains: Two delicate concepts in speech synthesis. International Journal of Speech Technology 6 (1), 57-71, 2003

[30] ČEPKO, J., Písomná práca k dizertačnej skúške, Katedra telekomunikácií, FEI STU, 2005, Bratislava

[31] TATHAM, M., LEWIS, E. Improving text-to-speech synthesis. Proceedings of the Institute of Acoustics, 1996, vol. 18 (9), pp. 35-42

[32] BLACK, A. W., TAYLOR, CHATR, P., A generic speech synthesis system. In Proc. of the International Conference on Computational Linguistics, 1994, Kyoto, Japan

[33] BLACK, A. W., CAMPBELL N., Optimising selection of units from speech databases for concatenative synthesis. Eurospeech95, vol. I, pp. 581-584, 1995, Madrid, Spain

[34] HUNT, BLACK, W.,Unit selection in a concatenative speech synthesis system using a large speech database. In Proc. of ICASSP, 1996, Atlanta, Georgia, pp. 373-376

[35] CERŇAK M., Využitie objektívnych meraní kvality pri korpusovej syntéze reči. Dizertačná práca, FEI STU, 2005, Bratislava

[36] HON, H., ACERO A., HUANG, X., LIU, J., PLUME, M. Automatic Generation of Synthesis Units from Trainable Text-To-Speech Systems. IEEE Proc. ICASSP, 1998, Seattle, pp. 293-206

[37] DONOVAN, R. E., WOODLAND, P. A, Hidden Markov-model based trainable speech synthesizer. Computer Speech and Language, 13(3):223-241, 1999

[38] YI, J., GLASS, J., Information-theoretic criteria for unit selection synthesis. In Proc. of ICSLP 2002, 2002, Denver, pp. 2617–2620

[39] LEE, K., COX, R.V., A segmental speech coder based on a concatenative TTS, Speech Communication 38 (2002) 89–100, 2001

[40] VRÁBEL, A.: Postspracovanie syntetizovanej slovenskej reči, Diplomová práca, Katedra telekomunikácií, FEI STU, 2005, Bratislava

[41] Text-to-speech native coding in a communication system, 2003,[Online, ]http://www.freepatentsonline.com/6681208.html

[42] VARGA, T., Optimalizácia syntézy reči podľa originálu, Diplomová práca, Katedra telekomunikácií, FEI STU, 2008, Bratislava

[43] DUDLEY, H.: Remaking Speech, J. Acoust. Soc. Amer., vol. 11, ASSP-24, p.169, 1939

[44] MUNSON, W., MONTGOMERRY, H.: A speech analyzer and synthesizer, J. Acoust. Soc. Amer.,vol. 222, 1950, p.678

[45] DUDLEY, H.: Phonetic pattern recognition vocoder for narrowband speech transmission, J. Acoust. Soc. Amer., vol. 30, pp. 733-739, 1958

[46] DELLER, J. R., PROAKIS, J. G., HANSEN, J. H. L.:Discrete-time processing of speech signals, 1993, Prentice Hall

[47] FITZ, K., WALKER, W., HAKEN, L.: Extending the McAulay-Quatieri Analysis for Synthesis with a Limited Number of Oscillators, ICMC Proceedings, 1992

[48] OLIVER, B. M., PIERCE, J., SHANONN, C.E.: The philosophy of PCM, Proc. IRE, vol. 36, November 1948, pp. 1324-1331,

[49] CCITT Reccomendation G.721, 32 kbit/sAdaptive Differential Pulse Code Modulation (ADPCM), in Blue Book, vol. III, Fascicle III.3, October 1988

[50] FANT, G.: Acoustic Theory of Speech Production, Gravenhage, 1960, The Netherlands: Mouton and Co.

[51] MAKHOUL, J.: Linear Prediction: A Tutorial Review, Proc. IEEE, vol. 63, no. 4, pp. 245-248, April 1972

[52] OPPENHEIM, A., SCHAFER, R.: Homomorphic analysis of speech, IEEE Trans. Audio, vol. AU-1, no. 6, pp.221-226, June 1968

[53] FLANAGAN, J., GOLDEN, R.: Phase vocoder, Bell Syst. Tech. J., vol. 45, p.1493, November 1966

[54] YU, E. W. M., CHAN, Ch-F.: Robust Multiband Excitation Coding of Speech Based on Variable Analysis Frame Sizes, VIII European Signal Processing Conference EUSPICO-96, September 10-13 1996, Trieste, Italy

[55] ATAL, B., REMDE, J.: A new model for LPC excitation for producing natural sounding speech at low bit rates, April 1982, Proc. ICASSP-82, pp. 614-617

[56] CARDINAL, J.: Quantization with Multiple Constraints, University Libre de Bruxelles, Faculty of Sciences, Departement of Informatics, PhD. Thesis, 2001

[57] SCHROEDER, M.R., ATAL, B.: Code-excited linear prediction (CELP): High quality speech at very low bit rates, in Proc. ICASSP-85, p.937, April 1985, Tampa, Florida,

[58] ELLIS, D., ROSENTHAL, D.: „Mid-level representations for Computanional Auditory Scene Analysis“, International Joint Conference on Artificial Intelligence, August 1995, Quebec,

[59] ALI, M.:Adaptive Signal Representation with Applications in Audio Coding, Ph.D. thesis, University of Minnesota

[60] LEVINE S.: Audio Representation for Data Compression and Compressed Domain Processing, Ph.D. thesis, 1998, Stanford University

[61] RODET, X.: Musical Sound Signal Analysis/synthesis: Sinusoidal +Residual and Elementary Waveform Models, IEEE Time-Frequency and Time-Scale Workshop, 1997, Coventry, Grande Bretagne

[62] VIRTANEM, V.: Audio signal modeling with sinusoids plus noise, MSc Thesis, Tampere University of Technology, August 2000

[63] FITZ, K., WALKER, W., HAKEN, L.: Extending the McAulay-Quatieri Analysis for Synthesis with a Limited Number of Oscillators. ICMC Proceedings, 1992

[64] IVANECKÝ J., Automatic transcription of slovak in computer speech recognition. In Alexandra Jarošová editor, Slovenčina a čeština v počítačovom spracovaní, pages 109 – 116, 2001, Bratislava, Slovakia, VEDA

[65] CERŇAK M., Využitie objektívnych meraní kvality pri korpusovej syntéze reči, Dizertačná práca, Katedra telekomunikácií, FEI STU, 2004, Bratislava

[66] STYLIANOU, Y., Applying the Harmonic plus Noise Model in Concatenative Speech Synthesis, IEEE Trans. Acoustics, Speech and Signal Processing, vol. 9, no. 1, pp. 21-29, January 2001

[67] STYLIANOU, Y., LAROCHE, J., MOULINES, E.: „High-Quality Speech Modification based on Harmonic + Noise Model, Proc. EUROSPEEECH, pp. 451-454, 1995

[68] OANCEA E., BADULESCU A., Stressed Syllable Determination for Romanian Words within Speech Synthesis Applications

[69] HUANG F., WU J., Study and Implementation of Intelligent E-learning System for Modern Spoken Chinese, Proceedings of the Eight International Conference on Machine Learning and Cybernetics, Baoding, July 2009

[70] DVONČ, L., a iní., Morfológia slovenského jazyka, Bratislava : Vydavateľstvo Slovenskej akadémie vied, 1966

[71] TURLÍKOVÁ, L., Tvaroslovník - databáza tvarov slov slovenského jazyka, Diplomová práca, Košice

[72] BIELIKOVÁ, M., NÁVRAT, P a kol., Štúdie vybraných tém softvérového inžinierstva 3., Bratislava : Vydavateľstvo STU, 2007, ISBN 978-80-227-2701-3

[73] CENTRUM VÝPOČTOVEJ GRAFIKY V BRATISLAVE, 2007, [Online, ] http://www.utf-8.sk, Je čas prejsť na UTF-8

[74] BRILL, E., A Simple Rule-Based Part of Speech Tagger, Proceedings of the third conference on Applied natural language processing University of Pennsylvania, 1992, [Online], http://acl.ldc.upenn.edu/A/A92/A92-1021.pdf

[75] ŠIMKOVÁ, M., Morfologická anotácia textov Slovenského národného korpusu, 2006, [Online], http://korpus.juls.savba.sk/files/tagset-www.pdf

[76] SLOVNESKÝ NÁRODNÝ KORPUS, r-mak-3.0. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2009, [Online], http://korpus.juls.savba.sk

[77] PSUTKA, J. a kol., Mluvíme s počítačem česky, Praha Academia, 2006, ISBN 80-200-1309-1.

[78] KRÁĽ, A., Pravidlá slovenskej výslovnosti: Systematika a ortoepický slovník, Martin, Matica slovenská, 2005, ISBN 80-7090-790-8.

[79] SYNAK, D, Analýza melodických kontúr slovenčiny, Bakalárska práca, Katedra telekomunikácií, FEI STU, 2008, Bratislava

[80] SIČÁKOVÁ, Ľ., Fonetika a fonológia pre elementaristov, Prešov: Náuka, 2002

[81] KONDELOVÁ A., Analýza prozodických vlastností slovenskej reči, Diplomová práca, Katedra telekomunikácií, FEI STU, 2010, Bratislava, FEI-5410-23262

[82] TÓTH J., Fonetická transkripcia skratiek pri syntéze reči, Diplomová práca, Katedra telekomunikácií, FEI STU, 2010, Bratislava, FEI-5410-29003

[83] VALENT M., Automatická detekcia slovných druhov v slovenskej vete, Diplomová práca, Katedra telekomunikácií, FEI STU, 2010, Bratislava, FEI-5410-22339

[84] GONŠOR J., Metodika opráv chybnej syntézy reči, Diplomová práca, Katedra telekomunikácií, FEI STU, 2010, Bratislava, FEI – 5410 - 28557

[85] IVANECKÝ J.: Automatická transkripcia slovenčiny v počítačovom rozpoznávaní reči, In: Slovenčina a čeština v počítačovom spracovaní, Zborník referátov zo seminára, Bratislava 26.-27. októbra 2001 (ISBN 80-224-0692-9)

[86] IVANECKÝ J., Nabelková M.: Fonetická transkripcia SAMPA a slovenčina, Jazykovedný časopis, 53, 2002, s. 81 - 95.

[87] TALAFOVÁ, R.: Syntéza reči v mobilnom telefóne, Diplomová práca, Katedra telekomunikácií, FEI STU, Bratislava 2007

[88] ROZINAJ, G., TURI NAGY, M., RYBÁROVÁ, R.: Sinusoidal Parametrization for Speech Synthesis in Mobile Phones In: WORLDCOMP'09 - The 2009 World Congress in Computer Science, Computer Engineering, and Applied Computing, July 13-16, 2009, Las Vegas,USA

[89] ROZINAJ, G., RYBÁROVÁ, R., MINÁRIK I.: Multimédia, Kurz C5, učebný text, IMProVET – Innovative Methodology for Promising VET Areas LIFELONG LEARNING PROGRAMME LEONARDO DA VINCI MULTILATERAL PROJECTS - TRANSFER OF INNOVATION 2011, GRANT AGREEMENT n° CZ/11/LLP-LdV/TOI/134011 (2011-2013)

[90] CAMPEANU, D., CAMPEANU, A.: An Objective Method to Assess the Perceptual Quality of Audio Compressed Files [available] http://ace.ucv.ro/sintes12/SINTES12_2005/SOFTWARE%20ENGINEERING/09.pdf

[91] MALÁ, T., KAČUR, J.: Moderné postupy kompresie audio signálov - algoritmus Ogg Vorbis. In: EE časopis pre elektrotechniku a energetiku. Roč. 15, mimoriadne č. (2009), s. 129–132, ISSN 1335-2547

[92] ITU-T recommendation P.862 [online] http://www.itu.int/rec/T-REC-P.862/

[93] ITU-R Recommendation BS.1387 [online] http://www.itu.int/rec/R-REC-BS.1387

[94] ALLEN B.: Cognitive differences in end user searching of a CD-ROM index. In: Proceedings of the 15th Annual International ACM SIGIR Conference on Research and development in information retrieval (SIGIR'92), ACM Press New York, NY, USA, 1992, s. 298–309

[95] COUVREUR, L., BOITE, J.M.: Speaker Tracking in Broadcast Audio Material in the Framework of the THISL Project In: Proceedings of 1999 Workshop on Accessing Information in Spoken Audio (ESCA-ETRW), Cambridge, UK, april 1999, s.84-89

[96] EEFTING, W., NOOTEBOOM, S. G.: Accentuation, information value and word duration: effects on speech production, naturalness and sentence processing In: V. J. Van Heuven & L. C. W. Pols (Eds.),Analysis and synthesis of speech: Strategic research towards high-quality text-to-speech generation, Berlin, Mouton de Gruyter. Speech Research, 1993, s. 225-240

[97] LARREUR, D., at al.: A real-time French text-to-speech system generating high-quality synthetic speech In: International Conference on Acoustics, Speech, and Signal Processing, 1989, s.309-312

[98] JUHÁR, J. a kol., Rečové technológie v telekomunikačných a informačných systémoch, EQUILIBRIA, Košice, 2011, 517 s. ISBN 978-80-89284-75-7.

[99] SULÍR, M., JUHÁR, J., ONDÁŠ, S.: Speech Synthesis Evaluation for Robotic Interface, In: Complex Control Systems. Vol. 11, no. 1 (2012),s. 64-69, ISSN 1310-8255

[100] SULÍR, M., JUHÁR, J.: Design of an Optimal Male and Female Slovak Speech Database for HMM-Based Speech Synthesis, In: Proc.Redžúr 2013 - 7th International Workshop on Multimedia and Signal Processing, May 1, 2013, Smolenice, Slovakia, ISBN 978-80-227-3921-4

[101] LEVICKÝ, D., Multimédiá a ochrana ich obsahu, Elfa, Košice, 2012, ISBN 978-80-8086-199-5

[102] LEVICKÝ, D., Kryptografia v informačnej bezpečnosti, Elfa, Košice, 2005, ISBN 80-8086-022-X

# Register


## A

AAC, 13, 73, 79, 80, 81, 82, 166
akronym, 121, 123, 125, 126, 128, 132
analýza
  akustická, 45
  audiosignálov, 49
  kontextová, 92
  lexikálna, 125
  lingvistická, 31
  morfologická, 109, 110, 115, 116, 122, 126
  morfologicko-syntaktická, 91
  sínusoidálna, 50
  syntaktická, 121
  tónovo-synchrónna, 53

## B

Barkove pásma, 48, 49, 54

## C

časovanie, 93, 107
CELP, 13, 66, 71
CHATR, 14, 34

## D

difón, 20, 28, 29, 156
DTW, 13, 42
dvojzmyselnosť
  sémantická, 121
  syntaktická, 121, 122

## F

filter
  formantový, 31
fonéma, 20, 26, 28, 29, 30, 42, 43, 45, 47, 59, 93, 96, 98, 99, 100, 101, 104, 141, 142
fonetická transkripcia, 7, 30, 32, 89, 91, 92, 93, 96, 99, 103, 120, 121, 131, 159
formant, 20, 21, 25, 42, 65, 88, 154
frázovanie, 95, 117
frekvencia
  fundamentálna, 20, 39, 40, 41, 53, 55, 56, 57, 58, 60, 61, 72, 133, 138, 143, 144, 146, 148, 154
  základná hlasivková, 13, 20, 29, 30, 34, 42, 43, 47, 53, 58, 94, 95, 138, 140, 141, 143, 144, 146, 157
  znelá, 55, 56, 58, 59, 61

## H

histogram, 149, 150, 151
HMM, 14, 28, 30, 34, 35, 53, 92, 95, 159

## I

inteligentné učenie, 106, 120, 131, 145
intenzita, 34, 39, 43, 93, 94
interpolácia
  kvadratická, 51, 141
intonácia, 94, 140, 141

## K

kodéry, 64, 66, 81
  hybridné, 64, 66
  rečovo špecifické (vokodéry), 64, 68
  tvaru vlny, 64, 66, 67
  vlnové, 66
kódovanie
  transformačné, 68, 73, 76
konvolúcia, 56, 69
korpusová metóda, 96
kvantizácia, 63, 66, 80

## L

lineárna interpolácia, 60
LTS, 14, 96, 103, 104

## M

magnitúda, 50, 51, 52, 68, 69
MBE, 14, 40, 72
MDCT, 14, 74, 77, 79, 81
model
  HNM, 14, 42, 43, 55, 56, 58, 59, 66, 72, 154, 157, 159
  parametrický, 40, 47
  sínusoidálny, 7, 39, 40, 41, 47, 48, 49, 50, 64, 146, 159
  SN, 15, 47, 48, 49, 53
  SNM, 146
modulácia
  pulzne kódová, 15, 63
MP3, 73, 74, 77, 78, 79, 80, 81, 82, 85, 86
MPC, 14, 71
MPEG-1, 73, 76, 77
MPEG-4, 73, 79, 81, 82

## N

normalizácia textu, 92, 121

## O

Ogg, 73, 74, 77, 78, 81, 82, 85, 86

## P

pásmová filtrácia, 67
prozódia, 7, 20, 28, 29, 30, 35, 38, 39, 40, 41, 43, 58, 59, 87, 88, 89, 91, 92, 93, 94, 95, 132, 133, 140, 141, 143, 144, 145, 156, 158, 159

## R

RELP, 15, 70
rezíduum, 48, 49, 54, 67, 70, 71

## S

SAMPA, 14, 30, 31, 32, 98, 99, 101, 120, 123, 131
SNK, 15, 108, 109, 110, 113, 129, 130, 133
spektrálna obálka, 40, 41, 42, 49, 51, 54, 60
spektrum, 20, 21, 30, 48, 51, 52, 53, 54, 55, 57, 59, 64, 66, 68, 69, 72, 74, 77, 80, 81
STC, 16, 72
syntetizátor
  formantový, 41
  modulárny, 81
  založený na akustickom spájaní jednotiek, 41
syntéza
  aditívna, 47, 49, 54, 61
  artikulačná, 28, 29
  difónová, 29, 156, 159
  formantová, 28, 29
  HMM, 28, 30
  na špecifickej oblasti, 29
  spájaním jednotiek, 28
  založená na jednotkovom výbere, 28
syntéza reči, 7, 15, 17, 19, 26, 27, 28, 29, 34, 35, 36, 39, 40, 41, 43, 47, 56, 86, 87, 88, 90, 111, 121, 159

## T

tagovanie, 93
text-to-speech, 16, 28, 35, 43, 44, 45, 90, 115
token, 107, 108, 109, 112, 121, 122, 123, 127, 129
tokenizácia, 91
transformácia
  diskrétna Fourierova, 13, 50, 51, 52, 53, 57, 68
  Fourierova, 50, 52, 54, 69
  rýchla Fourierova, 13, 77, 81
Twin VQ, 16, 73

## U

úroveň
  akustická, 39
  artikulačná, 39
  lingvistická, 39
  vnemová, 39

## V

vokodér
  homomorfný, 69
  kanálový, 65, 68
vzorkovanie, 63

## W

Wave, 75, 76

## X

XML, 16, 88, 89, 100, 103, 115, 116, 117, 119, 127, 143, 144, 159

## Z

zmena
  energie, 142
  tempa, 141
  základnej hlasivkovej frekvencie, 141

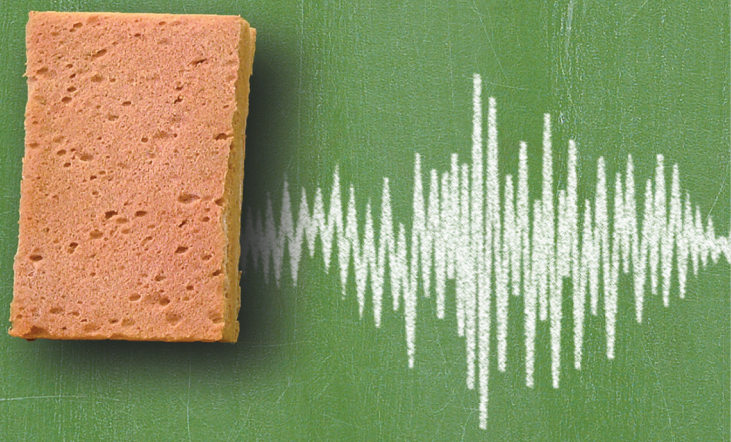
Ing. Renata Rybárová, PhD., doc. Ing. Gregor Rozinaj, PhD.

**METÓDY UČENIA PRE SYNTÉZU REČI**

Vydalo: Nakladateľstvo STU v Bratislave
Bratislava 2013

Rozsah 172 strán, 54 obrázkov, 20 tabuliek, 8,468 AH
Náklad: 120 výtlačkov, 120 ks CD
1. vydanie

ISBN 978-80-227-4033-3 (tlačená verzia)
ISBN 978-80-227-4034-0 (CD verzia)



ISBN 978-80-227-4033-3