VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
KOMPIUTERIJOS KATEDRA

Baigiamasis magistro darbas

Lietuvių kalbos balso sintezė vienetų parinkimo metodu

Atliko:
Kipras Kančys parašas

Vadovas:
dr. Pijus Kasparaitis

Vilnius
2017

Turinys

Santrauka

Šis darbas yra skirtas išbandyti iki šiol lietuvių kalbai sintezuoti nenaudotą vienetų parinkimo metodą su difonais. Ankstesni lietuviškai kalbantys balso sintezatoriai veikė kitais metodais, o vėlyviausias sintezatorius veikė vienetų parinkimo metodu su fonemomis. Tuo tarpu užsienio kalboms sintezuoti dažnai naudojamas tas pats metodas su difonais. Būtent tai ir buvo įgyvendinta šiame darbe. Tyrimo metu pastebėta, kad turėtas garso įrašas yra nepakankamas difoniniam sintezatoriui, o algoritmo sudėtingumas leidžia plėsti garsų bazę, nežymiai pailginant sintezatoriaus veikimo laiką. Galutiniame žmonių grupės teste lyginant foneminį ir difoninį sintezatorius truputėlį geresniu buvo pripažintas difoninis sintezatorius.

Summary

Lithuanian Text–to–Speech Synthesis Based on Unit Selection Method

The aim of this thesis is to test a new method for Lithuanian speech synthesis. Previous synthesizers that were based on unit selection method used phonemes as basic synthesis units. In contrast, this thesis will describe usage of unit selection method based on diphones. During the construction of diphones database it was found out that database used for phonemes is not big enough for diphones. Not only some of diphones were missing but also there was not enough copies of needed diphones from different parts of word or sentence. It was also clear that much large database could be used for diphones with low impact on synthesizer execution time. Even though at the final human listening test diphones synthesizer were a bit better then phonemes synthesizer.

Įvadas

Darbas yra skirtas lietuvių kalbos balso sintezatoriaus tobulinimui. Tokio sintezatoriaus reikalingumą puikiai atskleidžia lietuvių šneka valdomų paslaugų projektas *Liepa* [13], kuris skatina lietuvių kalbos naudojimo plėtojimą bendraujant su kompiuteriu. Vystomi įvairūs lietuvių kalbos panaudojimo projektai. Pavyzdžiui, lietuviškų žodžių tartuvas [10] (nežinantiems, kaip tarti lietuviškus naujadarus), naujienų portalų tekstų skaitymas lietuviškai. Nors projekto *Liepa* pagrindinė auditorija yra aklieji, tai galėtų tikti ir senyvo amžiaus žmonėms, kurių akys pavargusios. Jie galėtų klausytis tekstų, naujienų ar net knygų. Tuo taip pat galėtų pasinaudoti ir visi kiti, mėgstantys klausytis knygų ar kitų tekstų, kol dirba ar keliauja į darbą. Taip pat verta paminėti kitą *Liepos* vystomą projektą apie kompiuterių valdymą balsu [20], lietuvių kalba. Plėtojant šiuos projektus galima prisidėti prie neįgaliųjų ir senyvo amžiaus žmonių laisvalaikio gerinimo.

Šiuo metu veikiantys lietuvių kalbos sintezatoriai naudoja vienetų parinkimo metodą su fonemomis. Tai yra garso signalo surinkimas iš vientisų garsų – fonemų (garsai imami nuo garso pradžios iki garso pabaigos). Garso signalas sintezuojamas iš fonemų jas sujungiant į žodžius ir sakinius. Sujungus dvi fonemas iš skirtingų konteksų, jų jungtis gali nuskambėti neįprastai. Huang ir kiti [3] pastebi, kad kur kas natūraliau skamba balsas, kuris sudarytas jungiant difonus, t. y. garsų nuo vieno garso vidurio iki kito garso vidurio. Garsai kur kas labiau varijuoja, kinta kraštuose, nei per vidurį. Būtent šia idėja remiantis yra kuriami difoniniai sintezatoriai. Šio tyrimo metu ir buvo sukurtas difoninis sintezatorius, o vėliau ir palygintas su foneminiu analogu.

Pirmo semestro metu buvo paruošta difonų bazė. Buvo sprendžiama lietuviškų garsų dalinimo per pusę problema. Duslūs ir skardūs lietuviški priebalsiai reikalauja skirtingų metodų jų vidurio taško nustatymui. Buvo ieškomi garso bangų periodai, intensyvumo minimumai ir signalo taškai, kuriuose jie lygūs nuliui. Semestro eigoje taip pat buvo sukurtas sintezės vienetų parinkimo algoritmas. Rudens semestro metu algoritmas buvo integruotas į *Liepos* projekto aplikaciją, tai leido išgirsti algoritmo veikimą. Vėliau sintezatorius buvo tobulinamas pridedant įvairius balso sintezės aspektus.

1. Balso sintezatoriai

Balso sintezatoriai – tai įrenginiai, sugebantys paduotą tekstą perskaityti balsu. Šią technologiją įprasta vadinti tekstu-į-kalbą (angl. text-to-speach, TTS). Nepaisant to, kad mūsų kasdieniniame gyvenime klausytis kalbančių įrenginių yra gana įprasta, visgi reiktų atskirti žmogaus balso įrašą nuo žmogaus balso sintezavimo. Pavyzdžiui, GPS tariamos komandos ar autobuso stotelių pavadinimų pranešimai tėra žmogaus balso įrašai. Pagal apibrėžimą balso sintezatorius nuo balso įrašo skiriasi tuo, kad balso sintezatorius geba ištarti net ir tokius žodžius, kurie dar nė sykio nebuvo tarti. Puikūs pavyzdžiai būtų [1] *Apple*, *Amazon* ir *Microsoft* sukurti virtualūs asistentai *Siri*, *Amazon Echo* ir *Cortana*. Taip pat *Google* vertėjo balso sintezatorius. Nei vienas iš šių sintezatorių bent jau kol kas lietuviškai nekalba, tačiau egzistuoja ir lietuviškų sintezatorių, jie bus detaliau apžvelgti toliau šiame darbe. Visgi reikia pastebėti, kad balso sintezatorių ateitis šviesi, keturios didžiulės kompanijos *Apple*, *Google*, *Amazon* ir *Microsoft* turi savo balso sintezatorius ir daugelis kitų kompanijų juos plėtoja. Neabejotinai ateityje vis daugiau ir daugiau operacijų bus galima atlikti balsu.

Neregiai ir silpnaregiai bendrauti su kompiuteriu naudoja Brailio lenteles. Tai įrenginys 1 pav., kurio pagalba Brailio raštu galima perskaityti kompiuterio grąžinamą informaciją.

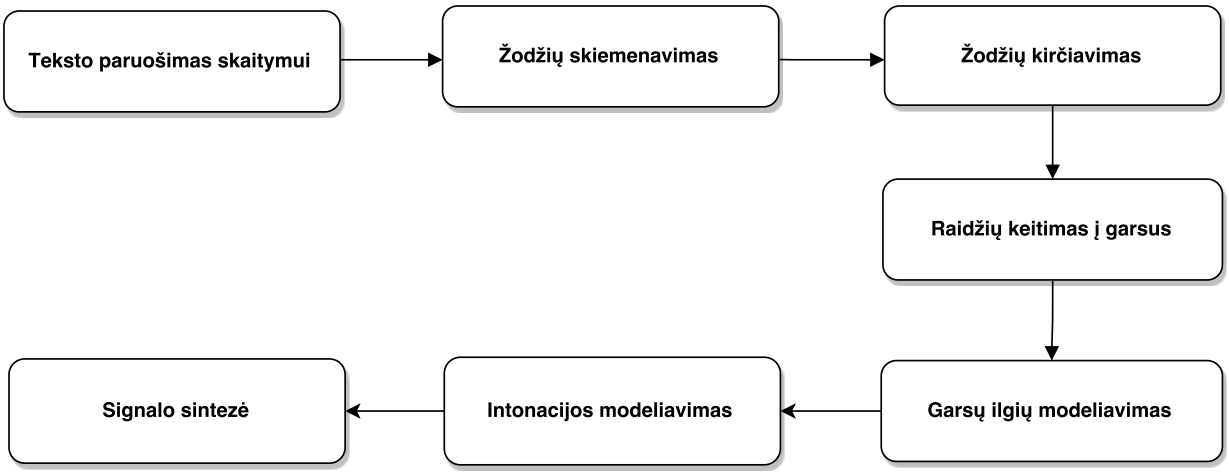
1 pav. Brailio skaitytuvas, iliustracija iš Boundless Assistive Technology tinklalapio

Alternatyva Brailio lentelėms yra kompiuteriniai balso sintezatoriai, kurie kompiuterio pateikiamą informaciją perskaito balsu. Kompiuteriai su *Windows* operacine sistema nuo seno turėjo sintezatorius, o šių dienų visos operacinės sistemos turi vienokį ar kitokį balso sintezatorių, taip pat juos galima parsisiųsti iš interneto. Sintezatoriai leidžia kompiuteriu naudotis akliesiems.

Artimiausiu metu sintezatoriai taps prieinami ir kitai suinteresuotai grupei – senjorams ir vyresnio amžiaus žmonėms, kuriems skaityti yra sunku ar nepatogu. Jau šiandien egzistuojantys interneto naršyklių plėtiniai leidžia pažymėtą tekstą skaityti ne tik anglišku, bet ir lietuvišku balsu. Tobulėjant žmogaus balso atpažinimo sistemoms, netrukus bus galima nuo kompiuterio įjungimo iki jo išjungimo su juo bendrauti vien balsu. Tad balso sintezatorių panaudojimas ir potencialas ateityje tik didės.

1.1. Balso sintezavimas

Sukurti kokybišką balso sintezatorių yra sudėtinga. Apskritai reikia apibrėžti, ką reiškia pasakymas – kokybiškas balso sintezatorius. Literatūroje aptinkamas terminas - tikras balsas (angl. true voice), t. y. balsas, kurį sunku atskirti nuo natūralaus žmogaus balso. Tačiau galima ir kitaip interpretuoti kokybišką balso sintezatorių. Tai turėtų būti toks sintezatorius, kurį pilnai suprastų kiekvienas žmogus. Kad sintezatorius skambėtų žmogiškai reikia atlikti daug teksto paruošimo ir sintezavimo užduočių, kurios matyti 2 pav.



2 pav. Balso sintezavimo užduotys

Teksto paruošimas skaitymui. Skaityti tekstą atrodo paprasta, tačiau tai yra kur kas sudėtingesnis uždavinys, nei gali pasirodyti. Tuo galima įsitikinti stebinti vaikus besimokančius skaityti. Taip yra todėl, kad dažniausiai parašytas tekstas gali būti dviprasmis ir tik naudojantis kontekstu galima suprasti tikrąją žodžių prasmę. Įvairūs skaičiai pasitaikę tekste kartais turi būti interpretuojami kaip datos, o kartais tiesiog kaip skaičiai. Įvairūs specialūs simboliai ir trumpiniai taip pat turi būti teisingai suprantami. Pavyzdžiui, data 2016 01 01 – turėtų būti skaitoma ne du tūkstančiai šešiolika, nulis vienas, nulis vienas, o du tūkstančiai šešioliktų metų sausio pirma diena.

Žodžių skiemenavimas. Žodžius reikia skiemenuoti, nes netaisyklingai ir neteisingose vietose suskiemenuotas žodis gali skambėti dirbtinai ar net netekti prasmės.

Žodžių kirčiavimas. Yra kalbų, kaip lenkų ar latvių, kuriose visada kirčiuojamas tas pats skiemuo, atitinkamai priešpaskutinis arba pirmas. Tačiau lietuvių kalboje, kaip ir daugelyje kitų, kirčio vieta žodyje kinta. Neteisingai sukirčiavus, balsas skambės ne tik nenatūraliai, bet ir daugeliu atvejų bus sunkiai suprantamas. Kai kurių žodžių prasmė priklauso nuo kirčio vietos, pavyzdžiui, br**ã**škės ar braškė**̃**s, paš**á**ukite ar pašau**̃**kite, **ì**škyla ar iškỹla.

Raidžių keitimas į garsus. Galiausiai reikia raides pakeisti į garsus. Lietuvių kalboje dažniausiai tariama taip, kaip rašoma, tačiau yra ir išimčių. Štai kelios taisyklės:

- Duslūs garsai *p, t, k, s, š* prieš skardžius garsus b, d, g, z, ž suskardėja. Pvz., žai[z]damas, nors žai**s**ti.
- Skardūs garsai *b, d, g, z, ž* žodžio gale ir prieš duslius garsus *p, t, k, s, š* suduslėja. Pvz., dir[p]ti, nors dir**b**a.
- Raidžių junginys *ia* dažnai tariamas kaip *e*.

Taip pat lietuvių kalboje turima ilgųjų ir trumpųjų balsių. Kai kurie iš jų žymimi tomis pačiomis raidėmis. Pavyzdžiui, trumpas e – nešu ir ilgas e – neša, trumpas a – darau ir ilgas a daro, trumpas o – tomas ir ilgas o – generolas. Šiems garsams turima po kelias fonemas, kurias reikia atitinkamai priskirti.

Garsų ilgių modeliavimas. Garsų ilgiai priklauso nuo daugelio faktorių,. Pavyzdžiui, garso vietos žodyje ar sakinyje.

Intonacijos modeliavimas. Žmonėms yra įprasta sakinius pradėti aukštesniu ir užbaigti žemesniu tonu. Klausiamuosiuose sakiniuose – priešingai, tonas kyla sakinio gale. Sintezatorius taip pat turėtų laikytis šių dėsningumų.

Aukščiau išvardyti iššūkiai šiame darbe nebuvo sprendžiami, buvo naudojamos jau turimos teksto paruošimo bibliotekos. Šio tyrimo metu buvo dirbama tik su signalo sinteze.

Signalo sintezė. Signalo sintezavimas tai galutinis balso sintezės uždavinys. Yra daug metodų naudojamų signalo sintezei, tačiau šiame darbe buvo taikomas vienetų parinkimo metodas.

1.2. Lietuviški balso sintezatoriai

Per kiek daugiau nei 20 metų buvo sukurti septyni lietuviško balso sintezatoriai. Jie naudoja įvairiausius sintezės metodus [7]. Plačiau apie kiekvieną iš sintezatorių:

- Appollo – pirmasis lietuviškai kalbantis sintezatorius 1994 m. sukurtas Didžiojoje Britanijoje. Priešingai nei kiti, jis buvo pamėgtas Lietuvos aklųjų bendruomenės. Balsas sintezuojamas formantinės sintezės metodu t. y. garso signalas yra generuojamas, tad tai pilnai dirbtinio garso sintezatorius.

- Aistis – pirmasis lietuvių sukurtas sintezatorius 1996 m. Šis ir kiti vėlesni sintezatoriai naudojo konkatenacinį metodą. Šio metodo esmė garso signalą suklijuoti iš mažesnių signalo dalių. Garsų bazę sudarė 480 įvairaus ilgio dalių, kurios buvo iškarpytos iš diktoriaus balso įrašų.

- Gintaras – čekų „RosaSOFT“ ir Pijaus Kasparaičio kurtas ir tobulintas sintezatorius. Balsas sintezuojamas taip pat konkatenaciniu metodu. Naudota 1500 įvairaus ilgio kalbos vienetų, kurių tonai ir ilgiai būdavo atitinkamai modifikuojami.

- Aistis 2 – balsas sintezuojamas konkatenaciniu metodu naudojant 5003 difonus, ilgius ir tonus papildomai modifikuojant.

- Egidijus – dar sykį naudotas konkatenacinis sintezės metodas su 6500 kontekstinių fonemų. Ypatingas dėmesys buvo skiriamas kirčiuotiems skiemenims išskiriant fonemas einančians prieš ir po kirčio. Taip pat visi dvibalsiai ir mišrieji dvigarsiai buvo laikomi savarankiškomis fonemomis. Garsų modifikavimo ir jungimo metodas viešai neskelbiamas.

- SINT.AS – balso sintezei naudotas vienetų parinkimo metodas su fonemomis. Sudarytos dviejų diktorių ištisinių įrašų bazės su 2 tūkst. sakinių, kuriuos sudaro apie 77 tūkst. garsų.

- Projektas LIEPA – Vykdant projektą buvo sukurtas keturių balsų sintezatorius. Sintezei naudojamas taip pat vienetų parinkimo algoritmas su fonemomis. Garsų bazę sudarė 5 000 sakinių, kuriuose – daugiau, kaip 161 tūkst. garsų.

1.3. Sintezatorių palyginimas

Turint ne vieną, o kelis balso sintezatorius, reikia mokėti juos palyginti. Schmidt–Nielsen [16] išskiria du pagrindinius sintetinio balso kokybės parametrus: objektyvų suprantamumą (angl. intelligibility) ir subjektyvų priimtinumą (angl. acceptability). Suprantamumas apibūdina, kiek teksto klausytojai sugebėjo suprasti. Pavyzdžiui, skaitytojas suprato 76% visų tekste buvusių žodžių. Atliekamas testas, kurio metu prašoma užrašyti ką tik išgirstą sakinį. Skaičiuojami teisingai parašyti žodžiai.

Priimtinumas yra skirtas nustatyti, ar sintetinis balsas yra malonus klausytojui, taip pat stengiamasi patikrinti ar klausytojas jį atskiria nuo naturalaus žmogaus balso. Šiam parametrui įvertinti galimi du testai. Pirmasis reikalaujau, kad klausytojas išgirdęs sintezatorių jį įvertintų penkiabalėje sistemoje. Kito testo metu dvejais sintezatoriais yra perskaitomas tas pats sakinys ir prašoma pasirinkti geriau nuskambėjusį sintezatorių. Taip pat galima lyginti sintezatorių su diktoriaus balsu. Šis testas buvo atliktas tyrimo metu, apie testo rezultatus galima skaityti 5.3 skyrelyje.

Kai pasirodė pirmieji sintezatoriai, ypatingai buvo stengiamasi gerinti suprantamumą, nes iš nesuprantamo sintezatoriaus nėra jokios naudos. Šiais laikais sintezatoriai jau yra pasiekę aukštą suprantamumo lygmenį, tad dabar tobulinamas priimtimumas [7].

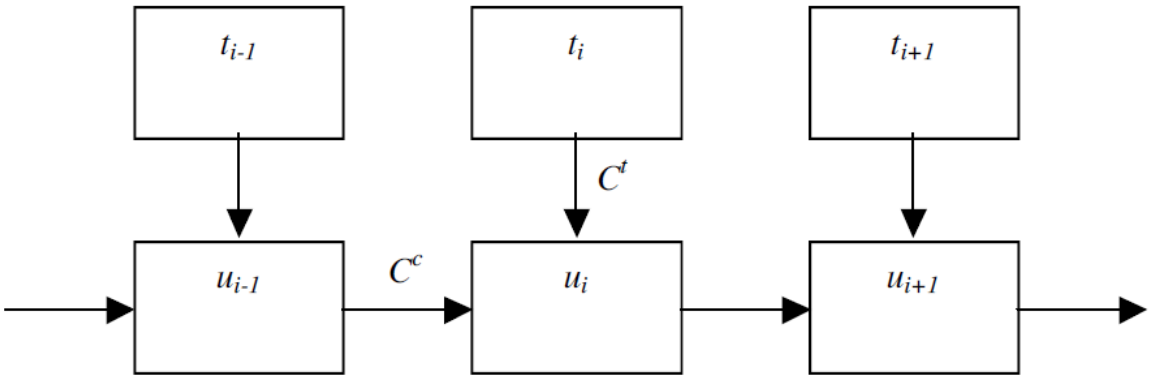
Atvejais, kai sintezatoriai yra sukurti naudojant tuos pačius balso įrašus arba panašius metodus, galima rasti būdų, kaip jų kokybę įvertinti statistiškai. Plačiau apie sintezatorių palyginimą 5 skyriuje.

2. Vienetų parinkimo metodas

Galima išskirti dvi pagrindines balso sintezavimo metodų grupes: formantinius ir konkatenacinius metodus. Pirmieji garso signalą generuoja dirbtinai, antrieji garso singalą sujungia iš žmogaus ištarto teksto dalių. Didžioji dauguma šiuolaikinių sintezatorių veikia konkatenaciniu metodu, tačiau *Google* dirbtinio intelekto labaratorijoje *DeepMind* neseniai buvo sukurtas sintezatorius formatiniu metodu. [11] teigiama, kad mašininių algoritmų pagalba generuojamo balso sintezatorius kokybe lenkia konkenacinius sintezatorius. Visgi šiai dienai dauguma sintezatorių veikia konkatenaciniais metodais, vienas jų vienetų parinkimo metodas (angl. unit selection).

Vienetų parinkimo metodu sintezuojamas balsas naudojantis diktorius balso įrašu. Šie garso įrašai būna apdorojami, skaidomi į smulkesnes dalis – sintezės vienetus. Taip gaunamas garsų rinkinys (angl. corpus). Kalbos sintezavimas tampa tarsi dėlionė, kurioje norimą žodį galima sudėti iš esamų detalių.

Remiantis [15], standartiniame vienetų parinkimo metode siekiama norimą (angl. target) užklausą $t^n = (t_1, ..., t_n)$ pakeisti vienetų seka $u^n = (u_1, ..., u_n)$, kurios vienetai u_i yra paimami iš garsų bazės. Klasikinis vienetų parinkimo metodas remiasi dviem kainų funkcijomis. Keitimo kainos funkcija (angl. target cost) $C^t(u_i, t_i)$, t. y. skirtumas tarp u_i – bazėje turimo ir t_i užklausoje esamo garso. Jungimo kainos funkcija (angl. concatenation cost) $C^c(u_{i-1}, u_i)$, kuri yra dviejų garsų (u_{i-1} ir u_i) jungimo kaina.



3 pav. Vienetų parinkimo metodo schema [15]

Reikia pastebėti, kad tyrime buvo naudota nestandartinė vienetų parinkimo metodo implementacija, kuri taip pat buvo naudota Pijaus Kasparaičio foneminiame sintezatoriuje, plačiau apie tai 2.4 skyrelyje.

[3] išskiriami šie pagrindiniai klausimai, kuriuos reikia atsakyti prieš pradedant kalbos sintezatoriaus konstravimą vienetų parinkimo metodu:

- Kaip pasirinkti dėlionės smulkiausią elementą?
- Kaip pasirinkti pradinį tekstą? Koks turėtų būti jo ilgis, kokie sakiniai ten turėtų atsidurti?
- Kaip rasti natūraliausiai skambančią garsų kombinaciją?

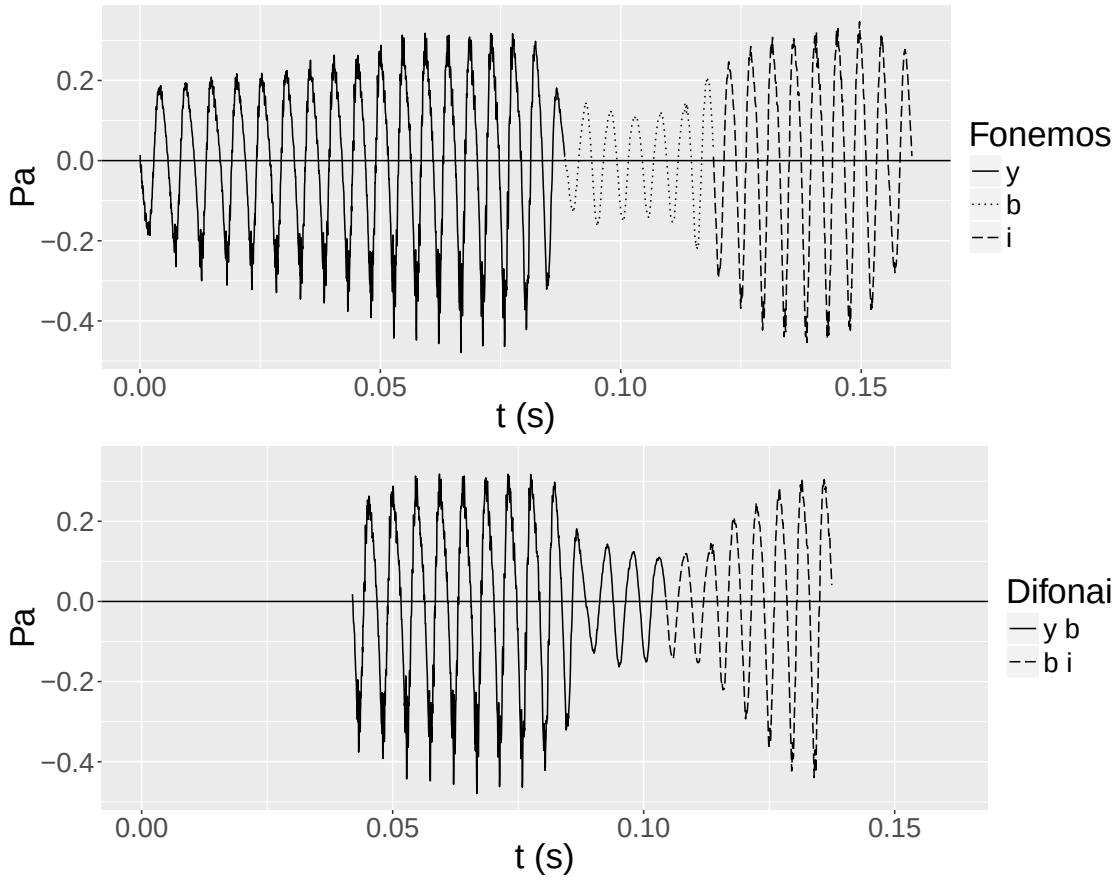
2.1. Bazinis elementas

Prieš konstruojant sintezatorių reikia apsispręsti dėl bazinio vieneto t. y. smulkiausio elemento (angl. base element, unit). Sudedant šiuos vienetus gaunami žodžiai ir sakiniai. Tarkime, jeigu

garsų bazėje egzistuoja absoliučiai visi žodžiai, galima surinkti, bet kokį tekstą. Tokiu atveju bazinis elementas būtų žodis. Tačiau tam reikėtų labai didelės įrašų bazės. Visgi turint duomenų bazę iš visų žodžių, sintezuojama šneka neskambėtų natūraliai, nes žodžių vieta sakinyje taip pat svarbi. Jei į sintezuojamo sakinio pradžią įdėdamas žodis iš sakinio galo, tai intonaciškai sakinys skambės keistai. Tad turėti pakankamą garsų bazę praktiškai nėra įmanoma, todėl įprastai naudojami smulkesni sintezės vienetai. Paul Taylor [19] išskiria daug galimų smulkiausio elemento tipų. Verta paminėti kelis:

- pusė fonemos – imama fonema nuo pradžios iki jos vidurio arba nuo vidurio iki pabaigos.
- fonema – imamas pilnas garsas nuo jo pradžios iki galo.
- difonas – imamas garsas nuo vienos fonemos vidurio iki kitos fonemos vidurio.

Yra ir kitų galimų variantų: pusė skiemens, pilnas skiemuo, žodžis ar net jų junginiai. Kiekvienas variantas turi savo pliusų ir minusų. Kuo didesnis vienetas, tuo didesnės duomenų bazės reikės. Kuo smulkesnis vienetas – tuo daugiau jungimo vietų sintezuojamame sakinyje, tuo keisčiau skambės kalba. Sintezatoriaus kokybė priklauso nuo kalbos, kuria sintezuojama, turimo įrašo ir sintezės vieneto. Mark Beutnagel ir kiti [2] savo tyrime lygino fonemų ir difonų sintezatorius ir nustatė, kad difoninis sintezatorius anglų kalbai skamba geriau nei foneminis. Tuo tarpu Kishore su kolegomis [8] Hindi kalbai rekomenduoja naudoti pusės fonemos vienetą, nors išbandė ir pilnos fonemos ir difono bei skiemens sintezatorius. Ankstesni lietuvių kalbos sintezatoriai vienetų parinkimo metodu naudojo fonemas. Šiame darbe buvo lyginama lietuvių kalbos sintezė iš fonemų su sinteze iš difonų.



4 pav. Fonemų ir difonų jungimo schemos

Huang ir kiti [3] yra pastebėję, kad difonų pagrindu kurti sintezatoriai skamba kur kas natūraliau, nei fonemų pagrindu pagaminti sintezatoriai. Pavyzdys paaiškina, kodėl taip yra. 4 pav. matyti žodžio *stat**ybi**ninkai* fragmentas. Viršutiniame grafike pavaizduotas garso signalas suskaidytas į fonemas /y/, /b/ ir /i/. Apatiniame į difonus /y b/ ir /b i/. Fonemų grafikuose matyti, kad fonema /y/ turės galiuką fonemos /b/, nes yra laikomasi garsų karpymo taisyklės – garso signalas gali būti kerpamas tik ten, kur jis yra lygus nuliui (plačiau apie garsų karpymo taisykles 3 skyriuje). Panašiai nutinka ir kitoje fonemos /b/ pusėje su fonema /i/. Tuo tarpu, kai garso signalą kerpamas pagal difonus, kirpimo taškai yra garso viduryje, kuriame nėra nieko kito tik pats garsas.

2.2. Pradinis garsų rinkinys

Imamas pasirinktas balso įrašas ir skaidomas į smulkesnius sintezės vienetus tarkime fonemas ar difonus. Sintezuojant tekstą ieškoma sintezės vienetų, kurie paeiliui padengtų norimus žodžius. Jei nerandami vienetai einantys paeiliui imami kiti vienetai, tačiau ir vėl siekiama, kad šie vėl eitų vienas paskui kitą. Taip galiausiai surenkami žodžiai ir sakiniai iš sintezės vienetų grupių. Kuo ilgesnis balso įrašas naudojamas, tuo daugiau žodžių ar sakinio fragmentų turima, tad nebereikia dėlioti žodžių iš garsų ar skiemenų. Tokiu atveju balsas skambės natūraliai. Kiekvienas žodis sudėtas iš garsų, kuriuos algoritmas parinks sudėti žodžiui, neskambės taip natūraliai, kaip ištartas diktoriaus. Norint išvengti žodžių sudarymo iš paskirų vienetų, reikia turėti didžiulę įrašų bazę. Augant įrašų bazei, ilgiau trunka garsų parinkimo algoritmai, tad reikia išlaikyti pusiausvyrą tarp sintezatoriaus kokybės, duomenų kiekio ir algoritmų spartos.

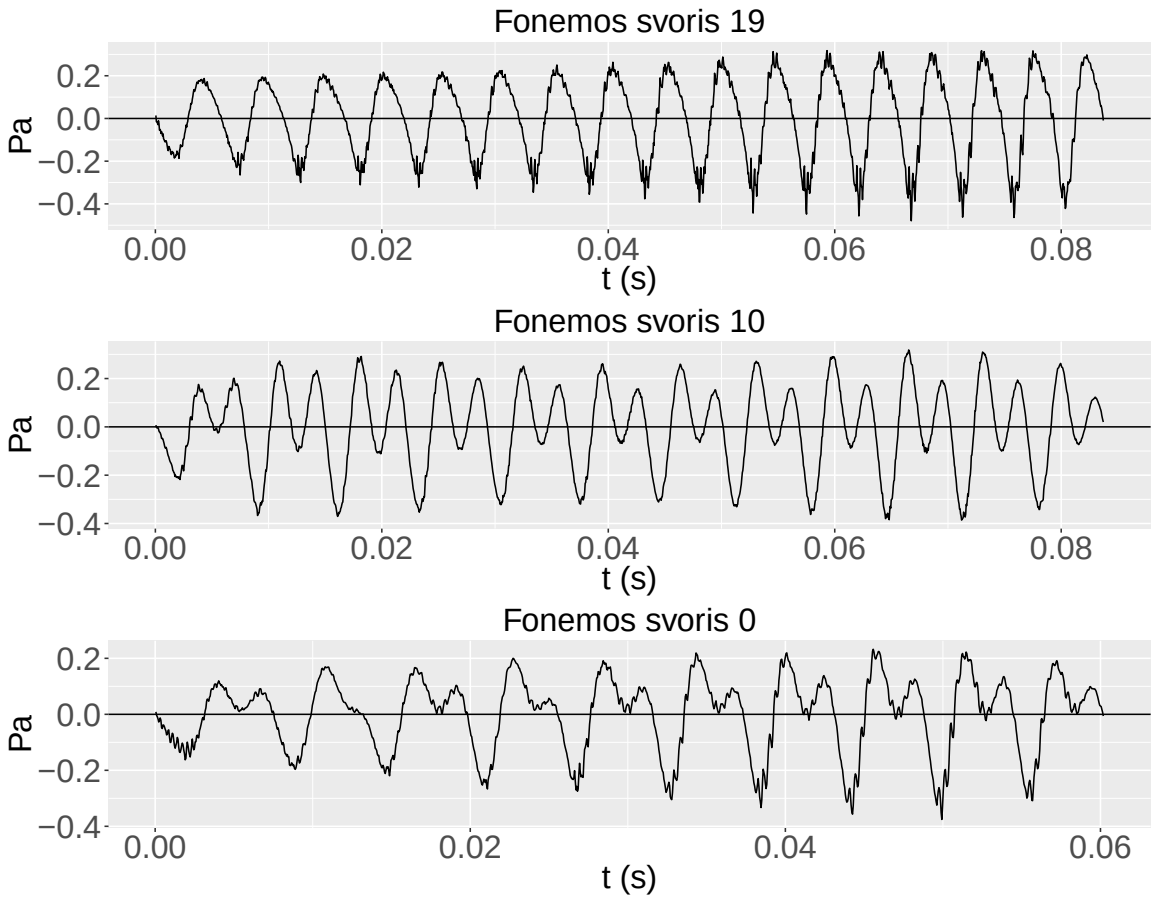
Sudarant garso įrašą reikia stengtis, kad jame būtų pakankamas kiekis dažnai kalboje pasitaikančių sintezės vienetų egzempliorių, bet tuo pačiu jame rastųsi pakankamas kiekis retesnių sintezės vienetų [2]. Sakinių, kuriuos reiktų įdėti į tekstą, parinkimas kiekvienai kalbai yra atskira užduotis. Lietuvių kalbos teksto ilgio ir sakinių parinkimo uždavinį sprendė dr. Pijus Kasparaitis [6]. Tyrimo metu buvo nuspręsta balso sintezatoriams naudoti trigarsių kombinacijas, kurios kuo geriau padengtų lietuvių kalbos pasitaikančius žodžius. Buvo parinkta apie 5 000 sakinių su 162 000 fonemų (maždaug 3 valandos skaitomo teksto). Būtent šis rinkinys buvo naudojamas darbe. Palyginimui, anglų kalbos Siri balso autorė Susan Bennett CNN interviu yra išsidavusi, kad balso įrašus įrašinėjo visą mėnesį po 4 valandas per dieną [14]. Tai būtų apytiksliai 80 valandų balso įrašų, skaičiuojant tik darbo dienas. Sintezatorių testavimui buvo naudojami 20 000 sakinių, kurie nebuvo naudoti garsų bazėje.

2.3. Garsų svoriai

Reikia pastebėti, kad tuos pačius garsus galima ištarti skirtingai. Tai daug kuo priklauso nuo intonacijos, tembro bei garsų, kurie eina prieš ir po tariamo garso. Tad natūralu, kad nagrinėjant fonemų bazę galima rasti atvejus, kai to paties garso signalai gerokai skiriasi. 5 pav. matyti fonemos /Ii/ (plačiau apie fonemų žymėjimą [5]) trys skirtingi signalai. Kaip matoma, skiriasi ne tik signalų periodai, trajektorija, bet ir jų ilgiai. Abu viršutiniai signalai trunka 0.08s tuo tarpu apatinio signalo ilgis 0.06s.

Taigi kuo labiau fonemos signalas skiriasi nuo kitų tos pačios fonemos signalų, tuo didesnis svoris jai turėtų būti priskiriamas. Reiškia sintezuojant bus vengiama naudoti fonemas, turinčias didesnius svorius, nes jos yra nutolusios nuo įprasto tos fonemos standarto.

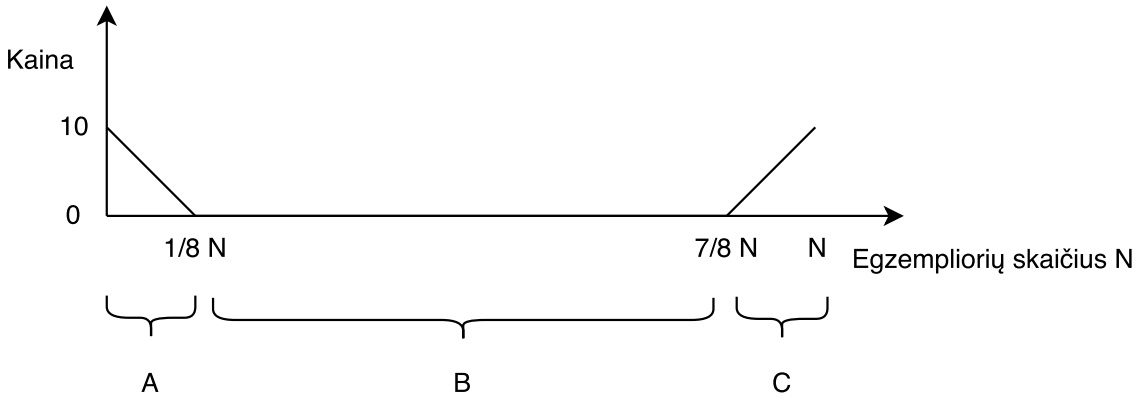
Garso signalą apibūdina šios charakteristikos: garso spektras, tonas, energija ir ilgis. P. Kasparaitis yra pasiūlęs kiekvienam garsui priskirti svorį įvertinant šias charakteristikas. Kiekvienai



5 pav. Fonemos |Ii| trys skirtingi signalai su jų svoriais

fonemai buvo skaičiuojami keturi parametrai: ilgis, energija (vidutinis standartinis nuokrypis), pagrindinis tonas ir spektras (14 Bark spektro koeficientų).

Tada visi tam tikros fonemos egzemplioriai (pvz., /aA/) buvo surūšiuojami pagal vieną iš parametrų. Tarkime, kad yra N egzempliorių. Tada fonemų svorių priskyrimas atrodytų, kaip pavaizduoda 6 pav.



6 pav. Fonemų svorių priskyrimo schema

Egzemplioriams, kurių numeriai surūšiuotame sąraše yra:

- A grupėje priskiriami svoriai, kurie tiesiškai mažėja nuo 10 iki 0;
- B grupėje priskiriam nuliniai svoriai;

- C grupėje priskiriami svoriai, kurie tiesiškai didėja nuo 0 iki 10;

Taip kiekvienai fonemai yra gaunami keturi svoriai, kurių reikšmės nuo 0 iki 10. Iš jų gaunam vieną reikšmę (nuo 0 iki 40) pagal formulę:

$$Svoris = 0.6 \times ilgioSvoris + 0.8 \times energijosSvoris + 1.2 \times tonoSvoris + 1.4 \times spektroSvoris$$

2.4. Garsų jungimo ir keitimo kaštų funkcija

Turint garsų bazę, reikia nuspręsti, kokiu būdu parinkti vienetus iš garsų bazės, kad norima frazė skambėtų kuo natūraliau. Tam yra naudojama garsų jungimo ir keitimo kaštų funkcija, kuri įvertina vieneto tinkamumą sintezuojamoje frazėje. Kaip sako pavadinimas, kaštai galimi dviejų tipų: jungimo ir keitimo. Viena dalis įvertina dviejų vienetų jungimą, kita įterpiamo vieneto kokybę ir tinkamumą.

Toliau pateikiamas pavyzdys iš foneminio sintezatoriaus, difoniniam sintezatoriui jungimo ir keitimo kaštų funkcijos buvo konstruojamos atitinkamai. Tegu užduotis yra sintezuoti garsą iš 3 fonemų *abc*. Tegu jau pasirinktas garsas *a* iš turimo garsų rinkinio. Tada prie jo reikia pridėti garsą *b*, tačiau šis garsas turimas kitame kontekste *ebd*. Remiantis Yi ir Glass [4] viena iš galimų kaštų funkcijų sujungiant *a* ir *b* gali būti tokia:

$$P(b) = C(a,b) + S_L([a]b, [e]b) + S_R(b[c], b[d]), \tag{2.1}$$

- C(a, b) – apjungimo kaštai (angl. concatenation cost)
- S_L([a]b, [e]b) yra kairios pusės keitimo kaštai (angl. substitution cost). Fonema b, einanti po fonemos a, keičiama į fonemą b, einančią po fonemos e.
- S_R(b[c], b[d]) yra dešinės pusės keitimo kaštai. Fonema b, einanti prieš fonemą c, keičiama į b, einančią prieš fonemą d.

Reikia paminėti, kad C(a, b) = 0, jeigu garsų rinkinyje a ir b eina viena po kitos.

2.5. Viterbi paieškos algoritmas

Viterbi algoritmas tai dinamiškas paieškos algoritmas, kuris originaliai buvo sukurtas dešifruoti sąsūkų kodus [17]. Tačiau šiuo metu aktyviai naudojamas daugelyje kitų sričių: kalbos sintezatorių, kalbos atpažinimo, bioinformatikos ir kitose. Foneminiame sintezatoriuje jis buvo naudotas kaštų funkcijos minimumui rasti. Viterbi algoritmo veikimas puikiai iliustruotas Adomo Kondroto [9] rašto darbe. Pavyzdyje sintezuojamas žodis ABAC. Lentelėje yra pateikiama pavyzdinė garsų bazė.

Fonemos	C1	B2	A3	A4	B5	C6	C7	A8	C9	C10	B11	A12
Svoriai	10	20	15	14	30	11	18	20	24	17	13	16

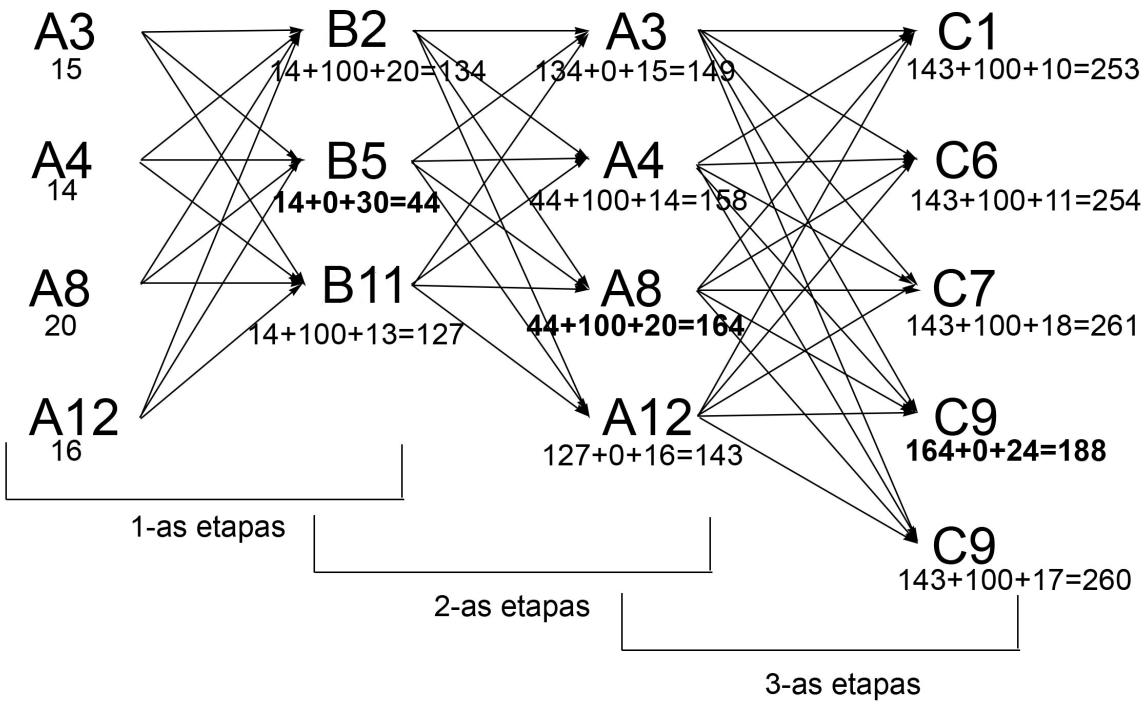
1 lentelė. Lentelė su svoriais

Fonemos raidė atitinką fonemą, matoma, kad egzistuoja 3 fonemos (A, B, C). Skaičius šalia fonemos tipo rodo jos poziciją garsų bazėje. Gretimų skaičių fonemos yra paimtos iš tos pačios

frazės. Tokiu atveju jų jungimo kaina bus lygi 0, priešingu atveju lygi 100 (tai galėtų būti ir koks nors kitas svoris). Taip pat pateikiamos kiekvienos fonemos svoris. Prisimenant garsų jungimo ir keitimo kaštų funkciją ši atrodytų taip. Garsų jungimo kaštai 100, jei garsai nėra imami vienas paskui kitą iš garsų bazės. Garsų keitimo kaštai tai fonemų svoriai.

7 pav. matyti schema, pagal kurią yra ieškoma geriausia vienetų seka. Keturi stulpeliai užpildyti fonemomis, kurios atitinka sintezuojamos frazės (ABAC) fonemas. Rodyklėmis pažymėti visi galimi keliai, kuriuos algoritmas praeis tam, kad rastų kaštų funkcijos minimumą. Tais atvejais, kai jungiamos iš eilės einančios fonemos, pavyzdžiui A4 ir B5, B2 ir A3, B11 ir A12 arba A8 ir C9, jungimo kaštai yra lygūs 0. Kitais atvejais jungimo kaina lygi 100.

Paveikslėlyje taip pat matyti atlikti skaičiavimai. Pirmajame stulpelyje po fonemomis surašyti tik jų pačių svoriai. Kituose stulpeliuose yra sumuojama ankstesnio stulpelio keitimo kaina, jungimo kaina (0 arba 100) ir atitinkamo stulpelio fonemos keitimo kaina. Kiekvieno etapo metu įsimenama tik geriausia ankstesnė kombinacija. Antrame stulpelyje geriausia kombinacija yra A4 ir B5, kaina 44 (nes jungimo kaina buvo lygi 0), tačiau trečiojo stulpelio A3 fonemos kaštai skaičiuojami jungiant ne B5, o B2 (nors ir jos kaina buvo 134). Taip yra todėl, kad B2 ir A3 jungimo kaina bus lygi nuliui. Galiausiai palyginame $134 + 0 + 15 < 44 + 100 + 15$, todėl algoritmas įsimena kelią B2 A3. Tokiu principu algoritmas surenka visą norimą frazę. Algoritmas užtikrina, kad būtų rastas kaštų funkcijos minimumas. Šio pavyzdžio atveju kelias būtų buvęs: A4, B5, A8 ir C9 t. y. frazė ABAC su galutine kaina 188.



7 pav. Viterbi algoritmo schema [9]

3. Garsų bazė ir jos paruošimas

Garso įrašas arba duomenų bazė yra vienas pagrindinių veiksnių, lemiantis sintezatoriaus veikimą. Imant ilgą įrašą bus turimas didžiulis kiekis žodžių, kurių sintezuoti nebereikės, tad kokybiškai sintezatorius skambės gerai, tačiau nukentės sintezatoriaus veikimo laikas. Turint mažą garsų bazę sintezatorius veiks greitai, tačiau daugelį žodžių reikės sujunginėti iš paskirų garsų, o tai paveiks sintezatoriaus kokybę.

3.1. Garsų bazė

Sintezatoriaus konstravimas buvo pradėtas nuo garsų bazės pasiruošimo. Kadangi darbo tikslas yra palyginti vienetų parinkimo metodą su fonemomis ir difonais, tai labai patogu sintezatorių konstravimui naudoti tą patį garso įrašą. Šiame darbe buvo naudotas *Liepos* projekto, diktorės Reginos balso įrašas. Plačiau apie tai, kaip sakiniai buvo parinkti į duomenų bazę, buvo kalbėta 2.2 skyrelyje.

2 lentelė. Pradinė garsų bazė

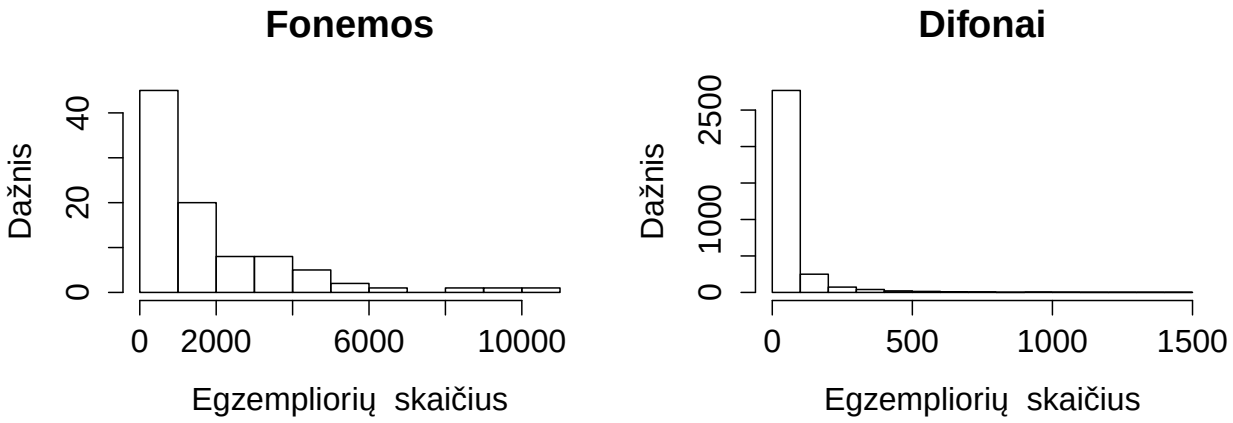
Diktorė	Regina Jokubauskaitė
Įrašų trukmė	3 val.
Žodžių skaičius	31396
Sakinių skaičius	5124
Vienetų skaičius	162448

3 lentelėje matyti, kaip ta pati duomenų bazė skiriasi sintezės vienetu pasirinkus fonemas ir difonus. Reikia atkreipti dėmesį, kad duomenų bazėje yra visos 92 Lietuvių kalboje pasitaikančios fonemos, tačiau joje trūksta, net 5981 teoriškai galimų difonų ($92 \times 92 = 8464$) t. y. beveik kas trečias difonas neegzistuoja turimoje garsų bazėje. P. Kasparaitis [5] teigia, kad Lietuvių kalboje pasitaiko 5003 difonai, kitos fonemų kombinacijos nepasitaiko arba pasitaiko nelietuviškuose žodžiuose. Taigi duomenų bazėje turima kiek daugiau nei pusė visų Lietuvių kalboje pasitaikančių difonų. Visgi sintezatorius turėtų sugebėti ištarti net pačias keisčiausias garsų kombinacijas, nes jos gali pasitaikyti tariant kitų kalbų žodžius. Tad laikysime, kad trūksta visų 5981 teoriškai galimų difonų. Trūkstamų difonų problemos sprendimas aprašytas 4.6 skyrelyje.

3 lentelė. Fonemų ir difonų statistikos

	Fonemos	Difonai
Skirtingų vienetų skaičius	92	3187
Vidutinis vienodų vienetų skaičius	1765	51
Vienodų vienetų skaičiaus mediana	1011	14
Trūkstamų vienetų skaičius	0	5981

Lyginant fonemų ir difonų garsų bazes taip pat pastebima, kad fonemų atveju yra turimas didelis kiekis fonemų egzempliorių. Tuo tarpu difonų egzempliorių skaičius yra kur kas mažesnis. Tai iliustruoja fonemų ir difonų egzempliorių histogramos (8 pav). Matyti, kad didžioji dauguma fonemų turi bent po 2 tūkstančius egzempliorių. Tuo tarpu difonų egzempliorių skaičius vos siekia šimtą.

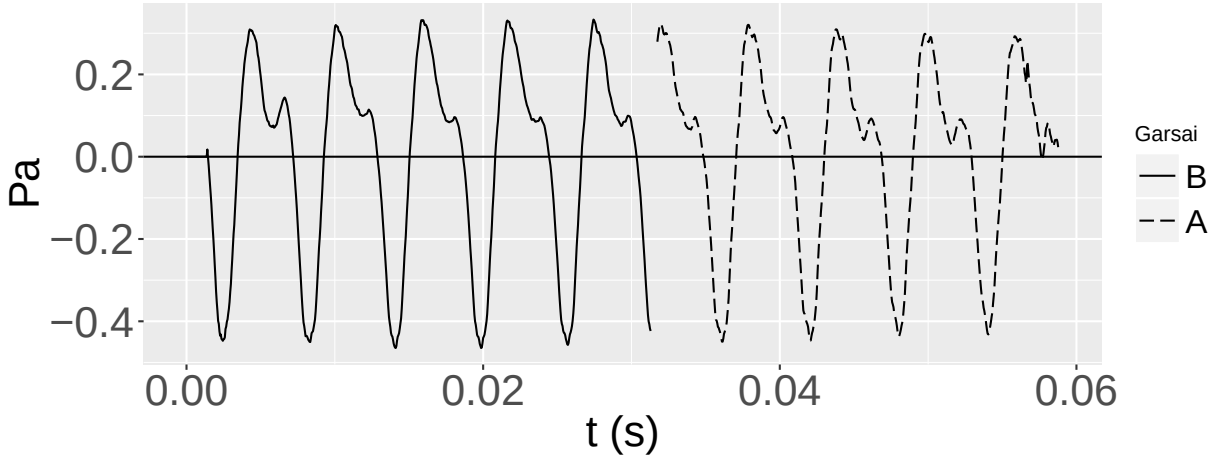


8 pav. Fonemų ir difonų egzempliorių skaičiaus duomenų bazėje histogramos

Žemiau matyti, kaip garso įrašas atrodo, kai būna atlikti įrašo paruošimo darbai. Tekstas suskiemenuotas, sukirčiuotas, raidės pakeistos į sintezės vienetus, šiuo atveju fonemas. Greta garso įrašo turime masyvą, kuriame sužymėti visų garsų pradžios ir pabaigos taškai. Papildomu "+" simboliu žymimi garsai esantys žodžio gale, simboliu "–" žymima skiemens pabaiga. Žemiau pateikiama sukirčiuota ir suskiemenuota frazė „kad statybininkai pasistengs“. Ši duomenų bazė buvo modifikuota ir pritaikyta darbui su difonais.

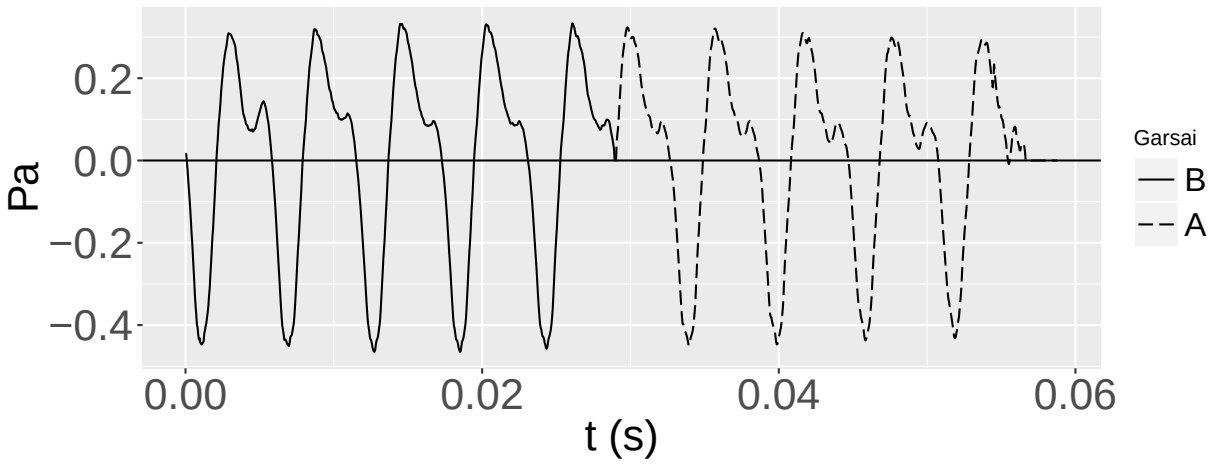
```
_+k a ts+t a-t' Ii-b' i-n' i n-k a j+p a-s' i-s' t' E N k s+_
```

Pirmoji užduotis buvo sukarpyti garsus nuo vieno garso vidurio iki kito garso vidurio. Garsų karpymui yra skirta programa *Praat* [12], kuri ir buvo naudojama. Prieš kerpant garsą reikia nustatyti kiekvieno garso vidurį. Tai atliekama imant aritmetinį vidurkį laiko momentų garso pradžioje ir pabaigoje. Tačiau kirpti garso tiesiog viduryje negalima. Pateikiamas pavyzdys su fonema /m/.



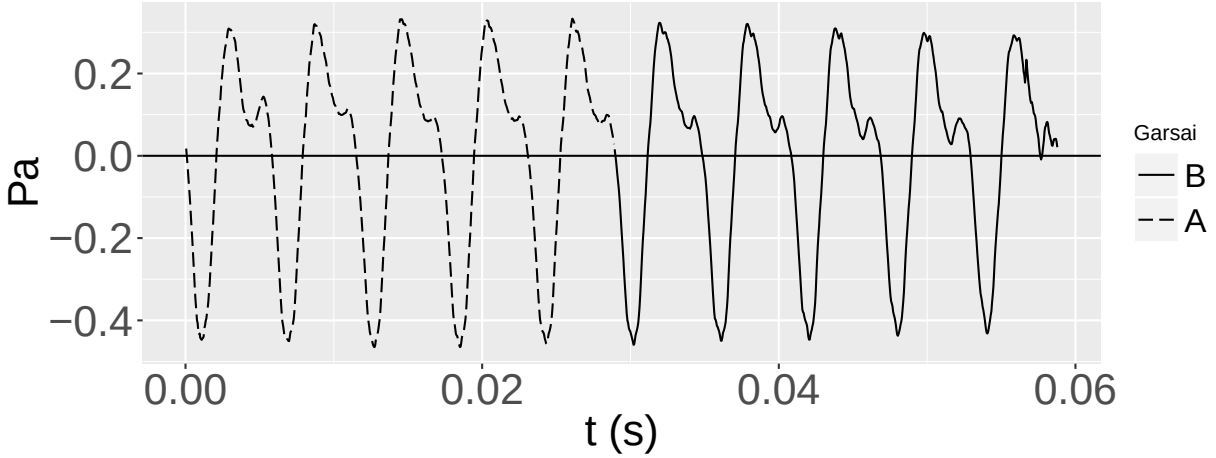
9 pav. Garso *m* signalų jungimas pirmas atvejis.

Kerpant dviejų fonemų /m/ signalus, o vėliau juos bandant sujungti galima gauti gauti plyšius tarp signalų, kurie matyti 9 pav. Tai nėra gerai, nes klausantis šioje vietoje pasigirstų trakštelėjimas. Kad taip nenutiktų, garsus leidžiama kirpti tik ten, kur jų signalai lygūs nuliui. Tada juos jungiat neliks plyšių.



10 pav. Garso *m* signalų jungimas antras atvejis.

Tačiau 10 pav. iliustruoja, kad fonemas jungiant taškuose, kuriuose abu signalai lygus nuliui gali būti neišlaikytas signalo periodiškumas. Tokių atvejų reikia vengti, todėl įvedama dar vieną taisyklė, garsai kerpami taškuose, kuriuose signalas lygus nuliui, ir pats signalas mažėja. To rezultatas – gaunamos tvarkingos garsų signalų jungtys, viena jų pavaizduota 11 pav.



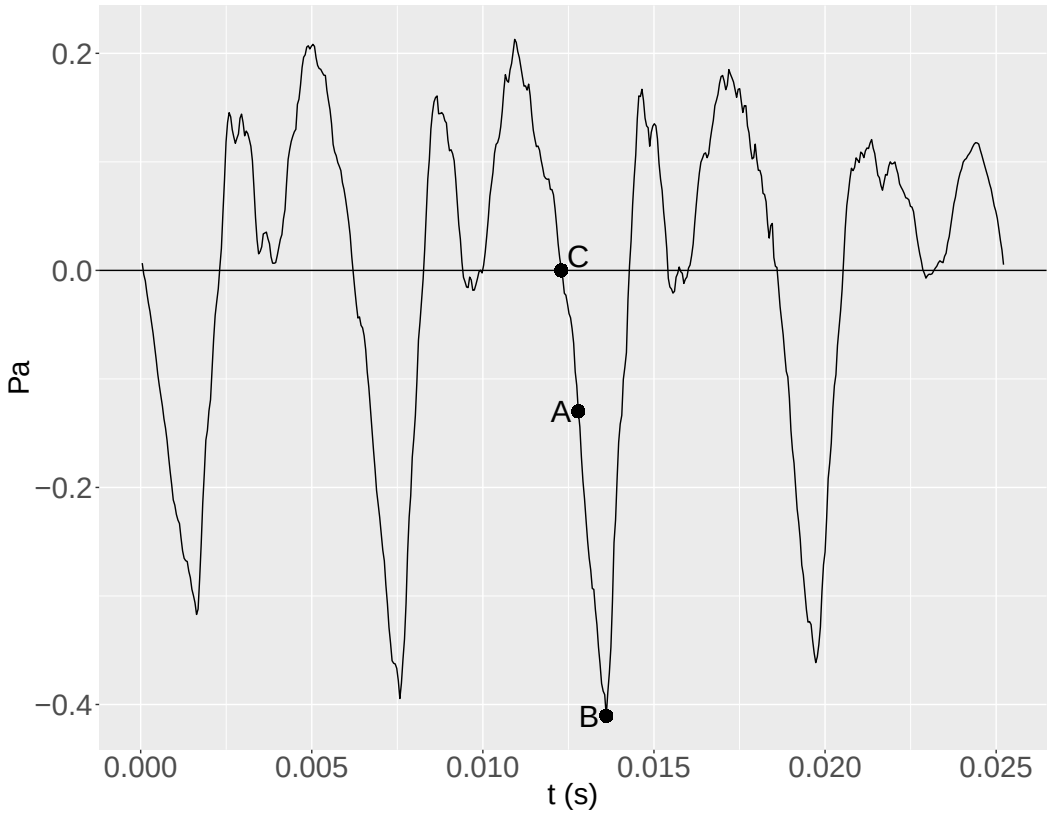
11 pav. Garso *m* signalų jungimas antras atvejis.

3.2. Balsiai, pusbalsiai ir skardieji sprogstamieji priebalsiai

Algoritmas, tinkantis daugumai garsų t. y. balsiams, pusbalsiams (l, m, n, j, v) ir skardiesiems sprogstamiesiems priebalsiams (b, d, g):

1. Rasti garso vidurį.
2. Aptikti garso periodą.
3. Aptikti minimumą garso periode.
4. Rasti artimiausią nulį kertantį tašką, kuriame signalas mažėja.

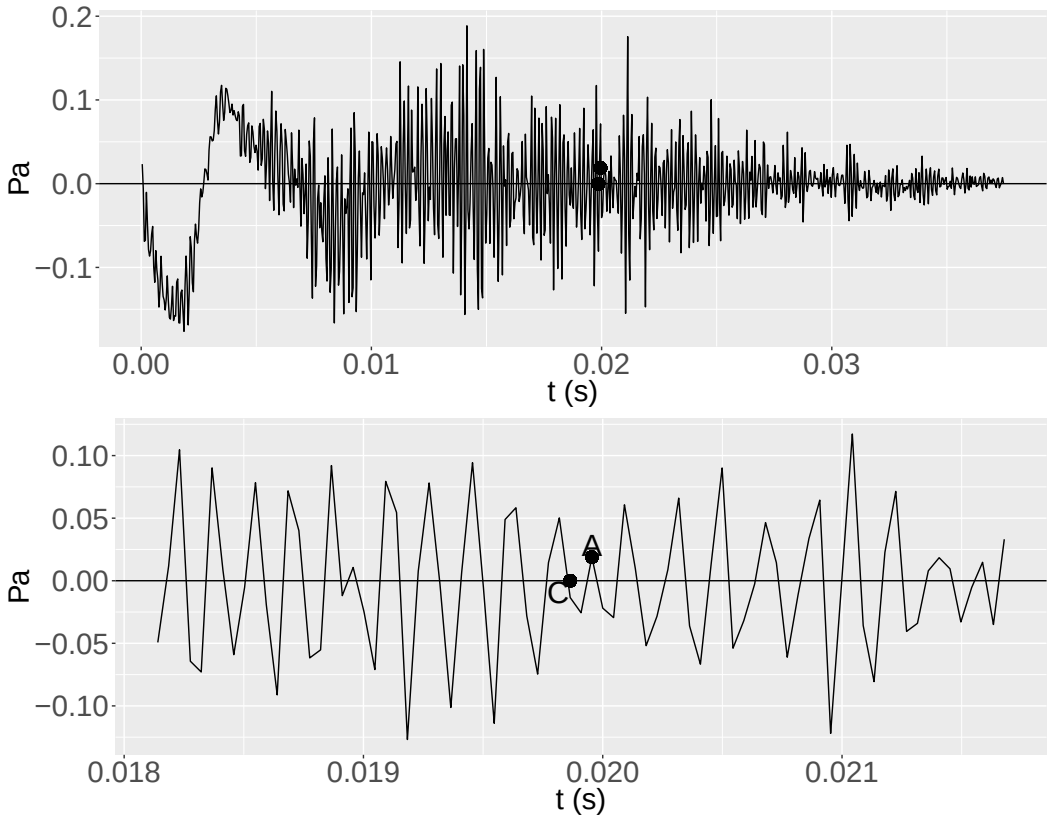
12 pav. matomi trys taškai. Taške A yra garso aritmetinis vidurys. Taške B rastas periodo minimumas. Taškas C yra vieta, per kurią reiktų kirpti šį garsą. Deja, ne visų garsų signalai turi tokius aiškius periodus.



12 pav. Garso *i* signalo pavyzdys.

3.3. Šnypščiantys duslieji s, š, f ir x garsai

Matyti, kad šie garsai turi aukšto dažnio signalus. Vieną jų smarkiai priartinus (žemiau esantis paveikslėlis), galima matyti signalo vidurį – tašką A ir šalimais pasirenkamą tašką C, kuriame signalas lygus 0. Šiems garsams jokie algoritmai netaikomi. Randamas signalo vidurys ir kerpama per šalia jo esantį artimiausią nulio tašką.



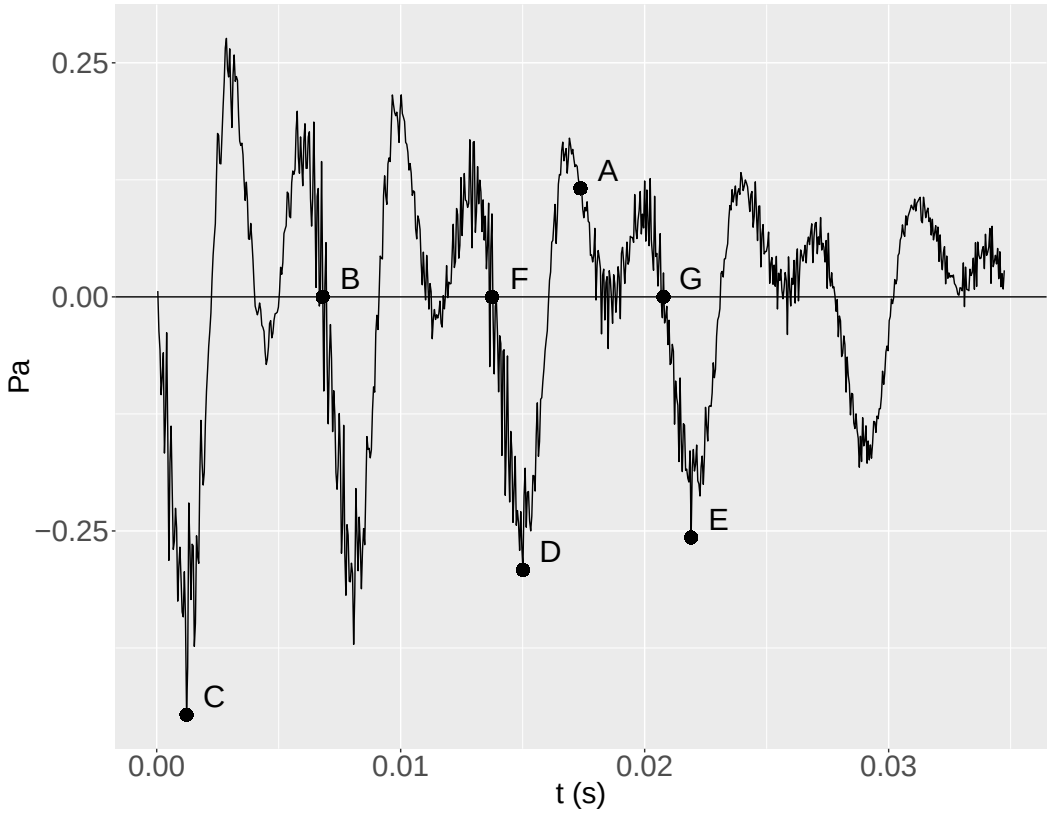
13 pav. Garso *s* signalo pavyzdys.

3.4. Šnypščiantys skardieji ž, z, dž, dz, h garsai

Nors šie garsai turi periodus, visgi pats signalas yra panašiai dygliuotas, kaip kad 14 ir 15 pav. Pateikiamas pavyzdys, kaip parinkti kirpimo vietą signalui 14 pav., kuris yra mažai dygliuotas, tačiau tuo pačiu metodu galima karpyti ir kur kas labiau dygliuotus signalus (15 pav.). Kaip įprastai siekiama, kad kerpant garsą būtų išsaugotas periodas. Reikia paminėti, kad šių signalų periodai išryškėja kraštuose, tuo tarpu garso viduryje periodai gali pranykti 15 pav. Taigi algoritmas šiems garsams yra toks:

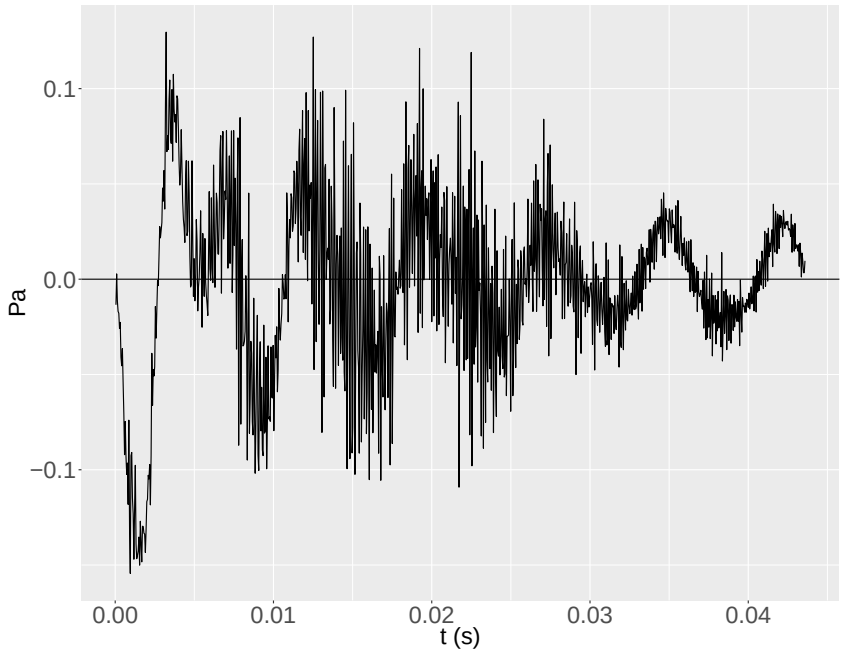
1. Aptikti signalo periodą garso pradžioje.
2. Apskaičiuoti periodo ilgį.
3. Rasti atstumą nuo periodo minimumo iki periodo pradžios.
4. Rasti garso vidurį.
5. Per periodo ilgio atstumą rasti signalo lokalius minimumus į kairę ir dešinę nuo garso vidurio.
6. Nuo rastų minimumų judėti į kairę pusę per atstumą rastą 3 punkte.
7. Aplink tuos taškus ieškoti, kur signalas mažėja ir kerta nulio ašį.
8. Parinkti tašką esantį arčiau vidurio.

Pagal aukščiau aprašytą algoritmą pirmiausiai randamas garso periodas (nuo garso pradžios iki taško B). Išmatuojamas jo ilgis (0.07s). Randamas periodo minimumas (taškas C) ir apskaičiuojamas atstumas iki garso pradžios taško (0.015s). Randamas signalo vidurys (taškas A). Į kairę pusę per 0.015s nuo taško A randamas pirmas minimumas (taškas D), dešinėje randamas antras minimumas (taškas E). Nuo šių taškų judėdama į kairę per 0.07s aptinkami taškus F ir G atitinkamai. Imamas esantis arčiau vidurio t. y. taškas G.



14 pav. Garso z signalo pavyzdys.

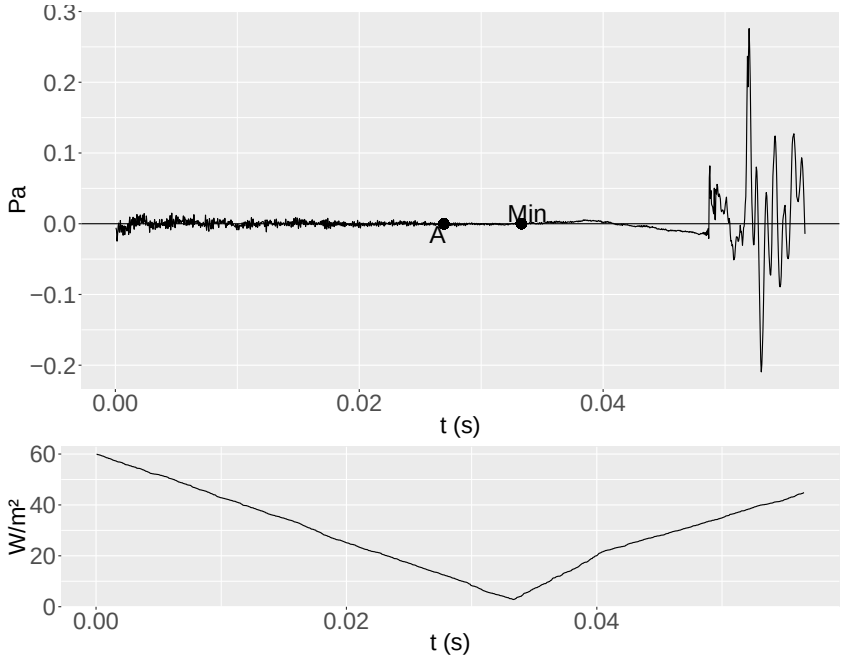
Šiame pavyzdyje vizualiai buvo aišku, ką reikėjo daryti. Tačiau daugelis šios grupės garsų signalų yra labai spygliuoti. 15 pav. matomas toks atvejis, visgi algoritmas randa tašką tinkamą kirpimui, net ir tokiame signale.



15 pav. Kitas garso *z* signalo pavyzdys.

3.5. Duslūs sprogstamieji k, p ir t garsai

Duslūs garsai gerokai skiriasi nuo ankščiau matytų garsų. Šie garsai dar vadinami sprogstančiais, jie turi ilgą pauzę, o pačiame gale pasigirsta garsas. Reikia paminėti, kad panašią savybę turi skardūs sprogstamieji garsai: b, d ir g. Visgi nagrinėtuose atvejuose šiems garsams puikiai tiko skardžių priebalsių metodas.



16 pav. Garso *p* signalo ir jo stiprio grafikai.

Taigi garsams k, p ir t rasti kirpimo vietai bus naudojamas garso stiprio signalas, kuris matomas 16 pav. Programa *Praat* leidžia rasti garso stiprio minimumą pasirinktoje signalo dalyje. Minimumas įprastai randamas duslių garsų pauzės dalyje. Šiame signale, kuris matyti 16 pav.

tai būtų taškas Min, o taškas A, kaip įprastai, yra garso vidurys. Šį sykį net nereikia, kad garso kirpimo vieta būtų arti vidurio. Imama ta vieta, kurioje garso stipris pasiekia minimumą.

Buvo apžvelgti skirtingų garsų signalai ir parinkti metodai garso kirpimo vietai nustatyti. Sukarpius visus garsus gaunama difonų bazė. Ta pati frazė, „kad statybininkai pasistengs“ pritaikius ją difonams atrodo taip:

```
_+ k|k a|a ts+|ts+ t|t a-|a- t'|t' Ii-|Ii- b'|b' i-|i- n'|n' i|i n-|
n- k|k a|a j+|j+ p|p a-|a- s'|s' i-|i- s'|s' t'|t' E|E N|N k|k s+|
s+ _|_ _+
```

Nors vaizdžiai iš signalų pavyzdžių galima matyti, kad buvo kerpama teisingose vietose, tačiau reikia suprasti, kad dalis šių jungčių gali skambėti neįprastai. Nes nors ir jungiama arti nulio, tačiau difono pradžioje ir pabaigoje gali skirtis difono tembras, spektras, intonacija ar tonas. Difonų svoriai turėtų gelbėti šioje vietoje.

4. Sintezatoriaus konstravimas

Paruošus difonų bazę buvo imtasi konstruoti sintezatorių. Tam reikėjo išspręsti šiuos uždavinius:

- Sukurti difonų parinkimo algoritmą;
- Algoritmą integruoti į *Liepos* projekto aplikaciją;
- Sukurti modulį, kuris skaičiuotų skiemenų vietos žodyje atitikimą;
- Sukurti modulį, kuris skaičiuotų žodžių vietos sakinyje atitikimą;
- Prie algoritmo pridėti difonų svorių kainų skaičiavimą;
- Prie algoritmo pridėti difonų jungimo kainų skaičiavimą;
- Išspręsti trūkstamų difonų problemą.

Reikia pastebėti, kad dalis sintezatoriaus funkcijų, nesusijusių su difonais (tokios kaip teksto vertimo į garsų seką, kirčiavimo, skiemenavimo), buvo perimtos iš *Liepos* projekto.

4.1. Difonų parinkimo algoritmas

Difonų parinkimui buvo realizuotas Viterbi paieškos algoritmas. Plačiau apie patį algoritmą buvo kalbėta 2.5 skyrelyje. Algoritmas įsimena visas difonų kombinacijas surinkti norimai žodinei kombinacijai. Kiekvienai kombinacijai suskaičiuojama kaina, o galiausiai pasirenkama kombinacija su mažiausia kaina.

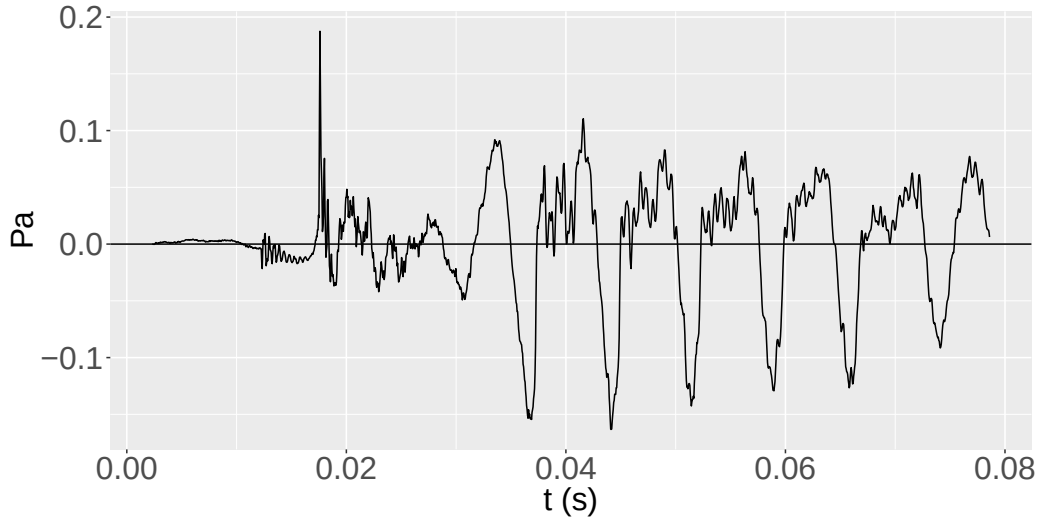
Nors buvo turimas foneminio sintezatoriaus fonemų parinkimo algoritmas su optimizacija, buvo nuspręsta realizuoti atskirą algoritmą difonų parinkimui, nes difonai ir fonemos turi nemažai skirtumų. Optimizacija, kuri buvo naudoja foneminiame sintezatoriuje, nebuvo pritaikyta difoniniam, nes kaip vėliau paaiškėjo sintezatoriaus veikimo laikas yra pakankamas, o tos pačios optimizacijos poveikis nebūtų toks stiprus kaip foneminiam sintezatoriui.

4.2. Difonų svoriai

Fonemų sintezatoriuje buvo naudoti fonemų svoriai, kurie apibūdina fonemos panašumą į kitas to paties tipo fonemas (tarkime, visas /aA/ fonemas). Apie fonemų svorių skaičiavimą parašyta 2.3 skyriuje. Atitinkamai difonams norėta priskirti svorius. Tačiau svoriai difonams nebuvo perskaičiuoti 2.3 aprašytu būdu, nes pasikartojančių difonų skaičius turimoje duomenų bazėje yra gerokai mažesnis nei pasikartojančių fonemų. Nors yra difonų, kurių pasikartojančių egzempliorių skaičius yra 1437 (tai difonas/s+ _/ t. y. garsas s žodžio gale), tačiau daugelis turimų difonų turi kur kas mažesnį pasirinkimą. Dar 400 difonų turi bent po 100 egzempliorių. Reiškia didžiajai daliai difonų būtų nekorektiška vienus difonus vadinti tipiniais ir kitus netipiniais. Plačiau apie difonų bazę ir difonų egzempliorių skaičius 3 skyriuje. Todėl buvo nuspręsta difonų svoriams priskirti difoną sudarančių fonemų svorių vidurkį. Tai turėtų neblogai atspindėti difonų kokybę.

4.3. Garsų vieta žodyje ir sakinyje

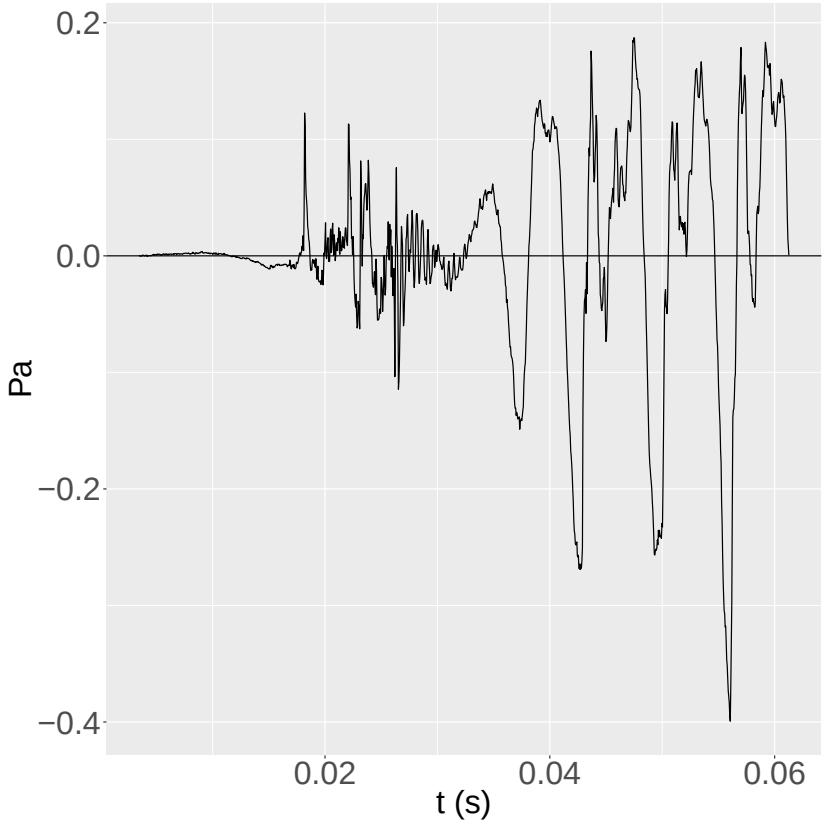
Difono vieta sakinyje ir žodyje yra svarbi. Difonai pasitaikantys žodžio pradžioje, gali skirtis nuo difono ištarto žodžio gale. Šis skirtumas dar labiau išryškėja imant difoną iš sakinio pradžios ir galo. Žemiau pateikiamame pavyzdyje 17 ir 18 pav. matyti dviejų difonų /t a/ signalai. Vienas paimtas iš sakinio pradžios, kitas iš sakinio galo. Kaip matyti skiriasi difonų ilgiai ir signalų stiprumai. Sakinio pradžioje garsai turi daugiau energijos ir yra trumpesni. Tuo tarpu sakinio gale garsai tampa ilgesni, tačiau turi mažiau energijos. Tai atitinkamai matyti pavyzdyje. Sakinio pradžioje difonas trunka 0.06 s, o sakinio gale 0.08 s. Signalo amplitudė didesnė sakinio pradžioje nuo 0.2 Pa iki -0.4 Pa, tuo tarpu sakinio gale nuo 0.2 Pa iki -0.2 Pa. Matoma, kad difono vieta sakinyje, taip pat ir žodyje yra svarbi. Todėl reikia sukurti modulį, kuris nustatytų difono vietą žodyje ir sakinyje bei palygintų su vieta, į kurią norima tą difoną įklijuoti. Atitinkamai difonams, kurių vieta žodyje ar sakinyje neatitinka, reikia priskirti kainas, dėl kurių algoritmas tokių difonų vengtų ir ieškotų difonų iš reikiamos žodžio ar sakinio dalies.



17 pav. Difono /t a/ signalas žodžio gale

Fonemų vieta žodyje skirstoma į tris atvejus. Pirmas skiemuo – žodžio priekis, paskutinis skiemuo – žodžio galas, skiemenys tarp pirmo ir paskutinio – žodžio vidurys. Jei žodis vienskiemenis – ketvirtas atvejis. Panašiai skirstomos fonemos sakinyje. Pirmas žodis – sakinio pradžia, paskutinis žodis – sakinio pabaiga. Žodžiai tarp pirmo ir paskutinio – sakinio vidurys. Sakiniai iš vieno žodžio atskiras atvejis.

Difonams šis paskirstymas netinka, nes difonas susideda iš dviejų fonemų, reiškia difonas gali tuo pačiu būti ir pirmame, ir antrame skiemenyje. Todėl kiekvienam difonui buvo priskiriami du indeksai. Pirmasis iš pirmos difono dalies ir antras iš antros. Visuose sintezatoriuose žodžio pozicijos sakinyje nesutapimo kaina buvo lygi 30, o skiemens pozicijos žodyje nesutapimo 10. Tačiau difonų atveju gali vieta sutapti pusėje difono, o kitoje nesutapti. Tad šios kainos tampa 15 ir 5 jei pusė difono sutampa, o pusė ne.



18 pav. Difono /t a/ signalas žodžio pradžioje

4.4. Difonų jungimo kainos

Buvo bandyti trys difonų jungimo kainų skaičiavimo būdai.

- Difonų jungimo kaina pagal garsų grupes
- Difonų jungimo kaina pagal difonų spektro skirtumus
- Difonų jungimo kaina pagal tono skirtumus

Tačiau difonų jungimo kainos skaičiavimas pagal tono skirtumą pasirodė prasčiausiai, tad plačiau šiame tyrime nebus aptariamas.

Šioje vietoje reikėtų paminėti dar vieną pastabą. Foneminiam sintezatoriui buvo pritaikytas optimizavimas, kuris gerokai jį pagreitino. Jei ateityje būtų norima optimizuoti difoninį sintezatorių reikia nepamiršti, kad tas pats optimizavimas veiks tik sintezatoriuije, kuriame difonų jungimo kaina priskiriama pagal garsų grupes.

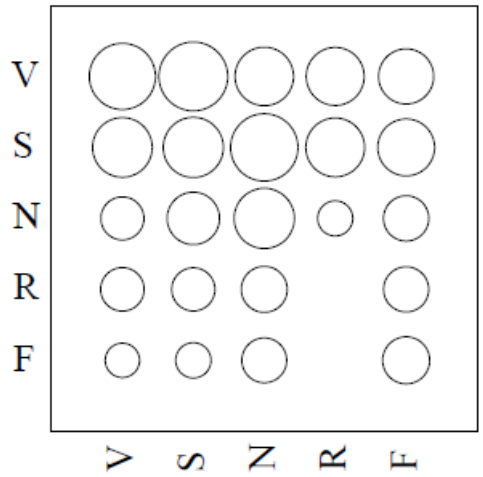
4.4.1. Difonų jungimo kainos pagal garsų grupes

J. Glass [4] pastebi, kad skirtingose garsų grupėse jungti difonus reikia skirtingomis kainomis. Jis teigia, kad vienose garsų grupėse jungtys skambės geriau, kitose prasčiau.

- Balsiai (angl. vowel) (a, e, i, u ...)
- Pusbalsiai (angl. semivowel) (j, w, ...)
- Nosiniai garsai (angl. nasal) (n, m, ...)

- Sprogstamieji garsai (angl. stop - realease sounds) (k, p, t, ...)
- Šnypščiantys (angl. fricatives) (s, z, ...)

19 pav. pateikiamos rekomenduojamos jungimų kainos. Pagal šią schemą buvo suskirstytos lietuviškos fonemos ir priskirtos kainos esančios 4 lentelėje. Kirčiuotiems garsams buvo nuspręsta pridėti papildomą + 10 kainą, kad būtų vengiama jungti per kirčiuotus difonus.



19 pav. Glass siūlomos jungimo kainų proporcijos [4]

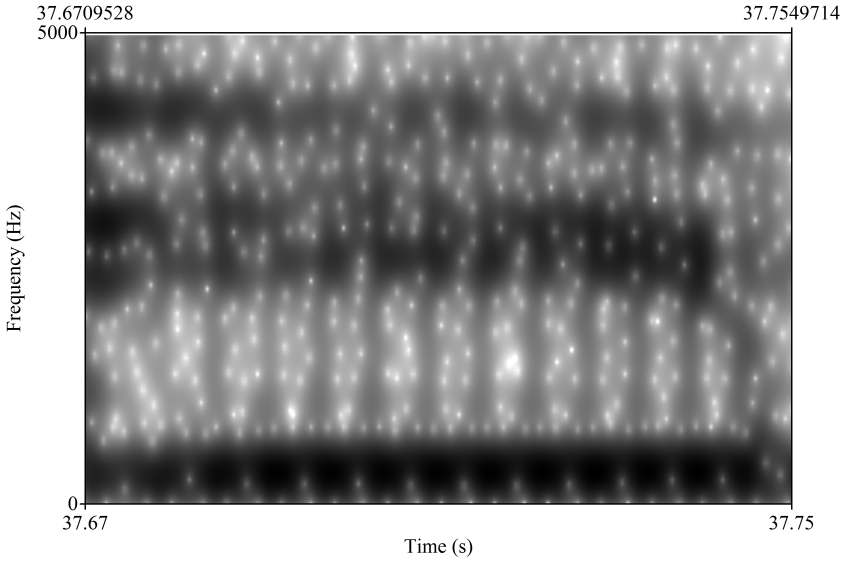
4 lentelė. Garsų grupių jungimo kainos

Garsų grupė	Kaina
Balsiai	90
Pusbalsiai	80
Nosiniai garsai	70
Sprogstamieji	30
Šnypščiantys	40

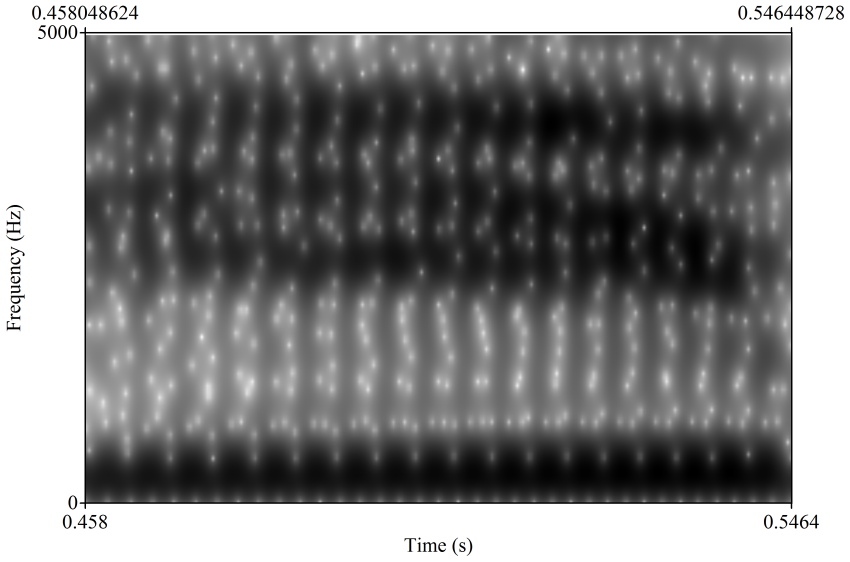
4.4.2. Difonų jungimo kainos pagal spektro panašumą

Garso spektras apibūdina vibracijas skirtinguose dažniuose. Vienodų fonemų, tuo pačiu ir difonų, spektrai gali gerokai skirtis vienas nuo kito. 20 ir 21 pav. matyti dvi tos pačios fonemos /Ii/ spektogramos. Spektogramoje juoda spalva žymi dažnių buvimą, balta - nebuvimą. Nors fonema ta pati, spektogramos gerokai skiriasi. Pirmoji spektograma gana blausi, joje didžiausias kiekis yra žemų dažnių. Tuo tarpu antrojoje yra panašus kiekis aukštų dažnių, kaip kad ir žemų.

Manoma, kad sintezatoriui jungiant vienetus jungimo kainą būtų galima skaičiuoti pagal spektro skirtumus. Kuo panašesni vienetų spektrai, tuo natūralesnė turėtų būti jungtis, t. y. tuo mažesnė turėtų būti jungimo kaina. Kiekvieno difono spektrui buvo suskaičiuoti 14 koeficientų, tad jungimo kaina buvo skaičiuojama pagal Euklidinį atstumą tarp dviejų difonų spektro koeficientų vektorių.



20 pav. Fonemos |Ii| spektograma



21 pav. Kitos fonemos *Ii* spektograma

4.5. Sintezatorių konfigūravimas

Kaip kad buvo minėta 2.4 skyrelyje surinktos difonų sekos kaina susideda iš dviejų komponenčių: jungimo ir keitimo kainų. Jungimo kainos parinkimui buvo bandyti du variantai. Konstanta – pagal difonų, ties kuriais jungiama, tipą, kaip kad siūlo J. Glass [4] ir kintanti kaina pagal difonų spektro koeficientų Euklidinį atstumą. Keitimo kaina susidėjo iš įterpiamo difono svorio, jo vietos žodyje ir sakinyje atitikimo kainų. Kiekviena komponentė turėjo papildomus svorius, kuriuos keičiant galima konfiguruoti sintezatorių. Fonemų sintezatoriaus koeficientai yra matomi 5 lentelėje.

5 lentelė. Fonemų sintezatoriaus kainų svoriai

	Koeficientas
Fonemų svorio	0.5
Keitimo kainos	1
Jungimo kainos	0.8

Testuojant difoninius sintezatorius buvo bandytos įvairios koeficientų kombinacijos. Galiausiai buvo parinkti koeficientai matomi 6 ir 7 lentelėse.

6 lentelė. Difonų sintezatoriaus pagal garsų grupes kainų svoriai

	Koeficientas
Difonų svorio	0.5
Keitimo kainos	1
Jungimo kainos	1

Kaip matyti, difoninių sintezatorių kainų svoriai skiriasi ties jungimo kaina. Kadangi difoninio sintezatoriaus pagal spektro panašumą, lyginami du vektoriai su 14 koeficinetų jų Euklidiniai atstumai gali patapti labai dideli skaičiai. Tad jiems priskiriamas mažesnis svoris.

7 lentelė. Difonų sintezatoriaus pagal spektro panašumus kainų svoriai

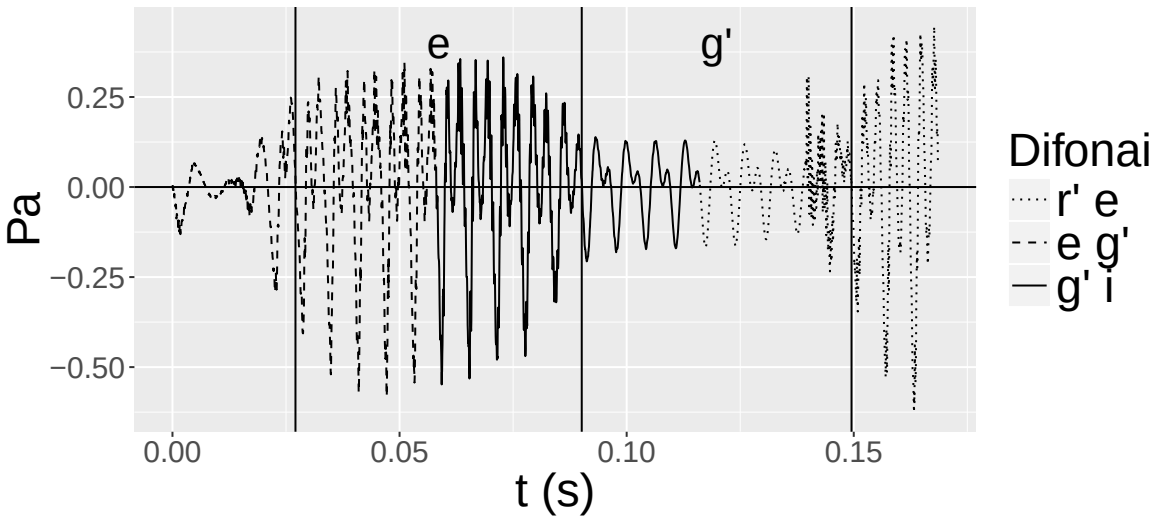
	Koeficientas
Difonų svorio	0.5
Keitimo kainos	1
Jungimo kainos	0.5

4.6. Trūkstamų difonų problema

3 skyriuje buvo pastebėta, kad didžiulis kiekis difonų, kurie teoriškai gali pasitaikyti sintezuojant tekstą, neegzistuoja turimoje duomenų bazėje. Laimei atvejai, kai prireikia difonų, kurių garsų bazėje nėra, yra reti. Atlikus testą su 20 000 sakinių, kuriuose buvo 798 tūkstančiai difonų, suskaičiuota, kad difonas nebuvo rastas 2895 kartus (0.36% atvejų). Šiais atvejais trūko 683 difonų, daugelio jų trūko kelis kartus. Mark Beutnagel [2] savo tyrime irgi pastebi, kad turimoje garsų bazėje trūksta daugelio galimų difonų. Tačiau tikėdamasis, kad tokie difonai retai pasitaiko, plačiau problemos nesprendžia.

Remiantis [5] yra siūlomi du sprendimo būdai. Pirmasis siūlo kiekvieną trūkstamą difoną pakeisti tokiu, kuris egzistuoja. Pavyzdžiui kietą priebalsį prieš minkštą priebalsį pakeisti į minkštą priebalsį, t. y. /b-b"/ keisti į /b"-b"/.

Kitas būdas trūkstamo difono atvejus spręsti imant ir sujungiant šalia esančius difonus fonemine jungtimi. Tarkime sintezuojama frazė yra *reg* iš žodžio *registracija*, o difonas /e g'/ neegzistuoja. Tokiu atveju imamas difonas /r' e/ ir prie jo pabaigos dar pridedama tai ko trūksta iki pilnos fonemos /e/. Taip pat daroma ir kitame difone /g' i/ imamas garso įrašas nuo fonemos /g'/ iki difono /g' i/ pabaigos. 22 pav. skirtingomis linijomis pažymėti trys difonai: /r' e/, /e g'/, /g' i/. Vertikalūs brūkšniai žymi fonemų /e/ ir /g'/ ribas.



22 pav. Difonų jungimas fonemine jungtimi

Šis metodas neveikia, jeigu eina du iš eilės neegzistuojantys difonai, tačiau tokie atvejai labai reti ir natūralioje kalboje praktiškai nepasitaiko. Ankščiau minėtame teste tokių atvejų buvo 11 (t. y. apie 0.001%). Tai buvo tokie nelietuviški žodžiai, kaip „Hebeštrait“ ar „Reichsministras“.

5. Sintezatorių įvertinimas ir palyginimas

Sukurti sintezatoriai buvo įvertinti ir palyginti su fonemų pagrindu kurtu sintezatoriumi. Visų pirma buvo palyginti sintezatorių algoritmai ir jų veikimo laikas, sintezuojant duotą tekstą. Taip pat buvo suskaičiuoti statistiniai sintezės įverčiai, kad galėtume įvertinti sintezatorių veikimą. Galiausiai buvo atliktas testas su žmonių grupe.

5.1. Algoritmų veikimo laikas

Kiekvienam iš trijų sintezatorių (foneminiam ir dviem difoniniams) buvo pateikta 20 000 sakinių, kuriuos reikėjo susintezuoti. Algoritmų veikimo laikas pateiktas 8 lentelėje.

8 lentelė. Algoritmų veikimo laikas

	Laikas
Fonemų sintezatorius	7595 s
Difonų sintezatorius su garsų grupėmis	3508 s
Difonų sintezatorius su spektru	3632 s

Kaip matyti foneminis sintezatorius užtruko beveik dvigubai ilgiau. Iš difoninių sintezatorių nežymiai, tačiau greičiau veikė sintezatorius su garsų grupėmis. Visgi net ir foneminis sintezatorius veikia pakankamai greitai, nes per 7595 s susintezuojami 20 000 sakinių, t. y. 0.36 s sakiniui susintezuoti. Šis laikas yra nereikšmingas palyginus su laiku, kurio prireikia perskaityti patį sakinį.

Pastebėkime, kad foneminis sintezatorius, apie kurį iki šiol buvo kalbėta yra optimizuotas (plačiau apie sintezatoriaus optimizaciją [6]). Tuo tarpu difoninis sintezatorius nėra optimizuotas. Taigi teoriškai būtų galima palyginti kombinacijų skaičių, kurį neoptimizuoti foneminis ir difoninis sintezatoriai turi patikrinti, kad rastų geriausią sintezės vienetų kombinaciją. Tam buvo naudoti tie patys 20 000 sakinių.

9 lentelė. Algoritmų sudetingumo statistikos

	Vidutinis kombinacijų skaičius
Fonemų sintezatorius	4164
Difonų sintezatorius	294

Tai reiškia, kad vidutiniškai turėtai tekstinei užklausai kiekvienai fonemai teko patikrinti 4164 egzempliorių ir išsirinkti geriausią. Tuo tarpu difoninis sintezatorius rinkdavosi iš 294 difonų. Tai didžiulis difoninio sintezatoriaus pranašumas. Kaip ir buvo minėta, foneminiam sintezatoriui galima naudoti optimizavimą, kuris gerokai pagreitina fonemų egzempliorių parinkimą. P. Kasparaitis [6] naudoja optimizaciją, kuri perrenkamų fonemų skaičių beveik prilygina, perrenkamų difonų skaičiui.

5.2. Sintezės statistinis įvertinimas

Pijus Kasparaitis [6] mini keletą įverčių, kurie tinka įvertinti vienetų parinkimo metodą. Buvo pasirinkti įverčiai, kurie tinka tiek fonemų, tiek difonų sintezatoriams. Jų buvo du:

- Paeiliui einančių vienetų ir visų užklausoje naudotų vienetų santykis procentine išraiška
- Vidutinis paeiliui einančių vienetų sekos ilgis

Testas buvo atliktas su tais pačiais 20 000 sakinių. Pirmojo įverčio rezultatai matyti 10 lentelėje. Difoninis sintezatorius naudoja ilgiausias sintezės vienetų sekas. Kiek trumpesnes difoninis su spektru sintezatorius, o trumpiausias sekas naudoja fonemų sintezatorius.

10 lentelė. Sintezatorių statistikos

	Sekos ilgis
Difonų su garsų grupėmis	2.55
Difonų su spektru	2.31
Fonemų	2.03

Antrojo įverčio rezultatai pateikiami 11 lentelėje. Rezultatai panašūs, didžiausią procentą paeiliui einančių vienetų turi difoninis sintezatorius su garsų grupėmis. Truputėlį mažiau difoninis sintezatorius su spektru. Mažiausias procentas yra foneminio sintezatoriaus.

11 lentelė. Sintezatorių statistikos

	Paeiliui einančių vienetų %
Difonų su garsų grupėmis	57%
Difonų su spektru	53%
Fonemų	50%

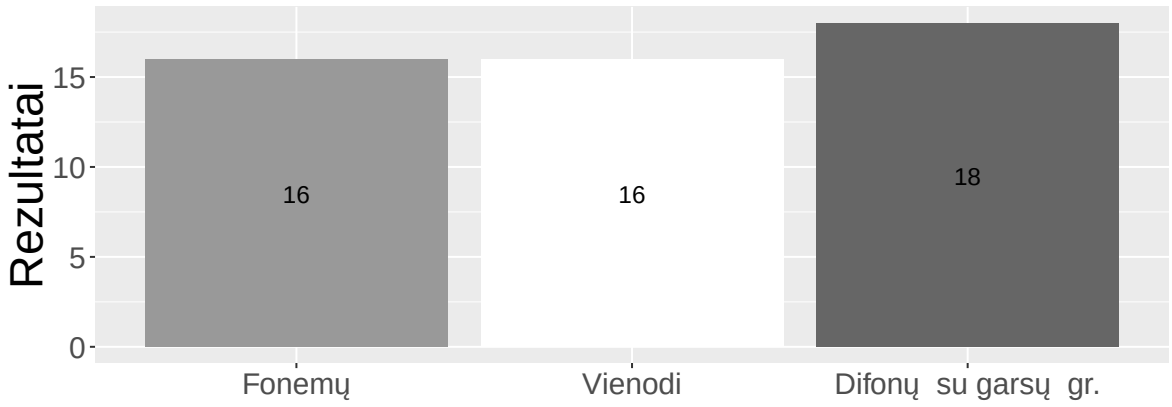
Abu šie įverčiai įvertina jungimų kiekį. Jei visa užklausa bus paimta iš turimos duomenų bazės ir nereikės nė vieno jungimo, ji skambės maksimaliai gerai. Iš kuo daugiau dalių bus sujungta užklausa, tuo didesnė rizika, kad pasitaikys jungtys, kurios skamba nenatūraliai. Taigi, kuo didesnė procentinė išraiška paeiliui einančių vienetų, tuo užklausa skambės natūraliau. Taip pat, kuo ilgesnė vidutinė paeiliui einančių vienetų seka, tuo geriau. Tačiau jeigu sintezatorius bus sukalibruotas taip, kad visada pasirinktų vienetus iš duomenų bazės siekdamas, kad sekos paeiliui einančių vienetų būtų kuo ilgesnės, bus susidurta su kita problema. Problema, kad dalis tų jungčių bus prastos, pavyzdžiui bus jungiama kirčiuotose balsiuose, kas yra blogiausias iš galimų variantų. Tad ilgos paeiliui einančių sintezės vienetų sekos nėra vienareikšmiškai geras požymis.

5.3. Žmonių grupės testas

Buvo atliktas testas su žmonių grupe, kurio tikslas buvo palyginti turimus sintezatorius. Buvo lyginami šie sintezatoriai: fonemų, difonų su garsų grupėmis ir difonų su spektru. Grupę sudarė 5 asmenys, kurių gimtoji kalba yra lietuvių. Jų amžius svyruoja nuo 20 iki 50. Iš jų dvi moterys ir trys vyrai. Nei vienas iš jų su balso sinteze nėra dirbę. Jie klausėsi 30 sakinių, kurie yra paimti iš lietuvių bendrinės kalbos, joje nepasitaikė nei asmenvardžių, nei vietovardžių ar kokių kitų

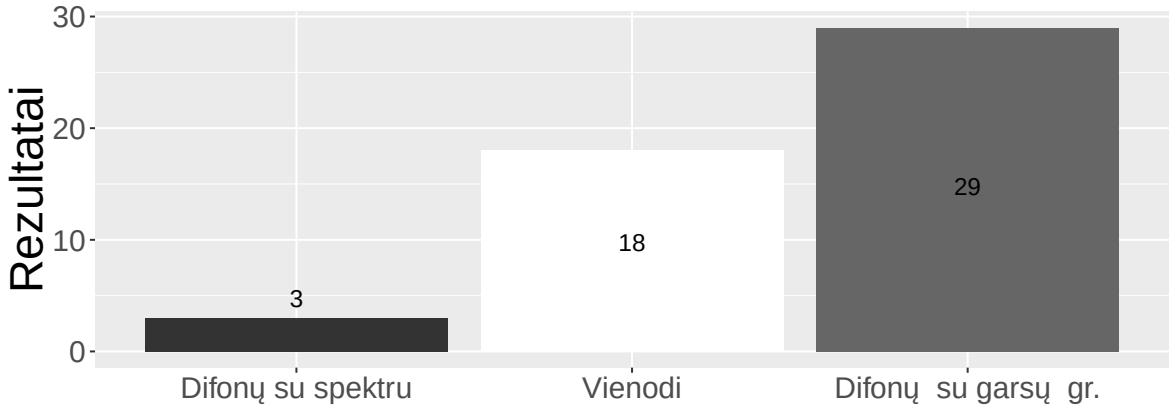
pavadinimų. Vidutinį sakinį sudarė 5 žodžiai. Nors teste dalyvavo tik 5 žmonės, tačiau to turėtų pilnai pakakti.

Taigi tas pats sakinys būdavo perskaitomas du kartus skirtingais sintezatoriais, klausytojų buvo prašoma pasirinkti vieną iš pasirinkimų, sintezatoriai skambėjo panašiai, pirmas skambėjo geriau, antras skambėjo geriau. Pirmuose 10 sakinių buvo lyginami fonemų ir difonų su garsų grupėmis sintezatoriai, kiti 10 sakinių buvo tariami difonų su garsų grupėmis ir difonų su spektru sintezatoriais ir paskutiniai 10 sakinių buvo ištarti difonų su spektru ir fonemų sintezatoriais. Sintezatorių pirmumas buvo atsitiktinai maišomas t. y. pirmas sakinys buvo skaitomas pirmiau fonemų, antras pirmiau difonų, trečias pirmiau fonemų, ketvirtas pirmiau fonemų, penktas pirmiau difonų sintezatoriumi ir taip toliau. Nors eiliškumas buvo atsitiktinai sumaišytas, tačiau kiekvienas sintezatorius vienodai kartų skambėjo pirmas arba antras.



23 pav. Difonų su garsų grupėmis ir fonemų sintezatorių palyginimas

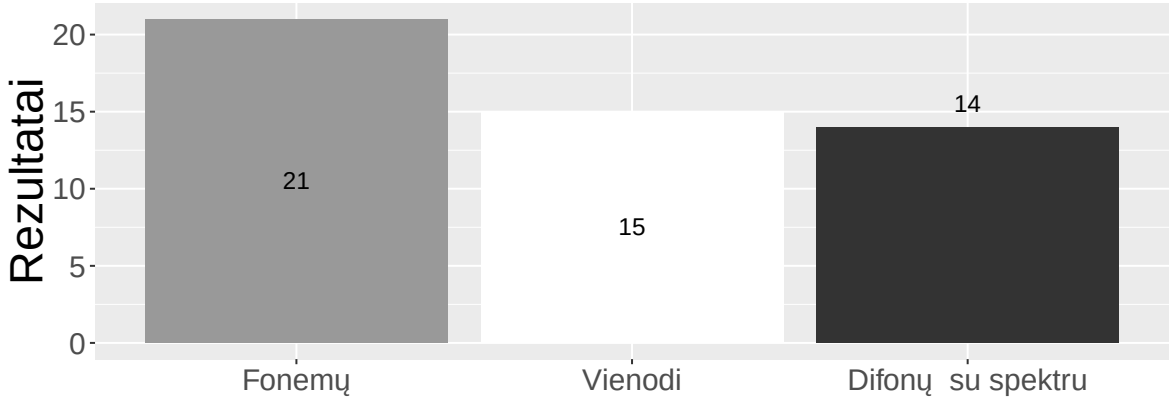
Pirmiausia buvo lyginami difonų su garsų grupėmis ir fonemų sintezatoriai. Gauti rezultatai matyti 23 pav. Į klausimą, kuris sintezatorius skamba geriau buvo atsakyta atitinkamai, difonų sintezatorius (18) nežymiai pasirodė geriau nei foneminis sintezatorius (16), tuo pačiu atsakymas, kad sintezatoriai nuskambėjo vienodai buvo pasirinktas 16 kartų. Suskaičiavus taškus kiekvienam sintezatoriui ir už atsakymą sintezatoriai nuskambėjo vienodai skiriant po pusę taško būtų gaunama, kad sintezatoriai labai panašūs t. y. 26 difonų ir 24 taškų fonemų sintezatoriui.



24 pav. Difonų su garsų grupėmis ir difonų su spektru sintezatorių palyginimas

24 pav. matyti dviejų difoninių sintezatorių palyginimo rezultatai. Akivaizdžiai garsų grupių sintezatorius (29) lenkia spektro sintezatorių (3). Tačiau ir vėlgi nuskambėjo sakinių, kurie

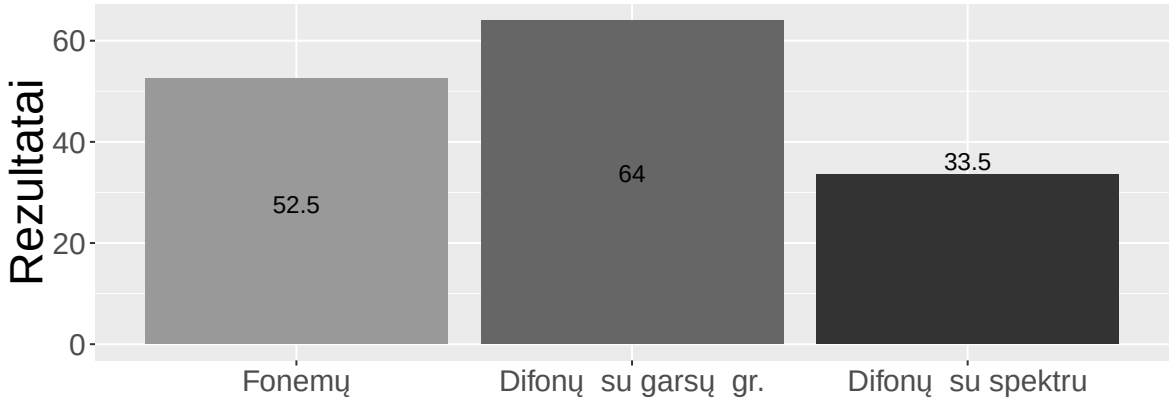
klausytojams nuskambėjo panašiai ir jie pasirinko variantą vienodi (18).



25 pav. Fonemų ir difonų su spektru sintezatorių palyginimas

25 pav. matyti fonemų ir difonų su spektru sintezatorių palyginimo rezultatai. Dar sykį difoninis spektrinis sintezatorius nuskambėjo prasčiausiai (14). Fonemų sintezatorius buvo pasirinktas 21 kartą, o sintezatoriai skambėjo vienodai 15 atvejų.

Apibendrintus rezultatus, kur klausytojui pasirinkus pasirinkimą "vienodi" abiem sintezatoriams būdavo skiriama po 0.5 taško, gaunami rezultatai matomi 26 ir 12 lentelėje.



26 pav. Apibendrinti rezultatai

Iš gautų rezultatų matyti, kad difoninis sintezatorius, kuris jungia difonus pagal spektro skirtumas yra prastas sprendimas ir reikia naudoti difoninį sintezatorių, kuris jungia garsus pagal garsų grupes. Tuo tarpu difoninis sintezatorius pasirodė šiek tiek geriau nei foneminis, tačiau skirtumas nėra didelis.

12 lentelė. Apibendrinti rezultatai

Fonemų	Difonų su spektru	Difonų su garsų grupėmis
52.5	33.5	64

Testo metu buvo atkreiptas dėmesys į tai, kad prastos difonų ar fonemų jungtys yra rečiau pastebimos, nei vieneto parinkimas iš netinkamos aplinkos. Kalbama apie vieneto iš žodžio priekio įdėjimas į žodžio galą, taip pat ir vieneto poziciją sakinyje.

Dar sykį prisimenant, kad teste dalyvavo 5 klausytojai, reikia pastebėti, kad pasitaikė sakinių, kai klausytojai pasirinkdavo visus tris variantus. T.y. kartais tas pats sakinys daliai klausytojų skambėdavo geriau pirmu sintezatoriumi, kitiems jie atrodydavo vienodi, o tretiems antras sintezatorius skambėdavo geriau. Tai nereiškia, kad klausytojai žymėdavo, kas papuola, tiesiog vieniems labiau užkliūdavo intonacijos ar kirčiavimo neatitikimai, kitiems vienetų jungimo neatitikimai.

Išvados ir rekomendacijos

Kalbos sintezatorių poreikis pasaulyje auga. Nors ankščiau balso sintezatoriai buvo naudojami tik neįgaliųjų, sintezuoto balso poreikis gerokai išaugo vystant balso atpažinimo bei asmeninių asistentų technologijas. Išmanūs įrenginiai valdomi balsu, asistentai sugebantys atsakinėti į užduodamus klausimus, įvairių tekstų skaitymas ir kita. Tai tik maža dalis to, kur ateityje bus girdimas sintezuotas balsas. Didžiosios pasaulio IT kompanijos (*Apple*, *Google*, *Microsoft*, *Amazon*) investuoja ir plėtoja teksto-į-kalbą technologijas. Lietuvių kalbos sintezatoriai taip pat vystomi ir tobulinami. Čia didžiulį darbą yra nuveikęs dr. Pijus Kasparaitis, kuris vienaip ar kitaip prisidėjo prie naujausių lietuvių kalbos sintezatorių kūrimo.

Vienetų parinkimo metodas yra pasaulyje gerai žinomas ir naudojamas. Keli lietuvių kalbos sintezatoriai taip pat yra kurti šio metodo pagrindu. Dr. Pijaus Kasparaičio šiuo metodu paremti sintezatoriai naudoja lingvistinius vienetus – fonemas. Tuo tarpu Huang ir kiti [3] pastebi, kad naudojant difonus vienetų jungimas turėtų būti natūralesnis. Būtent tokį sintezatorių sukurti buvo šio darbo tikslas.

Konstruojant difonų bazę iš foneminės bazės buvo nustatyta, kad įrašas, kuris yra pakankamas foneminiam sintezatoriui, visgi yra kiek per trumpas difoniniam sintezatoriui. Duomenų bazė yra per maža, nes ne tik trūksta daugelio teoriškai galimų difonų, tačiau taip pat nėra pakankamai difonų egzempliorių, kurie naudojami sintezėje. Didinti duomenų bazę difoniniam sintezatoriui leidžia ir nedidelis algoritmo sudėtingumas. Tai reiškia smarkiai padidinus duomenų bazę, sintezatoriaus veikimo laikas pakis nežymiai.

Tyrimo eigoje buvo sukonstruoti du difoniniai sintezatoriai su skirtingais difonų jungimo kainų parinkimais. Testo metu buvo nustatyta, kad garsų grupių sintezatorius ne tik skamba geriau nei spektro sintezatorius, bet jis ir veikia greičiau, o tuo pačiu jį dar galiam optimizuoti, priešingai nei spektro sintezatorių.

Šio darbo tikslas buvo palyginti foneminę ir difoninį sintezę lietuvių kalbai. Tai ir buvo įgyvendintai, žmonių grupės teste lyginant difoninius sitnezatorius su foneminiu, buvo užfiksuotas nežymus difoninio su garsų grupėmis sintezatoriaus pranašumas.

Ateities tyrimų gairės

Šio darbo tikslas buvo palyginti lietuvių kalbos sintezę naudojant fonemas ir difonus, tai buvo įgyvendinta šiame tyrime. Kalbant plačiau apie lietuvių kalbos balso sintezės tobulinimą galima išbandyti kitus sintezės metodus, kurie dar nėra bandyti lietuvių kalbai. [1] *Google Deep Mind* programa teigia pasiekę geresnę balso kokybę naudodami formantinius (generuojamas dirbtinis balsas) sintezės metodus. Šie metodai yra bandyti lietuvių kalbai [7], bet tai buvo prieš 20 metų. Galbūt dabar išplėtojus mašininius algoritmus būtų galima pasiekti, kur kas geresnių rezultatų.

Literatūra

[1] Aylett M. 2016. The future of voice synthesis after Google WaveNet debut. *Big Data*, 316-322 p.

[2] Beutnagel M., Conkie A. and Syrdal A. K. 1998. Diphone Synthesis Using Unit Selection. *Third ESCA/COCOSDA Workshop on Speech Synthesis*. 512-518 p.

[3] Huang X., Acero A. and Hon H. 2001. *Spoken language processing: a guide to theory, algorithm and system development*. Prentice-Hall, London.

[4] Yi, J., Glass, J. 2002. Information–theoretic criteria for unit selection synthesis. *Interspeech*, 2617–2620.

[5] Kasparaitis, P. 2005. Diphone Databases for Lithuanian Text-to-Speech Synthesis. *Informatica*, 16(2), 193-202.

[6] Kasparaitis P. and Abinderis T. 2014. Building Text Corpus for Unit Selection Synthesis. *Informatica*, 25(4), 551–562.

[7] Kasparaitis P. 2016. Lietuviškų balso sintezatorių palyginimas. *Kalbų studijos*, 28, 80-91.

[8] Kishore S. P. and Black A. W. 2003. Unit Size in Unit Selection Speech Synthesis. *EUROSPEECH* 1368-1320 p.

[9] Kondrotas A. 2015. *Lietuvių kalbos sintezė vienetų parinkimo metodu*. Baigiamasis bakalauro darbas, Vilniaus Universitetas.

[10] Lietuvių kalbos naujažodžių tartuvas. Projektas LIEPA. Prieiga per internetą: https:// liepa.raštija.lt/Tartuvas/Apie–tartuvą

[11] A. Oord, S. Dieleman, H. Zen ir kiti. 2016. Wavenet: A Generative Model for Raw Audio *arXiv:1609.03499v2*, Cornell University Library

[12] Praat. Prieiga per internetą: http://www.fon.hum.uva.nl/praat/

[13] Projektas LIEPA. Prieiga per internetą: https://www.raštija.lt/liepa

[14] Ravitz J. 2013. I'm the original voice of Siri. Prieiga per internetą: http://edition.cnn.com/2013/10/04/tech/mobile/bennett-siri-iphone-voice/

[15] Rudžionis A. 2001 Pagrindinių kalbos signalų technologijų plėtros ypatumai: sintezė. *Informacinės technologijos 2001* 460-464 p.

[16] Schmidt–Nielsen A. 1995. Intelligibility and Acceptability Testing for Speech Technology. *Applied Speech Technology*, 195–232.

[17] Stakėnas V. 2007. Kodai ir šifrai. Vilnius.

[18] Teksto sintezatorius. Projektas LIEPA. Prieiga per internetą: https://liepa.rastija.lt /Ieškotuvas/Teksto–sintezatorius

[19] Taylor P. 2009. *Text–to–speech synthesis*. Cambridge University Press, Cambridge.

[20] Valdytuvas. Projektas LIEPA. Prieiga per internetą: https://www.raštija.lt/liepa/ paslaugos–vartotojams/valdytuvas/7475