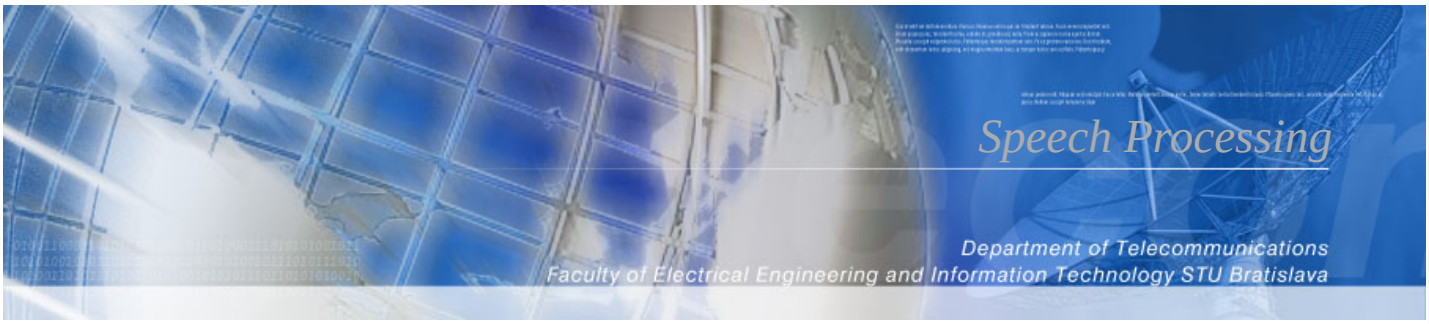

MENU

Home

TTS - Text To Speech

ASR - Automatic Speech Recognition

Dialog Systems [in Slovak]

Talking Head

Mobile Applications

Sinusoids - HNM System

Intelligent Terminal - Intelligent Kiosk

Workshop Redzur

Výber elementov syntézy

Štatistický výskum fonetických elementov v slovenčine

Prvým krokom pri výbere vhodných kandidátov na elementy syntézy je dôkladná štatistická analýza fonetických elementov daného jazyka. Jazykovedný výskum v oblasti kvantitatívnych charakteristík slovenčiny má na Slovensku nemalú tradíciu. Hlavným ťažiskom výskumu je frekvenčná analýza elementov (frekvencia elementov v texte), no skúma sa aj ich časová následnosť, či pravdepodobnosť ich asociácie, entropia, redundancia. Publikované štatistické výsledky zachytávajú väčšinou situáciu v písomnom prejave, zriedkavejšie sú štúdie ústneho prejavu.
Elementy slovenského jazyka je možné rozdeliť na ortografické a fonetické nasledovne **[SAV]**:

Ortografické elementy:

- graféma: element charakterizujúci základnú jednotku písomného systému (napr. slovo „včela“ sa skladá z nasledujúcich grafém: v, č, e, l, a).
- digram: element charakterizujúci dvojicu susediacich grafém (napr. slovo „včela“ sa skladá z nasledujúcich digramov: vč, če, el, la).
- trigram: element charakterizujúci trojicu susediacich grafém (napr. slovo „včela“ sa skladá z nasledujúcich trigramov: vče, čel, ela).

Fonetické elementy:

- hláska (fonéma): element charakterizujúci najmenšiu artikulačno-akustickú jednotku reči (napr. slovo „včela“ sa skladá z nasledujúcich hlások: f, č, e, l, a).
- dvojkombinácia hlások: element charakterizujúci dvojicu susediacich hlások (napr. slovo „včela“ sa skladá z nasledujúcich dvojkombinácií: fč, če, el, la).
- alofóna: predstavuje fonému v rôznych pozičných obmenách daných ľavým i pravým kontextom, v ktorom sa fonéma nachádza.
- difóna: element používaný pri signálovom spracovaní reči. Označuje úsek rečového signálu od polovice jednej hlásky po polovicu nasledujúcej hlásky. Názov difóny sa tvorí spojením znakov oboch hlások, ktoré obsahuje. Difóny sú využívané pri segmentácii reči najmä z toho dôvodu, že veľká časť akustickej informácie, dôležitá pre identifikáciu hlások, leží v prechodoch medzi hláskami. Výhodou difón je, že vo svojom centre zachovávajú prechodovú koartikulačnú informáciu a ustálené okrajové časti sú vhodné na spájanie s inými elementmi. Pri difónach sa berú do úvahy aj úseky ticha. Sú používané pri segmentácii začiatku a konca slova, kde rečový signál prechádza z ticha do prvej hlásky, resp. z poslednej hlásky do ticha. Preto sa difónový rozpis vždy začína a končí úsekom ticha (napr. slovo „včela“ sa skladá z nasledujúcich difón: /f, fč, če, el, la, a/).
- trojkombinácia hlások: element charakterizujúci trojicu susediacich hlások (napr. slovo „včela“ sa skladá z nasledujúcich trojkombinácií: fče, čel, ela).
- trifóna: element používaný pri signálovom spracovaní reči. Označuje úsek rečového signálu od polovice hlásky cez celú nasledujúcu hlásku až po polovicu ďalšej hlásky. Názov trifóny sa tvorí spojením znakov troch hlások, ktoré obsahuje. Pri trifónach sa podobne ako aj pri difónach berie do úvahy úsek ticha. Postup trifónovej segmentácie slova je rovnaký ako pri difónovej segmentácii, začína sa a končí úsekom ticha (napr. slovo „včela“ sa skladá z nasledujúcich trifónov: /fč, fče, čel, ela, la/).
- slabika: je fonetický útvar, ktorý obsahuje samohláskové jadro plus voliteľné počiatočné alebo koncové spoluhlásky, resp. skupiny spoluhlások. Slabika môže byť reprezentovaná aj samostatnou samohláskou. Môže obsahovať prechody spoluhláska – samohláska, ako aj prechody samohláska – spoluhláska vrátane väčšiny koartikulácií a iných fonologicko-fonetických javov vnútri svojich hraníc. Dĺžka slabiky nie je konštantná, je premenlivá. Fonologická teória zatiaľ neobsahuje pravidlá, ktoré by slabikovú hranicu presne určovali, preto sa v praxi stretávame s jej rôznymi definíciami.
- demislabika: rozdeľuje slabiku na 2 časti. Rozdelenie sa robí v samohláskovom jadre slabiky (zachovávajú sa spoluhláskové zhluky vo vnútri slabiky), čím sa behom reťazenia redukujú problémy s koartikuláciou. Jednotky sú potom typu KV alebo VK, kde K predstavuje žiadnu, jednu alebo viac spoluhlások a V je polovica samohlásky. Slovo „Vianoce“ takýmto spôsobom môžeme prepísať do tvaru: [via ia no o ce e]. Ich počet sa v anglickom jazyku pohybuje okolo 1000. Ukázalo sa, že sú vhodné pre syntézu nemčiny, v ktorej práve zhluky spoluhlások hrajú veľkú úlohu.

Z výsledkov frekvenčných analýz elementov uvedených v **[SAV]** je pre návrh databázy pre difónový rečový syntetizátor najdôležitejšia frekvenčná analýza dvojkombinácií hlások a difón (tab.1). Difóna a dvojkombinácia hlások sú elementy veľmi podobné. Dvojkombinácie hlások tvoria z hľadiska početnosti výskytu podmnožinu difón, pretože difóny môžu obsahovať naviac úsek ticha. Dvojkombinácia hlások je element používaný vo fonetickom výskume, pričom element difóna nachádza svoje uplatnenie pri segmentácii a spracovaní rečového signálu.

Dvojkombinácia hlások	Výskyt [%]	Dvojkombinácia hlások	Výskyt [%]	Dvojkombinácia hlások	Výskyt [%]
p r	1,37	o s	0,82	e n	0,64
o v	1,31	n o	0,80	s k	0,63
p o	1,25	ľ i	0,79	m e	0,63
n a	1,21	l a	0,77	t a	0,62
ň e	1,16	d o	0,76	ľ e	0,62
k o	1,10	v e	0,75	r i	0,62

s t	1,10	v o	0,73	j e	0,60
v a	1,06	o r	0,73	o b	0,56
r e	1,05	h o	0,72	e i<>	0,55
r o	1,04	v i	0,70	ť i	0,53
t o	1,02	l o	0,70	ď e	0,52
r a	0,96	ť e	0,68	m i	0,52
o u<>	0,88	s a	0,66	a k	0,51
o m	0,86	s ť	0,64	ka	0,51

Tab.1 Najčastejšie sa vyskytujúce dvojkombinácie hlások v slovenskom jazyku.

Syntetizátory sú podľa pravidiel obvykle konštruované tak, aby mohli generovať slová a vety z neobmedzeného slovníka. Stavebnými jednotkami, z ktorých je konštruovaná reč ich spájaním, sú tu najčastejšie alofóny, fonémy, difóny, demislabiky, či slabiky. Využívanie týchto základných jednotiek prináša so sebou výhody, ale i nevýhody.

Aplikácia fonémy ako konštrukčnej jednotky pre rečovú syntézu je výhodná z niekoľkých hľadísk. Predovšetkým, inventár rôznych foném je pomerne malý, takže nároky na ich vyhľadávanie a pamäť sú nízke. Ďalej sa pre každý jazyk dajú vytvoriť (produkčné) pravidlá a s ich pomocou sa dájú automaticky vytvárať bežné vety či slová daného jazyka z postupnosti foném. Tento proces sa nazýva **fonetická transkripcia.** Nevýhodou fonémy ako konštrukčnej jednotky je to, že fonéma je skôr logický reprezentant celej skupiny rečových zvukov – alofón. Tie obsahujú aj určitý stupeň koartikulácie a pre kvalitnejšiu syntézu je teda výhodnejšie využívať práve tieto stavebné jednotky. Aj pre alofóny sa pri syntéze dajú navrhnúť pravidlá pre tzv. alofonickú transkripciu, alebo pravidlá, ktoré by doplnili fonetickú transkripciu. Čím detailnejšie urobíme alofonickú transkripciu, t.j. čím viac rôznych alofonických jednotiek budeme využívať, tím bude syntetizovaná reč kvalitnejšia. Na druhej strane si vzrastajúci počet alofón vyžaduje väčšie množstvo pamäte a väčšie množstvo pravidiel. Rozumným kompromisom býva využívanie 100 až 200 alofonických variantov. Zvolené fonémy, či alofóny je potrebné veľmi starostlivo extrahovať zo signálu nahovorenej reči a zakódovať vo forme postupnosti parametrov vhodných pre riadenie formatového, či LPC syntetizátora. Pre zvýšenie kvality syntetizovanej reči je dobré aj pri spájaní alofón aplikovať istý interpolačný proces, aby boli vyhladené stopy náhlych zmien formantov či reflexných koeficientov.

Pri aplikácii difónov a demislabík ako konštrukčných jednotiek pri syntézy reči sa predpokladá, že spravíme najprv fonetickú transkripciu syntetizovanej správy, a potom zo slovníka difónov, poprípade demislabík, vyberieme zodpovedajúce jednotky a tie pospájame. Výhodou difónov je to, že nesú ako bolo už spomínané, dôležitú koartikulačnú informáciu. Keďže pri ich spájaní spájame rovnaké fonémy, redukujú sa podstatne aj požiadavky na interpoláciu prechodu. Demislabiky na rozdiel od difónov obsahujú podľa potreby celý počiatočný alebo koncový spoluhláskový zhluk, ktorý sa pomocou spájania difónov realizuje obtiažnejšie. Demislabiky môžu byť taktiež efektívne využité aj pri ovplyvňovaní rytmu reči. Bolo totiž dokázané, že na rytmus reči má významný vplyv práve dĺžka hlások vo vnútri zhluku spoluhlások. Napriek tomu sa počet difónov a demislabík sa odhaduje na 2000, čo prináša ťažkosti predovšetkým s ich extrakciou z reálnej reči, ale aj s ich uchovávaním a manipuláciou.

Slabika sa ako základná stavebná jednotka pri rečovej syntézy veľmi nepoužíva. Je to z toho dôvodu, že ich existuje veľmi veľký počet, ktorý je mnohonásobne väčší než počet difónov a demislabík. Aj keď sú koartikulačné efekty vo vnútri slabík prirodzene obsiahnute, v miestach ich spájania chýbajú a musia byť „dopĺňané“ interpoláciou.

Pokiaľ vytvárame konkrétne predhovory, ktoré by mali byť uložené v pamäti a podľa potreby syntetizované, je potrebné zostaviť reťazce zodpovedajúcich fonetických symbolov a doplniť ich o prozodické značky (označenie prízvuku, pauzy, a pod.). Pre jednotlivé vety predhovoru potom musíme vytvoriť samostatné obrazy priebehu základného hlasivkového tónu. Ak by bol vo fáze analýzy pripravený pre každý fonetický symbol a prozodickú značku odpovedajúci reťazec akustických príznakov k riadeniu formantového syntetizátora, popr. LPC syntetizátora, potom vytvorený reťazec znakov, doplnený o obraz základného tónu, popr. obraz intenzity, nesie dostatočnú informáciu pre rečovú syntézu danej správy. Pokiaľ pripravuje podobné predhovory konštruktér dopredu, môže samozrejme reťazec znakov podľa potreby upravovať tak dlho, až výsledný predhovor znie čo najprirodzenejšie. Obtiažnejšia situácia nastáva, ak má byť syntéza reči vykonávaná v reálnom čase priamo z predtým neznámeho písaného textu.

Literatúra:

[SAV] ROZINAJ G. a iní: Úloha výskumu a vývoja: Inteligentné rečové komunikačné rozhranie štátneho programu Budovanie informačnej spoločnosti – D 1.3 – Modul syntézy reči (Analýza súčasného stavu a návrh riešenia), Košice, jún 2004, s. 76 – 89