

KOMPIUTERINĖ LINGVISTIKA/ COMPUTATIONAL LINGUISTICS

Lietuvių kalbos difonų bazių optimizavimas

Pijus Kasparaitis, Žilvinas Ledas

crossref <http://dx.doi.org/10.5755/j01.sal.0.19.947>

Anotacija. Šiame darbe nagrinėjamas lietuvių kalbos difonų bazių, kurios naudojamos balso sintezėje iš teksto, sudarymas ir optimizavimas. Difonų bazių kūrimas pradedamas nuo difonų sąrašo sudarymo. Kai kurios fonemos negali sudaryti difonų dėl fonotaktikos apribojimų, tačiau jei vadovausimės vien fonotaktikos taisyklėmis, tai į sąrašą nepateks dalis difonų, reikalingų kitų kalbų žodžiams sintezuoti, tačiau pateks daug praktiškai nevartojamų difonų. Darbe atliktas difonų vartosenos dažnių statistinis tyrimas ir parodyta, kad sudarius difonų bazes tik iš dažniausiai vartojamų difonų ženkliai sumažėja bazių apimtis išlaikant pakankamai aukštą teksto padengimo difonais laipsnį.

Darbe taip pat aprašomas difonų bazės sumažinimas, remiantis garsų panašumu, kai vienas difonas keičiamas kitu panašiai skambančiu difonu, bei garsų ištempimu, kai trūkstamo difono vieta užpildoma ištempiant gretimus garsus. Atlikti audiciniai eksperimentai su difonais, sudarytais iš balsio ar dvigarsio prieš sprogstamąjį (garsų ištempimo atveju ir prieš friktyvinį) priebalsį. Nustatytos difonų klasės, kurioms taikant šiuos metodus sintezuoto garso kokybė nepablogėja, ir kuriems pablogėja nežymiai, bei įvertintas difonų bazių apimties sumažėjimas.

Raktiniai žodžiai: balso sintezė iš teksto, difonas, difonų bazės optimizavimas, garsų panašumas, difonų vartosenos dažniai.

Įvadas

Balso sintezės sistemose iki šiol bene populiariausias sintezės vienetas yra difonas – garso segmentas nuo vienos fonemos vidurio iki kitos fonemos vidurio (Taylor, 2009; p.425). Norint sudaryti difonų sąrašą, imamas nagrinėjamos kalbos fonemų sąrašas ir sudaromos visos galimos poros. Tada atmetamos tos, kurios yra negalimos, remiantis tos kalbos fonotaktikos taisyklėmis. Reikia nepamiršti, kad fonemų derinys, kuris negalimas žodžio viduryje, gali atsirasti dviejų žodžių sandūroje (sintezėje žodžių ribos ignoruojamos). Todėl gaunamas didelis difonų skaičius ir yra sunku sugalvoti pavyzdžių, iš kurių galima būtų iškirpti visus reikalingus difonus. Šiame darbe į galimų/ negalimų difonų problemą pasižiūrėta remiantis ne fonotaktika, o statistiniais vartosenos dažniais. Kiti autoriai jau yra nagrinėję kitokių lietuvių kalbos segmentų dažnius: fonemų (Karosienė & Girdenis, 1993), (Kazlauskienė & Raškinis, 2009), skiemenų (Karosienė & Girdenis, 1994), (Kazlauskienė & Raškinis, 2008), tačiau difonų dažniai dar nebuvo nagrinėti.

Difonų bazę naudojant balso sintezėje, gali pasitaikyti atvejų (nors ir retai), kai nerandamas reikiamas difonas. Vienas iš problemos sprendimo būdų – vietoje bazėje neegzistuojančio difono panaudoti kitą panašų, kuris bazėje yra (Taylor, 2009; p.517). Šis metodas naudojamas MBROLA difonų bazėse (Kasparaitis, 2005). Tačiau toks sprendimas ne visada geras, nes klausytojas gali pajusti, kad panaudotas ne tas garsas. Kitas problemos sprendimo būdas – trūkstamą difoną sintezuoti iš fonemų puselių, t. y. imti garso segmentą nuo fonemos vidurio iki pabaigos ir prijungti kitos fonemos segmentą nuo pradžios iki vidurio.

Bazėje reikiamų difonų gali nebūti ir tuo atveju, kai difonų bazę mažiname specialiai. Tai gali būti naudinga, siekiant įgyvendinti sintezę išmaniuosiuose telefonuose ar kituose ribotos talpos įrenginiuose. Difonų bazės mažinimas yra aktualus ir taupant laiką, kuris reikalingas tokioms bazėms paruošti. Be to, tokių sintezės sistemų kokybė tiesiogiai priklauso nuo difonų bazės kokybės, tai esant mažesnei bazei, kiekvienam individualiam difonui galima skirti daugiau laiko ir jį apdoroti kokybiškiau.

Nors difonų bazių sumažinimas paprastai reikalingas sintezuojant balsą vienetų parinkimo metodu, kurio esmė – saugoti po kelis to paties garso variantus ir parinkti tinkamiausią, tačiau galime laikyti, kad šiame darbe nagrinėjama panaši problema, tik renkama ne iš vienetų, o iš panašių garsų. Teigiama (Black & Taylor, 1997), kad iki 20 proc. sintezės vienetų pašalinimas, dažniausiai nedaro neigiamos įtakos sintezuotos kalbos kokybei (o kartais ją net pagerina), o iki 50 proc. galima pašalinti, nebijant ypač stipriai pakenkti sintezuotos kalbos kokybei.

Fonemų ir difonų žymėjimas šiame tekste

Šiame tekste bus naudojamas (Kasparaitis, 2005) aprašytas fonemų žymėjimas. Kirčiuoti ir nekirčiuoti garsai laikomi skirtingomis fonemomis. Visas 92 fonemų sąrašas yra toks:

/_/, /a/, /e/, /i/, /o/, /u/, /A/, /E/, /I/, /O/, /U/, /aa/, /ea/, /ee/, /ie/, /ii/, /oo/, /uo/, /uu/, /Aa/, /Ea/, /Ee/, /Ie/, /Ii/, /Oo/, /Uo/, /Uu/, /aA/, /eA/, /eE/, /iE/, /iI/, /oO/, /uO/, /uU/, /p/, /p/, /t/, /t/, /k/, /k/, /b/, /b/, /d/, /d/, /g/, /g/, /s/, /s/, /z/, /z/, /S/, /S/, /Z/, /Z/, /ts/, /ts/, /tS/, /tS/, /dz/, /dz/, /dZ/, /dZ/, /x/, /x/, /h/, /h/, /f/, /f/, /v/, /v/, /w/, /W/, /j/, /j/, /J/, /l/, /l/, /L/, /L/, /l/, /m/, /m/, /M/, /M/, /n/, /n/, /N/, /N/, /r/, /r/, /R/, /R/.

Čia trumpieji nekirčiuoti balsiai žymimi viena mažąja raide, pvz., /a/, trumpieji kirčiuoti balsiai žymimi viena didžiąja raide, pvz., /A/. Ilgieji nekirčiuoti balsiai žymimi dviem mažosiomis raidėmis, pvz., /aa/, ilgieji tvirtapradžiai balsiai žymimi viena didžiąja ir viena mažąja raidėmis, pvz., /Aa/, ilgieji tvirtagalieji balsiai žymimi viena mažąja ir viena didžiąja raidėmis, pvz., /aA/. Ilgieji balsiai „e“ žymimi dviem raidėmis /ea/, o ilgasis balsis „ė“ – raidėmis /ee/. Priebalsių minkštumas žymimas apostrofu, pvz., /b'/, priebalsiai „š“ ir „ž“ žymimi atitinkamai /S/ ir /Z/, o afrikatos „c“ ir „č“ žymimos atitinkamai /ts/ ir /tS/.

Tvrtagalių mišriųjų dvigarsių abu garsai žymimi didžiosiomis raidėmis, pvz., /A/ ir /L/. Nekirčiuota antroji sudėtinių dvibalsių, pvz. „ai“ ar „au“, dalis žymima atitinkamai /j/ ir /w/, o kirčiuota – /J/ ir /W/.

Difonai žymimi rašant dvi fonemas atskirtas brūkšneliu, pvz., /aA-d/, /t'-i/.

Difonų vartosenos dažnių statistinis tyrimas

Siekdami suprasti, kaip dažnai difonų bazėje nerandama reikiamų difonų, ir ar verta stengtis sukurti bazę, kurioje būtų visi galimi difonai, panagrinėkime difonų vartosenos sintezuojamame tekste statistiką. Eksperimentuose buvo naudojamas apie milijono (1.004.879) žodžių tekstynas (Anbinderis & Kasparaitis, 2009), kuris buvo automatiškai sukirčiuotas ir sutranskribuotas, gauti 6.684.599 difonai. Teksto kirčiavimui ir transkribavimui buvo panaudota P. Kasparaičio sukurta programinė įranga (Kasparaitis, 1999), (Kasparaitis, 2000). Buvo apskaičiuotas kiekvieno difono pavartojimų skaičius, o gauti rezultatai pirmiausiai palyginti su dviejose identiškose struktūros difonų bazėje (Kasparaitis, 2005) esančių difonų sąrašu (sąrašą sudaro 5003 difonai) ir difonų, kurie laikomi negalimais, sąrašu (3461 difonas). Nustatyta, kad bent po vieną kartą buvo pavartotas 51 difonas, kurie laikyti negalimais, pvz.: /j'-k/ (Željko), /d-f/ (Hardfortšyro), /h'-k'/ (Pahkel) ir t. t., o jų bendras pavartojimų skaičius – 5678 arba 0,085 proc.. Taip pat pastebėta, kad net 1013 iš 5003 (arba 20,25 proc.) bazėje esančių difonų nepavartoti nė karto. Be to, difonų vartosenos dažniai labai skiriasi. Dažniausiai vartojamas

2 lentelė. Mokymo ir testavimo duomenų padengimas difonais.

Mokymo duomenys	Padengimas, proc.	100.00	99.95	99.90	99.50	99.00
	Difonų skaičius	3989.4	3214.0	2993.8	2423.0	2149.6
	Dalis tarp visų difonų	47.13	37.97	35.37	28.63	25.40
Testavimo duomenys	Kiek skirtingų difonų nerasta	52.4	419.4	606.2	1165.2	1438.6
	Kiek iš viso difonų nerasta	74.4	901.0	1582.2	7083.8	13818.0
	Nerastų difonų dalis, proc.	0.006	0.067	0.118	0.530	1.034
	Kiek nepanaudota difonų	453.6	45.2	11.8	0	0

Difonų sintezė iš garsų puselių

Difonų pakeitimo metodo trūkumas yra tai, kad ne visada tarp likusių difonų pavyksta rasti panašų į trūkstamą. Pvz., atmetus difonus, kurie tekstyne buvo rasti mažiau nei 6

difonas /t'-i/ pavartotas 72978 kartus, o difonas /dz'-i/ pavartotas tik 36 kartus ir užima 2762 vietą. Kokią dalį teksto koks skaičius difonų padengia, pavaizduota 1 lentelėje.

1 lentelė. Teksto padengimas difonais.

Padengimas, proc.	50	80	90	95	98	99
Difonų skaičius	191	645	1017	1389	1854	2178

Tas faktas, kad kai kurie difonai pavartoti itin retai arba visai nepavartoti, o tarptautiniuose žodžiuose vis vien pasitaiko difonų, kurie šiaip yra negalimi lietuvių kalboje, leidžia manyti, kad neverta bandyti rankiniu būdu sudaryti galimų ir negalimų difonų sąrašų, o geriau pasikliauti statistiniais vartosenos dažniais. Buvo atlikti tokie eksperimentai: turimas milijono žodžių tekstynas padalintas į 5 maždaug vienodas dalis, viena dalis (vadinsime testavimo duomenimis) naudota testavimui, o likusios keturios (vadinsime mokymo duomenimis) – sąrašui sudaryti. Buvo apskaičiuojamas difonų vartosenos dažnis mokymo duomenyse, tada iš dažniausiai pavartotų difonų sudaromas sąrašas. Difonų skaičius sąraše parenkamas taip, kad sąrašui priklausantys difonai padengtų tam tikrą dalį mokymo duomenų. Testavimo metu randama, kokią dalį testavimo duomenų padengia sudarytas sąrašas. Rezultatai pateikti 2 lentelėje. Eksperimentai atlikti 5 kartus, testavimui imant vis kitą tekstyno dalį, todėl pateikti rezultatų vidurkiai.

Iš 2 lentelės galima matyti, kad mokymo ir testavimo duomenų padengimas difonais skiriasi nedaug, ir skirtumas santykinai mažėja mažinant padengimo procentą.

Tam, kad klausytojas nepajustų sintezuotos kalbos kokybės pablogėjimo, reikia, kad atvejai, kai nerandamas reikiamas difonas, būtų pakankamai reti. Remiantis 2 lentele, jei norėtume difonais padengti 99,9 proc. sintezuojamo teksto, tai reikėtų apie 3000 difonų (vietoje 5003 naudotų iki šiol), tuomet būtų nerasta apie 600 skirtingų difonų, tačiau jie sudarytų tik kiek daugiau nei 0,1 proc., o tai yra tik vienas atvejis iš tūkstančio.

kartus, jau tenka naudoti tokius (nelabai vykusius) pakeitimus: /dz-h/ → /dz-g/, /dZ'-dZ'/ → /dZ'-d'/, /dz-l/ → /ts-l/, /f-dz'/ → /v'-d'/, /f-S/ → /f-S'/, /f-x/ → /f-k/, /h-z/ → /x-s/, /ie-ie/ → /ie-j'/.

Kitas metodas – trūkstamus difonus sintezuoti iš fonemų puselių, kurias galima išsikirti iš tų pačių difonų. Infor-

maciją apie tai, iš kurio difono kokią puselę kirpti, galima saugoti arba kiekvienai fonemai nurodant difoną, iš kurio kirpti kairę ir dešinę puselę (pvz., /a/, /a- /, / -a/), arba tai nurodyti kiekvienam trūkstamam difonui (pvz., /a-k/, /a- /, /e-k/). Antrasis metodas žymiai lankstesnis, tačiau pakeitimų sąrašo sudarymas reikalauja daugiau pastangų. Difonų sintezė iš fonemų puselių turi dar ir tokį trūkumą, kad padaugėja garsų sujungimų, tačiau šis padidėjimas nežymus, jei reikiamo difono nerandama pakankamai retai. Išsamiau šio metodo šiame darbe nenagrinėsime.

Difonų bazės sumažinimas, remiantis garsų panašumu

Difonų pakeitimų metodą galima pritaikyti ir mažinant difonų bazės apimtį. Jei tarsime, kad kelių skirtingų fonemų puselės skamba vienodai (pvz., priebalsių „k“, „p“, „t“ tylos dalis prieš sproginimą), tai visus difonus, į kuriuos tos puselės įeina, galima pakeisti vienu difonu, nepriklausomai nuo vartojimo dažnio (pvz., vietoje /a-k/, /a-p/ ir /a-t/ vartoti tik vieną iš jų).

1.1. Difonai „balsis-sprogstamasis priebalsis“

Sprogstamųjų priebalsių kairės pusės skamba panašiai, todėl buvo tirta, ar galima difonus tipo „balsis-sprogstamasis priebalsis“ pakeisti kitais difonais su sprogstamaisiais priebalsiais. Nagrinėtos tokios skardžiųjų ir dusliųjų sprogstamųjų priebalsių grupės:

- B* = {B, B'}; kur
B = {/b/, /d/, /g/, /dz/, /dZ/};
B' = {/b'/, /d'/, /g'/, /dz'/, /dZ'/'};
- P* = {P, P'}; kur
P = {/p/, /t/, /k/, /ts/, /tS/};
P' = {/p'/, /t'/, /k'/, /ts'/, /tS'/'};

Visuose tolimesniuose eksperimentuose buvo naudota tik viena iš dviejų (G. Deksnio balso) difonų bazių (Kasparai-

3 lentelė. Apibendrinti skardžiųjų sprogstamųjų priebalsių grupės keturių audicinių eksperimentų rezultatai (čia V žymi balsį).

	V-b	V-d	V-g	V-b'	V-d'	V-g'	V-dz	V-dZ	V-dz'	V-dZ'
V-b	-	-	-	-	2	1	4	2	4	4
V-d	1	1	1	1	2	2	3	2	4	3
V-g	1	2	2	2	3	3	5	3	4	5
V-b'	-	1	-	-	3	5	5	4	6	3
V-d'	2	2	2	2	-	2	2	-	-	-
V-g'	-	-	-	-	-	-	-	-	-	-
V-dz	-	-	-	-	-	-	-	-	-	-
V-dZ	-	-	-	-	-	-	-	-	-	-
V-dz'	-	-	-	-	-	-	-	-	-	-
V-dZ'	-	-	-	-	-	-	-	-	-	-

Iš 3 lentelės matome, kad difonus, kurių antrasis dėmuo yra afrikata /dz/, /dZ/, /dz'/, /dZ'/'/, galima keisti tarpusavyje arba keisti difonais, kurių antrasis dėmuo /b/, /d/, /g/, /b'/, /d'/, /g'/'/, ir šis pakeitimas neturi įtakos girdimam garsui. Atsižvelgti į kietumą/minkštumą nereikia, pvz., /a-dZ'/'/

tis, 2005). Su programa PRAAT (<http://www.fon.hum.uva.nl/praat/>) difonai buvo kopijuojami vienas prie kito ir taip gaunami sintezuoti tokio tipo garsų junginiai: „pauzė balsis priebalsis balsis pauzė“. Atlikti 4 eksperimentai su skirtingais balsiais. Priebalsių grupei B* šiuos eksperimentus galima būtų nusakyti tokiais fonemų junginiais:

- /_ aA B aA _/;
/_ eA B' eA _/;
- /_ a B a _/;
/_ e B' e _/;
- /_ uU B uU _/;
/_ uU B' uU _/;
- /_ u B u _/;
/_ u B' u _/.

Difonais šie fonemų junginiai užrašomi taip:

- /_aA/, /aA-B/, /B-aA/, /aA-_ / ir t. t.

Pakeičiant antrąjį difoną visais galimais variantais grupės viduje (t. y., /aA-b/ keičiant į /aA-d/, /aA-g/, /aA-b'/, /aA-d'/, /aA-g'/, /aA-dz/, /aA-dZ/, /aA-dz'/, /aA-dZ'/'/), buvo sintezuojami garsų junginiai. Šie sintezuoti garsai buvo dviejų auditorių iš klausos įvertinti. Keturių eksperimentų su skirtingais balsiais rezultatai skardžiųjų sprogstamųjų priebalsių grupei pateikti 3 lentelėje. Stulpeliai nusako, koks difonas buvo panaudotas, o eilutės – vietoje kokio. Lentelės langeliuose įrašyti skaičiai rodo, kiek kartų sintezuotas junginys skambėjo nenatūraliai, t. y. vietoje nagrinėjamo priebalsio buvo girdimas kitas priebalsis. Kiekviename langelyje maksimalus skaičius gali būti 8, o „-“ reiškia, kad visi junginiai skambėjo natūraliai.

galima keisti į /a-g/. Šį faktą galima paaiškinti tuo, kad afrikatų sprogstamoji dalis yra pakankamai ilga.

Taip pat matome, kad difonus, kurių antrasis dėmuo /b/, /d/, /g/, /b'/, /d'/, /g'/'/, taip pat dažniausiai galima keisti vienus kitais neatsižvelgiant į kietumą/minkštumą. Natūraliai skamba 85,6 proc. atvejų (nenatūraliai skamba 41 iš 288),

keičiant tik kietaisiais priebalsiais – 89,6 proc. (blogai 15 iš 144), o keičiant priebalsiu /b/ – 92,7 proc. (blogai 4 iš 48). Keisti afrikatomis neverta, nes natūraliai skamba tik 67,2 proc. atvejų (blogai 63 iš 192). Dar verta atkreipti dėmesį į tai, kad net ir naudojant reikiamą difoną, 3 atvejai iš 80 skambėjo nenatūraliai (išskirta pajuodintu šriftu).

Antrosios (dusliųjų) sprogstamųjų priebalsių grupės audiocinių eksperimentų rezultatai čia nepateikiami, nes visais atvejais sintezuoti junginiai skambėjo natūraliai. Tai reiškia, kad difonus, kurių antrasis dėmuo yra iš aibės P*, galima laisvai keisti vienus kitais ir šis pakeitimas nepablogina girdimo garso kokybės, t. y. duslusis sprogstamasis priebalsis neturi įtakos prieš jį esančiam balsiui. Atsižvelgti į kietumą/ minkštumą taip pat nereikia, pvz., /a-tS'/ galima keisti į /a-p/.

1.2. Difonai „dvigarsis-sprogstamasis priebalsis“

Eksperimentai, panašūs į ankstesniame poskyryje aprašytus, buvo atliekami ir su sprogstamaisiais priebalsiais, einančiais po dvigarsio. Šiame tekste dvigarsiais vadinsime mišriuosius dvigarsius ir sudėtinius dvibalsius. Buvo nagrinėtos tos pačios dvi – skardžiųjų ir dusliųjų – sprogstamųjų priebalsių grupės.

Analogiškai su programa PRAAT buvo sintezuojami tokio tipo garsų junginiai: „pauzė dvigarsis priebalsis balsis pauzė“.

Pirmaisiai buvo atlikti 4 eksperimentai su skirtingais mišriaisiais dvigarsiais. Priebalsių grupei B* šiuos eksperimentus galima būtų nusakyti taip:

- /_ a l B A _/; /_ A L B a _/; /_ u l B U _/; /_ U L B u _/;
/_ e l' B' E _/; /_ E L' B' e _/; /_ u l' B' U _/; /_ U L' B' u _/;
- /_ a m B A _/; /_ A M B a _/; /_ u m B U _/; /_ U M B u _/;
/_ e m' B' E _/; /_ E M' B' e _/; /_ u m' B' U _/; /_ U M' B' u _/;
- /_ a n B A _/; /_ A N B a _/; /_ u n B U _/; /_ U N B u _/;
/_ e n' B' E _/; /_ E N' B' e _/; /_ u n' B' U _/; /_ U N' B' u _/;
- /_ a r B A _/; /_ A R B a _/; /_ u r B U _/; /_ U R B u _/;
/_ e r' B' E _/; /_ E R' B' e _/; /_ u r' B' U _/; /_ U R' B' u _/.

Difonais šie fonemų junginiai užrašomi taip:

- /_ -a/, /a-l/, /l-B/, /B-A/, /A- _/ ir t. t.

Pakeičiant trečiąją difoną galimais variantais grupės viduje atsižvelgiant į kietumą/ minkštumą (t. y., /l-b/ keičiant į

/l-d/, /l-g/, /l-dz/, /l-dZ/, o /l'-b'/ keičiant į /l'-d'/, /l'-g'/,

/l'-dz'/, /l'-dZ'/), buvo sintezuojami garsų junginiai.

Taip pat buvo atlikti dar 2 eksperimentai su sudėtiniais dvibalsiais. Grupei B* šiuos eksperimentus galima būtų nusakyti taip:

- /_ a j B A _/; /_ A J B a _/; /_ u j B U _/; /_ U J B u _/;
/_ e j B' E _/; /_ E J B' e _/; /_ u j B' U _/; /_ U J B' u _/;
- /_ a w B A _/; /_ A W B a _/;
/_ e w B' E _/; /_ E W B' e _/.

Analogiškai užrašius šiuos fonemų junginius difonais ir pakeičiant trečiąją difoną visais galimais variantais grupės viduje (t. y., /j-b/ keičiant į /j-d/, /j-g/, /j-b'/, /j-d'/, /j-g'/,

/j-dz/, /j-dZ/, /j-dz'/, /j-dZ'/), buvo sintezuojami garsų junginiai.

Šie sintezuoti garsai buvo dviejų auditorių iš klausos įvertinti ir rezultatai surašyti į lenteles, analogiškas 3 lentelei. Dėl didelės lentelių apimties jos čia nepateikiamos, tačiau apibendrinant galima daryti kelias išvadas.

Pirma, dusliųjų sprogstamųjų priebalsių grupės visų šešių eksperimentų visais atvejais junginiai skamba natūraliai. Keičiant difoną, į kurį įeina mišrusis dvigarsis, dar reikia atsižvelgti į kietumą/ minkštumą. Pvz., /l-tS'/ galima keisti į /l-p/, /l'-tS'/ – į /l'-p'/, o /j-tS'/ galima keisti tiek į /j-p/, tiek į /j-p'/.

Antra, skardžiųjų sprogstamųjų priebalsių grupės eksperimentų su mišriaisiais dvigarsiais atveju:

- Difonus, kurių antrasis dėmuo yra afrikata, beveik visada galima keisti difonais su sprogstamuoju priebalsiu. Atsižvelgti į kietumą/ minkštumą reikia. Pvz., /l-dz/ galima keisti į /l-b/, o /l'-dz'/ – į /l'-b'/. Natūraliai skamba 99,0 proc. atvejų (blogai 2 iš 192).
- Difonus, kurių antrasis dėmuo /b/, /d/, /g/, /b'/, /d'/, /g'/, galima keisti vienus kitais atsižvelgiant į kietumą/ minkštumą. Natūraliai skamba 92,7 proc. atvejų (blogai 21 iš 288). Dažniausia klaida, kai pakeitus difoną „mpriebalsis“ kitu difonu, garsas „m“ virsta garsu „n“.
- Pakeitimui nepatartina naudoti difonų, kurių antrasis dėmuo yra afrikata.

Trečia, skardžiųjų sprogstamųjų priebalsių grupės eksperimentų su dvibalsiais atveju:

- Vietoje dvibalsių „au“, „eu“ po pakeitimo dažnai pasigirsta „al“, „el“, todėl šioms dvibalsiams pakeitimų nenaudosime.
- Difonus, kurių antrasis dėmuo yra afrikata, visada galima keisti difonais su sprogstamuoju priebalsiu, neatsižvelgiant į kietumą/ minkštumą. Pvz., /j-dz'/ galima keisti į /j-b/.
- Difonus, kurių antrasis dėmuo /b/, /d/, /g/, /b'/, /d'/, /g'/, galima keisti vienus kitais atsižvelgiant į kietumą/ minkštumą. Natūraliai skamba 96,5 proc. atvejų (blogai 10 iš 288).

1.3. Difonų bazės sumažinimo galimybės, naudojant pakeitimus

Atsižvelgiant į difonų vartosenos dažnius, buvo sudarytas difonų sąrašas, vadovaujantis tokiais kriterijais:

1. Į sąrašą gali būti neįtraukiami tik tie difonai, kurie kirčiuojant 1 mln. žodžių tekstyną buvo aptikti mažiau kaip 6 kartus.
2. Į sąrašą įtraukti tik tie difonai, kurių balso įrašai yra difonų bazėje.
3. Į sąrašą įtraukti kai kurie difonai, kuriems sunku rasti gerus pakeitimus.

Remiantis šiais kriterijais gautas 3539 difonų sąrašas, kuris padengia 99,90 proc. mokymo duomenų, ir 4925 pakeitimai.

Tarp šių 3539 difonų yra 825 tokie, kurių antras dėmuo yra iš aibių B* arba P*. Jei pakeitimus naudosime vietoje difonų, kurių kairė pusė yra balsis arba dvigarsis, o dešinė pusė yra iš aibės P* arba fonemos /dz/, /dZ/, /dz'/, /dZ'/, sąrašas sumažėtų 367 difonais ir liktų 3172. Jei pakeitimus naudosime vietoje difonų, kurių kairė pusė yra balsis arba dvigarsis, o dešinė pusė yra iš aibių P* ir B*, sąrašas sumažėtų 571 difonu ir liktų 2968. Pasiękti geri rezultatai leidžia tikėtis, kad pakeitimus galima naudoti ir vietoje difonų, kurių kairėje pusėje bet kokia fonema, o dešinė pusė yra iš aibių P* ir B*. Šiuo atveju sąrašas sumažėtų 681 difonu, liktų 2858 difonai, o pakeitimų būtų 5606.

Difonų bazės sumažinimo įverčiai apibendrinti 4 lentelėje.

Garsų ištempimo metodas

Šiame darbe siūlomas metodas, leidžiantis sumažinti ne tik difonų skaičių, bet ir jungimų skaičių. Tarkime, kad difonų sekoje /l-a/, /a-b/, /b-e/ neturime difono /a-b/, tuomet galima pratęsti difono /l-a/ antrąją dalį, difono /b-e/ pirmąją dalį ir taip užpildyti trūkstamą difoną.

Mūsų manymu, šiuo metodu galima pakeisti difonus, kuriuos sudaro balsiai ar dvigarsių antrieji dėmenys prieš sprogstamuosius ir friktyvinius priebalsius. Be to, toks garsų padalinimas į nesikertančias aibes užtikrina, kad šį metodą niekada neteks taikyti dviem gretimoms difonams užpildyti.

Metodo tinkamumas vėlgi buvo tikrinamas programa PRAAT kopijuojant difonus vieną prie kito. Trūkstami difonai buvo užpildomi du kartus ištempiant dalį prieš tai esančio garso ir dalį po to esančio garso. Jeigu nagrinėtume tik balsį „a“ ir dvigarsius „al“ ir „el“, tai šiuos eksperimentus galima būtų nusakyti taip, kad buvo sintezuojami tokio tipo garsų junginiai:

- /_ C a b A _/; /_ C A b a _/;
/_ C aa b aA _/; /_ C aA b aa _/;
- /_ a l b A _/; /_ A L b a _/;
/_ e l' b' E _/; /_ E L' b' e _/.

4 lentelė. Difonų bazės sumažinimo galimybių apibendrinimas.

Tipas	Di-fonų sk.	Pakeitimų sk.	Proc. nuo visų difonų	Proc. nuo orig. bazės	Pastabos
Originali bazė	5003	3461	59,1	100	Originali difonų bazė, žr. (Kasparaitis, 2005).
Dažniausi difonai	3539	4925	41,8	70,7	4.3 skyrelio 1, 2 ir 3 kriterijai. Padengia 99,90 proc. atvejų.
Dažniausi difonai atmetus panašius	3172	5292	37,5	63,4	Pakeitimai vietoje difonų, kurių kairė pusė yra balsis arba dvigarsis, o dešinė pusė yra iš aibės P* arba fonemos /dz/, /dZ/, /dz'/, /dZ'/.
	2968	5496	35,1	59,3	Pakeitimai vietoje difonų, kurių kairė pusė yra balsis arba dvigarsis, o dešinė pusė yra iš aibių P* ir B*.
	2858	5606	33,8	57,1	Pakeitimai vietoje difonų, kurių kairė pusė – bet kokia fonema, o dešinė pusė yra iš aibių P* ir B*.
Dažniausi difonai ir ištempimo metodas	2572	4396	30,4	51,4	1496 difonai, kurie padengia 22,40 proc. atvejų, sintezuojami naudojant difonų ištempimą.

Difonais šie garsų junginiai užrašomi taip:

- /_ C/, /C-a/, /b-A/, /A-_/;
/_ C/, /C-A/, /b-a/, /a-_/;
/_ C/, /C-aa/, /b-aA/, /aA-_/;
/_ C/, /C-aA/, /b-aa/, /aa-_/;
- /_ a/, /a-l/, /b-A/, /A-_/;
/_ A/, /A-L/, /b-a/, /a-_/;
/_ e/, /e-l'/, /b'-E/, /E-_/;
/_ E/, /E-L'/, /b'-e/, /e-_/.

Čia C – atsitiktinai parinktas kietas priebalsis, o vietoje /b/ buvo imamos visos aibių B*, P* ir Z*, kur Z*={/z/, /Z/, /h/, /s/, /S/, /x/, /f/, /z'/, /Z'/, /h'/, /s'/, /S'/, /x'/, /f'/}, fonemos.

Trūkstamas difonas, kuris turėtų būti tarp antrojo ir trečiojo difonų, buvo sintezuojamas programos PRAAT trūkmės keitimo funkcija PSOLA ištempiant antrojo difono antrąją dalį ir trečiojo difono pirmąją dalį. Šie sintezuoti junginiai buvo dviejų auditorių iš klausos įvertinti ir rezultatai surašyti į lenteles. Dėl didelės lentelių apimties jos čia nepateikiamos, tačiau apibendrinant galima daryti kelias išvadas.

Pirma, garsų, kurie yra kintantys, t. y. „ie“, „uo“, „r“, negalima kokybiškai ištempti. Likusius garsus dažniausiai galima ištempinėti. Natūraliai skamba 95,1 proc. atvejų.

Antra, aiškaus dėsningumo, ar šis metodas geriau tinka skardiesiems ar dusliesiems, kietiesiems ar minkštiesiems, sprogstamiesiems ar pučiamiesiems priebalsiams, nebuvo pastebėta.

Trečia, metodas geriau tinka nekirčiuotiems garsams. Kirčiuotiems garsams reikėtų taikyti papildomą algoritmą amplitudžių šuoliams sumažinti. Taikant šį metodą, paprastai vieni garsai nevirsta kitais, tačiau palyginti su pakeitimų metodu spragsėjimų ir kitokio triukšmo sintezuojame signale padaugėja.

Taigi, jei difonams, kurių kairė pusė yra balsis arba dvigarsis (išskyrus „ie“, „uo“ ir „r“), o dešinė pusė – visos aibių B*, P* ir Z* fonemos, naudotume garsų ištempimo metodą, tai sąrašas liktų 2572 difonai, kurie padengia 77,50 proc. atvejų, o 1496 difonai, kurie padengia 22,40 proc. atvejų, būtų sintezuojami naudojant difonų ištempimą. Taip pat būtų naudojami 4396 pakeitimai, kurie padengia likusius 0,10 proc. atvejų (4 lentelė).

Išvados

Sudarant difonų bazę pagal difonų vartosenos dažnius, o ne pagal kalbos fonotaktikos taisykles, galima sumažinti reikiamų difonų skaičių iki 70,7 proc. originalios bazės dydžio, o nerastų difonų padaugės tikrai 0,015 proc. (nuo 0,085 proc. iki 0,1 proc.).

Be to, difonų bazę galima sumažinti, remiantis garsų panašumu. Iš atliktų audicinių eksperimentų galima daryti tokias išvadas:

- Difonus, kurių antrasis dėmuo yra /p/, /t/, /k/, /p'/, /t'/, /k'/, /ts/, /tS/, /ts'/, /tS'/ galima laisvai keisti vienus kitais, neatsižvelgiant į kietumą/minkštumą.
- Difonus, kurių antrasis dėmuo /dz/, /dZ/, /dz'/, /dZ'/, beveik visada galima keisti difonais, kurių antrasis dėmuo /b/, /d/, /g/, /b'/, /d'/, /g'/ . Natūralus skambėjimas gautas visais atvejais po balsio ar dvibalsio ir 99,0 proc atvejų po mišriojo dvigarsio.
- Difonus, kurių antrasis dėmuo /b/, /d/, /g/, /b'/, /d'/, /g'/, taip pat dažniausiai galima keiti vienus kitais. Natūraliai skamba 85,6 proc. atvejų po balsio, 96,5 proc. atvejų po dvibalsio ir 92,7 proc. atvejų po mišriojo dvigarsio.

Kai trūkstamus difonus bandoma užpildyti ištempiant greitimus garsus, buvo pastebėta, kad:

- Kintančių garsų „ie“, „uo“ ir „r“ ištempti negalima.
- Likusius garsus dažniausiai galima ištempti – natūraliai skamba 95,1 proc. atvejų.

Pijus Kasparaitis, Žilvinas Ledas

Optimization of Lithuanian Diphone Databases

Summary

Creation and optimization of the Lithuanian diphone inventory used for concatenative text-to-speech synthesis is studied in this paper. Creation of diphone database starts with compilation of a list of diphones. This is not a trivial problem, because some diphones are not valid. If valid diphones are deduced using language phonotactics rules, some diphones which needed to synthesize foreign words are omitted. Besides, a lot of practically unused diphones are added to the list. Statistical diphone usage analysis was performed in this work. The results of this analysis imply that using statistically motivated diphone inventory pruning one can compile a much smaller inventory while keeping a very high text coverage.

Diphone inventory pruning using phone similarity (one diphone is substituted with another that sounds similarly) and using phone stretching (a missing diphone is synthesized by stretching phones of adjacent diphones) is described, as well. Listening experiments with diphones that contain a vowel or diphthong followed by stop consonant (or fricative, for phone stretching) were executed. Groups of diphones were identified for which synthesized speech quality is not degraded (and for which quality is degraded only marginally) when using two described methods. In addition, potential diphone inventory reduction was estimated.

Straipsnis įteiktas 2010 11
Parengtas publikuoti 2011 11

Apie autorius

Pijus Kasparaitis, fizinių mokslų daktaras, Vilniaus universiteto Matematikos ir informatikos fakulteto Kompiuterijos katedros docentas. Per pastaruosius 5 metus paskelbė 7 publikacijas, penkias iš jų recenzuojamuose žurnaluose.

Mokslinių tyrimų sritis apima balso sintezę iš teksto ir kitas kompiuterinės lingvistikos sritis.

Adresas: Vilniaus universitetas Matematikos ir informatikos fakultetas Kompiuterijos katedra, Naugarduko g. 24, 03225 Vilnius, Lietuva.

El. paštas: pkasparaitis@yahoo.com

Žilvinas Ledas, dirba Vilniaus universiteto Matematikos ir informatikos fakulteto Programų sistemų katedros asistentu.

Mokslinių tyrimų sritis apima balso sintezę iš teksto.

Adresas: Vilniaus universitetas Matematikos ir informatikos fakultetas Programų sistemų katedra, Naugarduko g. 24, 03225 Vilnius, Lietuva.

El. paštas: Zilvinas.Ledas@dict.lt

Taikant difonų pakeitimo metodą, galima difonų skaičių sumažinti iki 63,4 proc. originalios bazės dydžio nepabloginant sintezuoto garso kokybės, ir sumažinti iki 57,1 proc. nežymiai pabloginant kokybę.

Pritaikius garsų ištempimo metodą, difonų skaičių galima sumažinti iki 51,4 proc. originalios bazės dydžio, tačiau sintezuotame garse gali atsirasti papildomų triukšmų.

Literatūra

1. Anbinderis, T., Kasparaitis, P., 2009. Lietuvių kalbos homografų vienareikšminimas remiantis leksemų ir morfologinių pažymų vartosenos dažniais. *Kalbų studijos/ Studies about Languages*, 14. Kaunas: Technologija, pp.25–31.
2. Black, A. W., Taylor, P., 1997. Automatically Clustering Similar Units for Unit Selection in Speech Synthesis. *Eurospeech 97*, 2, Rhodes, Greece, pp.601–604.
3. Karosienė, V., Girdenis, A., 1993. Bendrinės kalbos fonemų dažnumai. *Kalbotyra*, 42(1). Vilnius: Vilniaus universiteto leidykla, pp. 28–38
4. Karosienė, V., Girdenis, A., 1994. Lietuvių bendrinės kalbos skiems tipų dažnumai. *Kalbotyra*, 43(1). Vilnius: Vilniaus universiteto leidykla, pp.34–42.
5. Kasparaitis, P., 1999. Transcribing of the Lithuanian Text Using Formal Rules. *Informatika*, 10(4). Vilnius: VGTU Leidykla, pp.367–376.
6. Kasparaitis, P., 2000. Automatic Stressing of the Lithuanian Text on the Basis of a Dictionary. *Informatika*, 11(1). Vilnius: VGTU Leidykla, pp.19–40.
7. Kasparaitis, P., 2005. Diphone Databases for Lithuanian Text-to-Speech Synthesis. *Informatika*, 16(2). Vilnius: VGTU Leidykla, pp.193–202.
8. Kazlauskienė, A., Raškinis, G., 2008. Lietuvių kalbos fonologinio skiemens struktūrinių modelių dažnumas. *Žmogus ir žodis*, 10(1). Šiauliai: Šiaulių universiteto leidykla, pp.24–31.
9. Kazlauskienė, A., Raškinis, G., 2009. Bendrinės lietuvių kalbos garsų dažnumas. *Respectus Philologicus*, 16(21). Vilnius: VPU leidykla, pp.169–182.
10. Taylor, P., 2009. *Text-to-Speech Synthesis*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511816338>