RAAMAT SÜNNIB KUULATA

MEELIS MIHKLA, INDREK HEIN, ANDRUS HIIEPUU,
INDREK KIISSEL, RAIVO RUUSALEPP, URMAS SINISALU

1. Sissejuhatus

Elektronilisel kujul raamatute levik, raamatukogu kirjalike materjalide digiteerimine ja kõnetehnoloogiliste vahendite areng on teinud võimalikuks pakkuda raamatukoguteenuseid ja e-raamatuid sünteeskõnes etteloeetuna ning ilmuvaid ja ilmunud paberile trükitud väljaandeid heliraamatute vormis. Seda teenust pakutakse mitmes veebikeskkonnas. Eesti Rahvusraamatukogu (RR) digitaalarhiivi DIGAR kasutajakeskkonnas on võimalus spetsiaalse audiovaatori abil lasta endale ette lugeda raamatuid, ajakirjaartikleid ja muud elektronilisel kujul eestikeelset trükisõna. Elisa Raamatu rakendus sisaldab ligi 2000 eestikeelset e-raamatut, mida lisaks nutitelefonil või tahvelarvuti ekraanilt lugemisele saab ka kuulata. Eesti Keele Instituudi kõneserveris on avaliku teenusena helindamissüsteem Vox Populi, millega on kõigil võimalus mis tahes tekste helindada: laadides artikli, ettekande, raamatu või mõne muu teksti serverisse üles ja saades vastu helifaili. Kõik need süsteemid on ühistöö¹ vili. Järgnevas artiklis keskendutakse peale süsteemide kirjelduste veel tekstitöötlus- ja hääldusprobleemidele ning tekstiliigenduse esiletoomisele sünteeskõnes. Analüüsitakse, mil määral ja kuivõrd on kõnesüntesaatori sisendtekstis leiduvad lühendid, numbrid ja mittetähelelised ühendid teisendatavad õiges vormis sõnadeks ning võõrnimed lähtekeelega võimalikult sarnaseks häälduseks. Raamatute ettelugemisel peab sünteeskõne ideaaljuhul andma edasi võimalikult täpselt ka tekstiliigendust. Liigenduse optimaalseks markeerimiseks kõnes viidi läbi mitmeid tajuteste; samuti tehti teste selleks, et leida sünteeskõne jaoks optimaalsed akustiliste parameetrite (häälekõrgus, kõnetempo ja kõne valjus) väärtused pealkirjade ja otsese kõne esiletoomiseks.

1.1. Sihtgrupp

Eestlased on tuntud oma suure lugemuse ja hea lugemisoskuse poolest, loeme palju raamatuid ja ajakirjandust. Seda kinnitavad mitmed OECD algatatud PISA-testid ja Euroopa Komisjoni aruanded. Ometi on ka meie ühiskonnas palju inimesi, kellele lugemine ühel või teisel moel raskusi valmistab. Vaeglugejate arv Eestis pole täpselt teada, aga mitmesuguste uurimuste ja PISA-testide tulemuste põhjal on düslektikute määr ühiskonnas üsna täpselt prog-

¹ Ühisprojektides osalesid Eesti Keele Instituut, Eesti Rahvusraamatukogu, Eesti Pimedate Raamatukogu, Eesti Pimedate Liit, Elisa Eesti AS ja Eesti Digiraamatute Keskus. Projekte rahastati riikliku programmi „Eesti keeletehnoloogia 2011–2017” toel.

noositav. Euroopa Komisjoni tellimusel 2012. aastal valminud aruandest (European Commission 2012) selgub, et Eesti on Euroopa Liidus kirjaoskuse tasemelt Soome järel teisel kohal. Aruandes sisalduva PISA-testi tulemused meie lugemisoskuse kohta kõditavad küll eestlaste kui väikerahva edevusnärvi, kuid rõõmustamiseks pole siiski põhjust, sest ka meil on lugemisoskus puudulik 13,3 %-l inimestest. Euroopa Liidus tervikuna puudub igal viiendal 15-aastaselt noorel ja peaaegu 75 miljonil täiskasvanul elementaarne lugemiska kirjutamisoskus (European Commission 2012). Aruande tulemuste põhjal võib heliraamatute ja digiarhiivide helindamise vajadus olla Eestis küllaltki suur – iga seitsmes või kaheksas eestlane kuulub sellesse keelelise erivajaduse sihtrühma.

Teise suure keeleliste erivajadustega sihtrühma moodustavad vaegnägijad ja pimedad. Eestis on umbes 6000 vaegnägijat, kelle nägemislangus ei võimalda igapäevaeluks vajaliku teabe hankimiseks kasutada traditsioonilisi vahendeid. Eesti Pimedate Raamatukogus (EPR) salvestatakse näitlejate abiga aastas ca 160 heliraamatut. Nägemispuudega inimestele heliraamatute kättesaadavaks tegemiseks on EPR-is välja arendatud veebiraamatukogu. Vastavalt autoriõiguse seadusele saavad veebiraamatukogust heliraamatuid kuulata ja alla laadida ainult EPR-i registreeritud lugejad. Eestis ilmub aastas umbes 3500 eestikeelset raamatut. Seega EPR-i praegune suutlikkus raamatuid helindada katab ära 4–5 % Eestis publitseeritavatest raamatutest. Elektroonilisel kujul raamatute levik, arhiveerimine ja kõnetehnoloogiliste vahendite areng on loonud erivajadustega inimestele eeldused suuremal määral osa saada ilmunud kirjandusest ja trükimeediast.

Raamatute kuulamisvõimalus pakub ka tavainimestele alternatiivi olukordades, mil paberramatu lugemine või nutiseadme ekraani jälgimine on võimatu (näiteks autot juhtides, reisides, trenni tehes või koeraga jalutades). Juhul kui raamat jääb huvitava koha peal pooleli, sest kohe on vaja Tallinnast Tartusse sõita, siis Elisa Raamatu rakendusega saaks sõidu ajal pooleli jäänud kohalt kuulamise vormis jätkata (ning sõidu lõppedes võiks omakorda naasta tavapärase lugemise juurde). Kuulamisliides võib abiks olla eakatele inimestele, kelle nägemisteravus on langenud, ja ka lastele, kes ei oska veel piisavalt hästi lugeda. Ka eesti keele võõrkeelena õppijad võiksid helindamis-teenusest huvitatud olla.

1.2. Õiguslikud alused

ÜRO puuetega inimeste õiguste konventsiooni² artikkel 30 punkt 1c kohustab muuhulgas tagama erivajadustega inimestele juurdepääsu raamatukoguteenustele ja trükimeediale. Pimedate, vaegnägijate ja düslektikute jaoks eeldab see spetsiaalset audiovaaturi olemasolu raamatukogu digitaalsete arhiivmaterjalide kodulehel. Hiljuti riigikogus vastu võetud säilituseksemplari seadus (RT I, 07.07.2016, 1) kohustab alates 1. jaanuarist 2017 kirjastusi Eestis esmakordselt avaldatud trükise algmaterjali failina või samaväärse sisuga faili koopiana kohe edastama rahvusraamatukogu elektroonilise loovutussüsteemi kaudu. Elektroonilisel kujul raamatufailide ja heliraamatute genereerimise süsteemi olemasolul on võimalik automaatselt uusi paberversioonis ilmunud

² Eesti ratifitseeris selle riikidevahelise kokkuleppe 21. märtsil 2012 (RT II, 04.04.2012, 6).

trükiseid teisendada heliraamatute vormi. Selline võimalus põhimõtteliselt tagab uute raamatute ja trükiste üheaegse kättesaadavuse nii tavainimestele kui ka keeleliste erivajadustega inimestele heliraamatute vormis ning on vastavuses rahvusvaheliselt tunnustatud võrdse kohtlemise printsiibiga (Eestis kehtib võrdse kohtlemise seadus, vt RT I, 06.07.2012, 22).

1.3. Sünteeskõnes vs. inimkõnes ettelõetud heliraamatud

Sünteeskõnes heliraamatute automaatne genereerimine eeldab raamatute elektroonilisi tekstifaile ja vastava keele süntesaatorit. Eesti Rahvusraamatukogu kogudest on juba praegu väga suur hulk eestikeelset kirjasõna digiteeritud ja need materjalid on elektroonilisel kujul kasutatavad ja töödeldavad. Digiteeritakse isegi tänapäeva inimestele raskesti loetavaid ja mõistetavaid gooti kirjas eestikeelseid trükiseid. Säilituseksemplari seaduse jõustumine garanteerib, et kogu Eestis trükistena ilmuv kirjasõna on elektroonilisel kujul rahvusraamatukogu digitaalarhiivis talle ja põhimõtteliselt keeleliste erivajadustega inimeste jaoks kasutatav.

Käesoleva sajandi esimesel kümnendil on paljude keelte kõnesüntesaatorite väljundkõne kvaliteet jõudnud kasutajate poolt aktsepteeritud tasemele, mis võimaldab neid kasutada mitmesugustes rakendustes. Ka eestikeelseid kõnesüntesaatoreid on viimasel kümnendil Eesti keeletehnoloogia riikliku programmi toel süstemaatiliselt edasi arendatud. Loodud on neljal erineval meetodil põhinevaid süntesaatoreid, mida on juba kasutatud paljudes rakendustes: subtiitrite helindamine, uudiste lugeja nutitelefoni, veebisõnastike helindamisliides jms (Mihkla jt 2012; Mihkla jt 2014).

Sünteeskõnes heliraamatute loomiseks on saadaval nii kommerts- kui ka vabavara (ABC – Audio Book Creator, Dolphin Publisher jt). Sellise tarkvara kasutamist raamatukogudes piirab eelkõige see, et seda pole võimalik paindlikult integreerida olemasolevatesse raamatukogude infosüsteemidesse ja erinevatesse tekst-kõne teisendajatesse, samuti pole osa eestikeelse kõnesünteesi mooduleid neis valmissüsteemides rakendatavad.

Kindlasti ei suuda sünteeskõne lähitulevikus inimkõnega loomulikkuse, meeldivuse ja ilmekuse osas võistelda. Näitlejate ja professionaalsete diktorite esituses heliraamatuid luuakse nüüd ja kindlasti ka edaspidi. Otsest konkurentsi sünteeskõnes heliraamatute poolt pole vaja karta. Küsimus on pigem vahendites ja ressurssides. Praegu näitlejate abiga produtseeritavate heliraamatute arv jääb alla viie protsendi Eestis aasta jooksul ilmuvatest raamatutest. Aga tekst-kõne teisenduse ja heliraamatute genereerimise automaatsed süsteemid loovad võimaluse kõik ilmuvad trükised ja e-raamatud ning neis sisalduv info sünteeskõnes ettelõetuna kätte saada põhimõtteliselt üheaegselt nende trükis ilmumisega. Enim huvipakkuvad teosed (nt need, mida on RR-i kõneserveris kõige rohkem helindada lastud) ilmuvad heliraamatutena näitlejate või diktorite esituses.

Näitlejate ettelõetud heliraamatud pakuvad omakorda huvi ka kõnesünteesi arendajatele. Lugejate ja kuulajate poolt hinnatud raamatute ettelugejate hääli saab edukalt kasutada uute sünteeshääle treenimiseks ja loomiseks (Haag 2011; Prahallad, Black 2011). Ehkki sünteeshääle pole inimhääle üksühene kloon, võib doonorhääle meeldiv variatiivsus rikastada sünteeshääle kõla ja ilmekust.

2. Tekstide helindaja Vox Populi

Tekstide helindaja Vox Populi, Elisa Raamatu rakendus Iselugeja ja DIGAR-i helindamisliides on põhimõtteliselt sarnase funktsiooni ja struktuuriga tekst-kõne teisendajad. Kuna Vox Populi on neist kõige laiemate kasutusvõimaluste ja funktsioonidega, siis artiklis kirjeldatakse üksikasjalikumalt just seda, Eesti Keele Instituudis loodud tekstide helindajat. Vox Populi, rahva hää, koosneb kahest komponendist: toimetaja- ja sünteesiliidesest (vt joonist 1). Toimetajaliideses analüüsitakse sisendteksti ning koostatakse nimekiri arvatavatest võõrnimedest ja tundmatutest märgijadadest. Võõrnimede häälduse andmebaasi *online*-keskkond võimaldab hääldusvasteid lisada ja ka võõrnimede hääldusõigsust kontrollida. Sünteesiliides võimaldab valida erinevate sünteeshäälte vahel ning muuta konkreetse hääle kõnetempot. Nende kahe liidese koostöös teisendatakse tekstifailid sünteeshäälega helifailideks.

Joonis 1. Toimetaja- ja sünteesiliidesega tekstide helindaja Vox Populi.

Tekst-kõne teisendusprotsess on liigendatud hõlpsasti mõistetavateks etappideks. Joonise 2 vasakpoolses klahviveerus on näha teisendusprotsessi erinevad etapid loogilises järjestuses. Esiteks tuleb kasutajal tekstifail serverisse saata, siis tellida selle ettevalmistus kõnesünteesiks, mille käigus erinevad failiformaadid teisendatakse puhtaks tekstiks, tuvastatakse võimalikud erihääldust vajavad sõnad (nt võõrnimed) ning asendatakse lühendid jm mittetekstilised märgijadad loetava tekstiga. Seejärel on soovi korral võimalik muuta sõnade hääldusi, nt *John* asemel *Džon*, *O/Ü* asemel *oo üü*. Kui kõik tundub sobivat, tuleb valida sünteeshää ja lugemiskiirus, käivitada kõnesüntees ja viimaks loodud helifail ka alla laadida. Kui tulemuses midagi kõrva riivab, võib alati hääldusi muuta ja uuesti sünteesida või parandada algteksti ja selle uuesti süsteemile saata. Serveripõhine süsteem töötab autonoomselt, ei pea brauserit lahti jätma ega ootama, millal töö valmib. Kui uuesti sisse logida ja sama fail avada, on süntees kas lõppenud ja allalaadimisklahvid aktiivsed või on olekuteadete kastis kirjas, mitu protsenti tööst on juba valminud. Laias laastus kulub 10 kilobaidi suuruse tavateksti sünteesiks minut, kuid see sõltub serveri koormusest ja veel rohkem valitud lugemiskiirusest (mida suurem kiirus, seda väiksemad on helifailid ja seda kiirem nende salvestamine ning ühendamine).

Joonis 2. Tekstide helindaja Vox Populi teksti sätete näide ekraanitõmmisena.

Heliraamatute automaatse genereerimise oluliseks komponendiks on emakeelne kõnesüntesaator, mille abil sünteesiliides teisendab raamatuteksti ortoeepiliseks kõneks. Eestikeelset kõnesünteesi on arendatud neljal erineval meetodil: formantsüntees ja difoonsüntees ning Markovi peidetud mudelitel ja üksuste valikul põhinev kõnesüntees (Mihkla jt 2012). Tekstide helindamisel kasutab Vox Populi Markovi peidetud mudelitel põhinevaid HTS-sünteeshäält, kuna kõne ladusus ja väljundkõne prosoodia on parim just sellel meetodil. Vox Populi tekstisätetes saab valida tekst-kõne teisenduseks sobivat sünteeshäält (meessünteeshääl Tõnu ning kolm naissünteeshäält Eva, Liisi ja Riina) ja ka teksti ettelugemise kiirust kuue erineva kõnetempo (tavaline, kiirem, väga kiire, jube kiire, aeglane, väga aeglane) seast (vt joonist 2). Tavatempoks peetakse harilikult kõnet, mille kiirus jääb vahemikku 120–140 sõna minutis, heliraamatute soovitavaks kõnetempoks on 150–160 sõna minutis. Sünteeshääl Tõnu räägib ja loeb tekste teistest pisut kiirema tempoga, 165 sõna minutis. Tema kiire ja väga kiire lugemiskiirus on vastavalt 195 ja 244 sõna minutis, mis on tavalisest ca 18 % ja 48 % kiirem. Jube kiire lugemistempo, mis on tavalisest ligi kaks korda kiirem, on mõeldud neile pimedatele, kes suudavad suulist teavet eriti kiiresti omandada. Nimelt on vastavad uuri- mused tõestanud, et osa pimedaid inimesi suudab ajus kõnest arusaamist suu- nata neil muidu jõude seisvasse nägemisinfort töötlevasse ajukoore piirkonda (Moos jt 2008; Hertrich jt 2010). See võimaldab neil hästi aru saada ka kõnest, mis on tavatempost kaks või enam korda kiirem ja mis nägijatele tundub aru- saamatu mulina. Kõnetempo eelistuste uurimisel Eestis täheldasime ka ise, et meiegi pimedate hulgas on sellised eriliste võimetega inimesed täiesti ole- mas (Mihkla jt 2011).

2.1. Tekstitöötlus ja võõrnimede häälde andmebaas

Vox Populi sisendtekst läbib esmalt morfoloogilise analüüsi,³ mille väljundi põhjal tuvastatakse mittetäheleliised ühendid (numbrid, geograafilised koordinaadid, erimärgid jms), lühendid, tundmatud tähejaded ning võõrnimed. Kõnesünteesi jaoks tuleb lühendid ja mittetäheleliised ühendid teisendada loetavateks sõnadeks. Kõnesüntesaator teisendab ka ise numbreid ja erimärke loetavaks tekstiks. Paraku on see lihtsustatud teisendus, mis ei arvesta konteksti, milles lühendid, numbrid ja erimärgid paiknevad, et neid õiges vormis esitada, ega raamatutekstide eripära. Raamatute ettelugemine ja tekstist õigesti arusaamine nõuab oluliselt üksikasjalikumat numbrite, erimärkide, lühendite ja võõrnimede analüüsi ja töötlust.

Numbrijadade töötlusel on kõige keerulisem geograafiliste koordinaatide ettelugemine (nt Tallinna Raekoja platsi geograafilised koordinaadid on *59° 26' 14" N, 24° 44' 43" E* → *viiskümmend üheksa kraadi kakskümmend kuus minutit neliteist sekundit põhjalaiust ja kakskümmend neli kraadi nelikümmend neli minutit nelikümmend kolm sekundit idapikkust*) ning sageli ka sporditulemuste õige esitamine (nt maratonijooksu aegu esitatakse mitmel erineval moel: *2:24.04* või *2.24.04* või *2.24:04* → helindamissüsteem peab kõigil nendel juhtudel häälega teatama aega *kaks tundi kakskümmend neli minutit ja neli sekundit; 4 x 100 m teatejooks*, kus tähemärki *x* loetakse *korda* → *neli korda saja meetri teatejooks*). Ka aastavahemike dešifreerimine on tihti komplitseeritud, nt *2001–2004* võidakse tekstis esitada kui *2001–04* ning süsteem peab aru saama, et see ei tähista kahe tuhande esimese aasta aprillikuud, vaid aastate vahemikku; aastavahemiku ülemine piir võib olla tähistatud ka kolme punktiga *2012–...* → *kahe tuhande kaheteistkümnendast kuni tänaseni*. Grupeeritud numbrijadadest tuleb teatud juhtudel moodustada tervikuid (nt kui suuri arve esitatakse kolmest gruppina *22 555 474* → *kakskümmend kaks miljonit viissada viiskümmend viis tuhat nelisada seitsekümmend neli*), aga samas tuleb ka otsustada, kas lugeda numbrijada arvuna või üksiknumbrite jadana (nt numbrite grupeerimist esineb ka telefoninumbrites (*617 7544*) ja vahel ka pangakontonumbrites (*EE 90 99 00123456789012*), aga neid numbrijadasid pole mõtet grupeerida ega suurte arvudena ette lugeda, vaid ikka numberhaaval). Arvsõna õiget käänat aitab määrata naabersõnade käänete arvestamine (nt *Ehitis valmis 2011. aastal*; morfoanalüüsi väljund annab *aasta+l //_S_ sg ad, //*, seega ka järgarv *2011* on alalütlevas käändes → *kahe tuhande üheteistkümnendal*). Vahel on vormitunnused lisatud numbrijadade külge, kas sidekriipsuga või ilma (*134-ne* → *saja kolmekümne neljane, 100s* → *sajas*). Teatud sõnad (eriti mäarsõnad) numbrite ja arvude ees või järel tingivad konkreetse arvsõna vormi: kui numbrile või arvule eelneb *üle* või järgneb *paiku* või *võrra*, siis tuleb lugeda number või arv omastavas käändes (nt *üle 5 m* → *üle viie meetri, kohtume 3 paiku* → *kohtume kolme paiku, nihkusime 2 koha võrra edasi* → *nihkusime kahe koha võrra edasi*). Kui numbri või arvu ees on sõna *kell*, *neist* või *joobes* (või *joove*), on number nimetavas käändes (*kell 23* → *kell kakskümmend kolm, neist 3 oli katki* → *neist kolm oli katki, tuvasitati joove 1,2 %* → *tuvastati joove üks koma kaks promilli*). Erilist tähelepanu nõuab murdarvude (ehk komaga arvude) käänamine, mis on küllalt keerukas

³ Rakendatakse Filosoofi avatud lähtekoodiga vaba tarkvara Vabamorf, kasutatav ja muudetav ka kommertsrakenduste osana (<https://github.com/Filosoof/vabamorf>).

ja mille puhul kirja pilt pole alati üheselt tõlgendatav (nt 2,3 % võrra → *kahe koma kolme protsendi võrra*, mõlemad arvsõnad on omastavas käändes, mis aga ei kehti arvu 0,3 korral → *null koma kolme* vs. *nulli koma kolme*).

Vox POPULI – muuda oma tekst helifailiks

Uudiseid ei ole, vedelen siin jõude.

Muuda tekstis esinevate sõnade hääldusi

AIDS	<input type="text" value="aids"/>	<input type="button" value="Kuula"/>	<input type="text" value="Lühend, häälda nagu on"/>
Bastille	<input type="text" value="Bastii"/>	<input type="button" value="Kuula"/>	<input type="text" value="Nagu hääldus ütleb"/>
Fere	<input type="text" value="feer"/>	<input type="button" value="Kuula"/>	<input type="text" value="Nagu hääldus ütleb"/>
Hering	<input type="text" value="Hering"/>	<input type="button" value="Kuula"/>	<input type="text" value="Sõna, häälda nagu on"/>
IBM	<input type="text" value="ai bii emm"/>	<input type="button" value="Kuula"/>	<input type="text" value="Inglise lühend, tähthaaval"/>
Manchester	<input type="text" value="Manchester"/>	<input type="button" value="Kuula"/>	<input type="text" value="Kirjaveega, õige kuju hääldusväljal"/>
Pierre	<input type="text" value="P'äär"/>	<input type="button" value="Kuula"/>	<input type="text" value="Nagu hääldus ütleb"/>
Rouge	<input type="text" value="ruuž"/>	<input type="button" value="Kuula"/>	<input type="text" value="Nagu hääldus ütleb"/>
Rudolf	<input type="text" value="Rudolf"/>	<input type="button" value="Kuula"/>	<input type="text" value="Sõna, häälda nagu on"/>
Tracy	<input type="text" value="trassii"/>	<input type="button" value="Kuula"/>	<input type="text" value="Nagu hääldus ütleb"/>
de	<input type="text" value="dõ"/>	<input type="button" value="Kuula"/>	<input type="text" value="Nagu hääldus ütleb"/>
la	<input type="text" value="la"/>	<input type="button" value="Kuula"/>	<input type="text" value="Sõna, häälda nagu on"/>

Hääldused on üle vaadatud, alusta sünteesi

Joonis 3. Väljavõte võõrnimede häälduse andmebaasist (hääldussõnastikust).

Tekstis ja võõrnimedes esineb teatud erimärke, mis omakorda nõuavad spetsiaalset töötlust. Kui plussmärgi + ees ja taga on tühikud, siis hääldatakse see plussina ($a + b = c$), ilmateadetes ($+20^\circ$) jäetakse plussmärk pigem ette lugemata. Et võõrnimesid originaalkeelele lähedaselt hääldada, tuleb osa *unicode*-märke meie alfabeedi aabetegega asendada: tšehhi *č* → *tš*, läti *ā* → *aa*, türki *ı* → *õ* jne. Helindatavas tekstis leiduvatest võõrnimedest ja tundmatutest tähejadadest koostatakse teksti analüüsi käigus nimekiri (vt joonist 3). Selles nimekirjas saavad teksti helindamist tellinud kasutajad võõrnimede ja tundmatutele tähejadadele sobivaid hääldusvasteid lisada ja ka olemasolevaid hääldusi kuuldeliselt üle kontrollida. Nimekirja viimases veerus on konkreetset võõrnime, lühendit või tundmatut tähejada iseloomustav tunnus. Vaikeväärtusena pakutakse välja „nagu hääldus ütleb”, st võõrnime või tähejada hääldatakse andmebaasi hääldusväljal oleva transkriptsiooni järgi (nt *Tracy* → *trassii*, *Fere* → *feer*). Võimalikud valikuvariandid on veel: „sõna hääldada nagu on” (nt *Rudolf* → *Rudolf*, *Hering* → *hering*), st võõrnime või tähejada hääldatakse eesti keele hääldusreeglite järgi; „eesti lühend, tähthaaval” (nt *SA* → *es aa*, mitte *sihtasutus*, kuna *SA*-lühendeid võib tekstides olla erinevaid ja ka asesõna *sa* võib vahel tekstis rõhutamise eesmärgil olla suurte tähtedega kirjutatud; seega on mõistlik loobuda *SA* konkreetsest lühendivastest); „lühend, hääldada nagu on” (nt *AIDS* → *aids*), „inglise lühend, tähthaaval”

(nt *IBM* → *ai bii emm*) või „kirjaveaga, õige kuju hääldusväljal” (nt *Manchester* = *Manchester* → *mantšester*, *Intagram* = *Instagram*). Võõrnimedele lisatakse käändelõpp sageli ülakomaga (nt *Bastille’sse* → *bast’iisse*, *Pierre Rouge’i* → *piäär ruuži*), tekstitöötluses on probleemiks see, et tekstides võidakse kasutada 4–5 erinevat ülakoma. Ka liitvõõrnimede hääldus ja käänamine nõuab erilist tähelepanu (nt *Grande-Rue’s* → *grand rüüs*, *krahv de la-Fere’is* → *döla’feeris*, *proua de Bois-Tracy’iga* → *döbu’aa trass’iiga*). Võõrnimede hääldusbaas täieneb pidevalt, sisaldades juba praegu üle 10 000 võõrnime, lühendi ja märgijada hääldust.

2.2. Tekstiliigenduse markeerimine kõnes

Loetavuse huvides on trükise tekst üldjuhul esitatud kindla struktuuriga. Tekst võib olla jaotatud osadeks, jaotisteks, peatükkideks, jagudeks, lõikudeks, loetelupunktideks jne. Osadel, peatükkidel ja jagudel võivad olla pealkirjad ja veel ka alapealkirjad. Teksti sees on esile toodud tegelaste otsese kõne lõigud, tekst võib sisaldada allmärkusi, tsitaate ja viiteid. Raamatute ettelugemisel peaks sünteeskõne ideaaljuhul suutma edasi anda võimalikult täpselt ka tekstiliigendust. Kõigile tekstiliigenduse finessidele pole alati kõnes võimalik tähelepanu osutada, kuna nad on sageli elektroonilisest tekstist raskesti tuvastatavad. Vox Populi helindamissüsteemis otsustasime keskenduda pealkirja esiletoomisele, otsesele kõnele ja lõigupiiri markeerimisele. Ent ka pealkirjad ja otsese kõne lõigud pole alati üheselt määratavad. Ehkki visuaalselt on pealkirjad tekstis kergesti eristatavad, pole elektroonilisest tekstifailist nende automaatne tuvastamine üldsegi nii lihtne. Pealkirjade lõpus ei ole üldjuhul kirjavahemärke ja nad on tavaliselt eraldi real, aga ka mõned üksikud laused võivad olla eraldi real ja vahel kasutatakse ka pealkirjade lõpus kirjavahemärke. Otsese kõne märgistuseks kasutatakse erikujulisi märke ning samuti esineb tekste, kus otsese kõne esiletoomiseks ei kasutatagi erimärke.

Joonis 4. Hinnangud pauside pikkusele pealkirja ja uudisteksti vahel.

Selleks et teada saada, millised on lugejate-kuulajate eelistused tekstiliigenduse esiletoomiseks kõnes, koostasime viis tajutesti. Testijateks olid nii Pimedate Liidu liikmed kui ka nägijad, kokku osales testides 20 inimest. Kuulajad pidid viiepallisüsteemis hindama vastava tekstiliigenduse esiletoomise sobivust kõnes. Pealkirjade markeerimise uurimiseks kõnes tegime teste lühemate uudislõikudega, millel olid iseloomustavad pealkirjad. Pealkirjadega uudislõike luges ette kõnesüntesaator, varieerisime pealkirja ja põhiteksti vaheliste pauside pikkust ning kõnetempot pealkirja lugemisel. Joonisel 4 on toodud keskmised hinnangud erinevatele pealkirja ja teksti vaheliste pauside kestustele. Optimaalne paus on umbes ühesekundiline, aga pausi kestus sõltub ka pealkirja pikkusest, st pikemal pealkirjal võiks paus põhitekstiga olla natuke pikem. Joonisel 5 on võrreldud pealkirja kõnetempot põhiteksti ettelugemiskiirusega. Uudiste põhiteksti lugemisel oli kõnetempo 1.1, mis vastab lugemiskiirusele 181 sõna minutis. Pealkirjade kõnetempodest kõige kõrgemad keskmised hinded anti variandile 0.94, mis vastab lugemiskiirusele 155 sõna minutis. Seega optimaalne on variant, kui pealkirju loetakse *ca* 15 % aeglasemalt kui põhiteksti.

Joonis 5. Hinnangud pealkirja kõnetempole võrrelduna põhiteksti kõnetempoga (uudisteksti loeti kõnetempoga 1.1 = 181 sõna minutis).

Testisime ka variante, kus pealkirju loeti põhitekstist erineva sünteeshäälega või sama sünteeshäälega, aga pisut erineva tämbriga (konkreetselt hääle erineva väändekoefitsiendi väärtusega). Kõige paremad keskmised hinded sai meeshääle variant väändekoefitsiendiga 0.57, mis kujutab endast sügavamalt häält kumedama tämbriga, selle kõrval on variant 0.53 – kime-dama, porgandihäälega tämber (joonis 6). Loomuliku tämbriga sünteeshääle väändekoefitsiendi väärtus on 0.55. Kuna ükski pealkirja lugemise hääle vahetus ega tämbri muutmise variant silmapaistvalt head hinnangut ei saanud, jäime helindamissüsteemis selle juurde, et üks ja seesama ning moonutamata sünteeshääle loeb nii pealkirju kui ka põhiteksti, pealkirja ja põhiteksti vahel on sekundiline paus ja pealkirju loetakse 15 % aeglasema kõnetempoga kui põhiteksti.

Joonis 6. Hinnangud sellele, kui pealkirju loeti põhitekstist (põhiteksti luges sünteeshääl Tõnu) erinevate sünteeshääldete või sama hääle erinevate tämbritega (hääle väändekoeffitsiendi väärtused: 0.53 – kime tämber, 0.57 – kume tämber).

Otsese kõne markeerimiseks muutsime sünteeshääle parameetreid, et hääle erineva tooni (häälekõrguse ja valjuse) põhjal oleks võimalik eristada otsest kõnet nii saatelausest kui ka muust tekstist. Esmalt muutsime otsese kõne häälekõrgust saatelausest pooltoonide võrra (joonis 7). Parimad keskmised hinnad sai see variant, kus otsese kõne häälekõrgus oli saatelausest 2,5 pooltooni kõrgemal. Teises testiseerias (joonis 8) muutsime otsese kõne ja saatelause valjust (–3 dB, 0 dB, 3 dB, 6 dB). Siin langes parim keskmine hinnang variandile, kus otsene kõne loetakse 3 detsibelli ehk signaali tasandil 1,4 korda valjemalt. Seega testitulemuste põhjal on optimaalne variant, kui otsest kõnet markeeritakse 2,5 pooltooni kõrgema ja 3 dB valjema sünteeshäälega.

Joonis 7. Keskmised hinnangud otsese kõne häälekõrgusele võrreldes saatelause helikõrgusega.

Joonis 8. Keskmised hinnangud otsese kõne valjusele (intensiivsusele) võrreldes saatelause intensiivsusega.

3. Kasutajakeskkond DIGAR, helindamise kõneserverid ja Elisa Raamatu rakendus Iselugeja

DIGAR on Eesti Rahvusraamatukogu kasutajakeskkond, mis pakub juurdepääsu digitaalarhiivis talletatud väljaannetele. Rahvusvahelistele pikaajalise säilitamise standarditele vastav digitaalne arhiiv kogub, säilitab ja teeb kättesaadavaks raamatuid, perioodikat, Eesti avaliku sektori institutsioonide trüki ilmuvate väljaannete trükifaile ja Internetis ilmuvaid võrguväljaandeid. Raamatute, ajakirjade ja muude trükiste lugemisel ja sirvimisel on tekstide helindamiseks vajalik DIGAR-i horisontaalses menüüs valida tekstirežiim, kuna kõnesüntees eeldab sisendis tekstifaili, mitte pildi kujul raamatu lehekülge. Seejärel ilmub valikutesse Kõnesüntesaatori-klahv (vt joonist 9), millele klõpsamisel käivitub nn audiovaatur: Mängi-klahvile vajutamisel saadetakse ekraanil oleva raamatu või ajakirja lehekülje tekst kõnesüntesaatori sisendisse ning mõne hetke pärast kostab arvuti heliväljundis (kõlarites või kõrvaklappides) tekstile vastav helifail. Võimalik on valida ka teksti ettelugemise kiirust (tavaline, aeglane, kiirem). Praegu õnnestub DIGAR-is reaalajas kuulata trükifaile ja e-raamatuid, kus on tagatud 100 % korrektne tekst. Enamik digiteeritud vanemast materjalist, mis on gooti kirjas, ei ole kahjuks täna veel kuulatav, kuna nende tekstide tärgtuvastus (OCR) ei ole piisavalt hea kvaliteediga, et võimaldada vigadeta ettelugemist. Küll aga ollakse jõudmas ajajärku, kus interaktiivne ettelugemine hakkab saama raamatukogu igapäevaseks teenuseks.

Tekstide helindaja Vox Populi ja DIGAR-i helindamisliides realiseeriti kahe eraldiseisva kõneserverina (vt joonist 10). Rahvusraamatukogu tekst-kõne server teenindab eelkõige veebikeskkonna DIGAR jooksvaid tekst-kõne teisendusi. Lisaks saavad Eesti Pimedate Raamatukogu registreeritud kasutajad edaspidi esitada uute ja ka olemasolevate raamatute helindamistellimusi

rahvusraamatukogule. Raamatukoguteenusena paralleelselt on Eesti Keele Instituudis avalikus kasutuses kõneserver Vox Populi, mille abil on võimalik kõigil soovijatel mis tahes eestikeelseid tekste helindada: artikli, ettekande, raamatu või mõne muu teksti faili saab üles laadida ja helifaili vastu saada. Algselt oli plaan ühe kõneserveriga läbi ajada, aga paraku ei või rahvusraamatukogu trükifaili oma süsteemidest välja lubada ning nende serverit ei saa ka juhuslike tekstide helindamisega ilmaasjata koormata. Seetõttu installeeriti kaks sarnast, kuid erinevale kasutajaskonnale mõeldud kõneserverit.

Joonis 9. Kuvatõmmis DIGAR-i kasutajakeskkonna veebilehest.

Joonis 10. Kõneserverid tekstide helindajatenä.

Mõlema kõneserveri funktsioon on selge ja konkreetne: sisendtekst tuleb teisendada väljundkõneks ehk helifailiks (joonis 10). Heliraamatute juures on veel oluline kasutaja toimetajaliides, et oleks võimalik tõlkeraamatutes oleva-

tele võõrnimedele ja tundmatutele lühenditele hääldusvasteid lisada. Esialgu oli plaanis võimaldada ka rahvusraamatukogu DIGAR-i audiovaaturis kasutajatele sellist võõrnimede hääldusvastete lisamist. Aga sellest tuli praktilistel kaalutlustel loobuda, sest on võimatu tagada, et kõik kliendid sihipäraselt seda toimetamisvõimalust kasutaksid, ja süsteemile järelevaatajaks andmebaasi toimetaja ametikoha loomine polnud mõistlik.

Elisa Raamatu rakendus Iselugeja on eestikeelse kõnesünteesi esimene kommertsrakendus, ehkki e-raamatute kuulamisvõimaluse eest lisaraha ei küsita. Ligi 2000 eestikeelset e-raamatut on juba praegu Elisa Raamatu rakenduse kaudu lugejate kasutuses. Iselugeja abil on neid võimalik lasta endale kas meessünteesihäälel (Meelis) või naissünteesihäälel (Pille) ette lugeda (vt joonist 11). Võimalik on valida ka lugemiskiirust kas aeglasema (joonisel 11 olevat liugurit teo pildi poole nihutades) või suurema kõnetempoga (liugurit jänese pildi poole nihutades). Erinevalt Vox Populist või DIGAR-i helindamisliidesest pole Iselugeja serveripõhine rakendus. Elisa Raamatu rakendusega on eestikeelne kõnesüntesaator nutitelefoni või tahvelarvutisse installeeritud. Seega näiteks kui olete enne reisi e-raamatud oma nutiseadmesse alla laadinud, saate lennurežiimis või *offline*'is raamatuid ekraanilt lugeda või lasta endale ette lugeda ja kõrvaklappidega kuulata kogu pika lennureisi vältel.

Joonis 11. E-raamatu lehekülg Elisa Raamatu rakenduses Iselugeja.

4. Kokkuvõte ja edasised tööd

Eestikeelsed kõnesüntesaatorid, tekstide helindajad ja iselugejad rikastavad trükisõna esitlusvormi ning raamatute sünteeskõnes kuulamine avardab inimeste võimalusi saada osa kirjalikust teabest. Ehkki raamatute sünteeskõnes ettelugemine ei pruugi olla päris süütu nähtus, kuna sünteeskõnes ette tulevad hääldus- ja rõhuvead võivad mõjutada ka meie endi keelekasutust, ei saa ometi tehnoloogia arengule kätt ette panna. Kõnesünteesi ja helindamissüsteemide arendamine ja täiustamine toimub pidevalt nii sünteeskõne kvaliteedi ning analüüsi- ja sünteesimeetodite kui ka süsteemide funktsionaalsuse osas, et pakkuda kasutajatele kvaliteetsemat ja usaldusväärsemat teenust.

Kindel on see, et juba olemasolevad lahendused pakuvad suurele hulgale keeleliste erivajadustega inimestele (vaegnägijad, düslektikud, eakad inimesed ja väikesed lapsed) lisavõimalust osa saada raamatukogu digitaalsele kujule teisendatud paberkogudest ning ilmunud ja ilmuvatest eestikeelsetest raamatutest ja trükistest. Ehkki praegu on raamatukogude kogud heli vormis kättesaadavad vaid Eesti Rahvusraamatukogu ja Eesti Pimedate Raamatukogu vahendusel, on loodud tarkvaralahendused vabavaralised ja rakendatavad ka teistes raamatukogudes, muuseumides või arhiivides. Lähitulevikus tekstide helindajad ja heliraamatute automaatsed genereerijad inimkõnes heliraamatute loomise vajadust veel ei kõrvalda. Enimnõutud ja populaarsemate ilukirjanduslike raamatute ettelugemine näitlejate esituses jätkub ka edaspidi. Võrdse kohtlemise printsiibist lähtuvalt on aga põhimõtteliselt oluline, et uued ilmunud trükised oleksid keeleliste erivajadustega inimestele tavainimestega üheaegselt kättesaadavad.

Elisa Raamatu rakendusega on paljud eestikeelsed ilukirjanduslikud raamatud kolinud riulitest nutiseadmetesse, mis loodetavasti toob järjest enam noori inimesi raamatute lugemise ja kuulamise juurde. Analoogiline protsess on toimumas õpikute ja töövihikutega. Ka need on ranitsatest arvutisse kolimas digiõpikute ja -vihikutena (nt Astra veebikeskkond, mille rakendusele lisatakse õpikutekstide kuulamisvõimalus).

Erilist tähelepanu nõuab audiovormis raamatute väljundformaad. Olemasolevatel tekst-kõne süsteemidel on väljundkõne suunatud kas otse kasutajaseadme heliväljundisse või siis helifailidesse (mp3- või wav-formaadis), millel puudub otsene seos raamatu trükiformaadi lehekülje või e-raamatu ekraanilehega. Multimeedia vormis raamatute väljaandmiseks on kujunemas uus e-raamatute publitseerimisformaad ePub3, mis võimaldab loogiliselt ühendada raamatu tekstid, pildid, sünkroonsed helifailid ja ka videofailid ühtseks meediatervikuks. Sellise ühtse formaadi rakendamisel võime julgesti öelda, et raamatut sünnib mitte ainult lugeda ja kuulata, vaid ka liikuvate piltide kujul vaadata.

Kõnesünteesi ja raamatute helindamise alast arendus- ja uurimistööd on toetanud riiklik programm „Eesti keeletehnoloogia (2011–2017)”, institutsionaalne uurimisteema IUT35-1 „Kõnestiilid, lauseprosoodia ja fonoloogiline varieerumine: kirjeldus, teooria ja modelleerimine” ja Euroopa Regionaalarengu Fond (Eesti-uuringute tippkeskus).

Võrguviited

Astra. www.opik.ee

DIGAR = Rahvusraamatukogu digiarhiiv. <http://www.digar.ee>

Elisa Raamat. <https://www.elisa.ee/et/eraklient/teenused/elisa-raamat>

EPR = Eesti Pimedate Raamatukogu veebiraamatukogu. <http://veebiraamat.like.ee/veebiraamat/>

Vabamorf. Eesti keele morfanalüsaator. <https://github.com/Filosoft/vabamorf>

Vox Populi. <http://heliraamat.eki.ee/voxpathuli/>

Kirjandus

European Commission 2012 = Final report of EU high level group of experts on literacy. Luxembourg: Publications Office of the European Union, 2012. http://ec.europa.eu/dgs/education_culture/repository/education/policy/school/doc/literacy-report_en.pdf (23. I 2017).

Haag, Kathrin 2011. HMM-based Speech Synthesis from Audio Book Data. Master of Science. Speech and Language Processing School of Philosophy, Psychology and Language Sciences. University of Edinburgh.

Hertrich, Ingo, Dietrich, Susanne, Ackermann, Hermann 2011. Cross-modal interactions during perception of audiovisual speech and nonspeech signals: An MRI study. – *Journal of Cognitive Neuroscience*, kd 23, nr 1, lk 221–237.

Mihkla, Meelis, Hein, Indrek, Kiissel, Indrek, Orusaar, Margit, Räpp, Artur 2011. Kõnetempo eelistused ja audiosüsteem nägemispuudega inimestele. – *Keel ja Kirjandus*, nr 5, lk 334–342.

Mihkla, Meelis, Hein, Indrek, Kalvik, Mari-Liis, Kiissel, Indrek, Sirts, Risto, Tamuri, Kairi 2012. Estonian speech synthesis: applications and challenges. – *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference „Dialogue”*. Toim A. E. Kibrik. Moskva: PITY, lk 443–453.

Mihkla, Meelis, Hein, Indrek, Kiissel, Indrek, Räpp, Artur, Sirts, Risto, Valdna, Tanel 2014. A System of Spoken Subtitles for Estonian Television. – *Human Language Technologies – The Baltic Perspective. (Frontiers in Artificial Intelligence and Applications 268.)* Toim Andrius Utkas, Gintarė Grigonytė, Jurgita Kapočiūtė-Dzikienė, Jurgita Vaičenonienė. IOS Press, lk 19–26.

Moos, Anja, Hertrich, Ingo, Dietrich, Susanne, Trouvain, Jürgen, Ackermann, Hermann 2008. Perception of ultra-fast speech by a blind listener – Does he use his visual system? – *Proceedings of the 8th Seminar on Speech Production, ISSP 2008*. Toim Rudolph Sock, Susanne Fuchs, Yves Laprie. INRIA, lk 297–300.

Prahalad, Kishore, Black, Alan W. 2011. Segmentation of monologues in audio books for building synthetic voices. – *IEEE/ACM Transactions on Audio, Speech and Language Processing*, kd 19, nr 5, lk 1444–1449.

Puuetega inimeste õiguste konventsioon ja fakultatiivprotokoll. – *Riigi Teataja II*, 04.04.2012, 6.

Säilituseksemplari seadus. – *Riigi Teataja I*, 07.07.2016, 1.

Võrdse kohtlemise seadus. – *Riigi Teataja I*, 06.07.2012, 22.

Books for listening

Keywords: audio books, speech synthesis, prosodic markers of text structure, phonetic database of foreign names, text-to-speech interface for library materials

The article describes the text-to-speech system Vox Populi, the Iselugeja (Self-Reader) application of Elisa Raamat, and the interactive service of reading aloud library materials in DIGAR user environment. Digitation of written library materials, electronic book circulation and the developments in speech technologies now enable the linguistically challenged – people suffering from visual impairment or dyslexia – to receive library services in the form of texts read aloud in synthetic speech or of audio books instead of paper books (no matter whether published or still in print). Text-to-speech systems consist of two components: editor interface and synthesis interface. The editor interface analyses the input text and makes up a list of supposed foreign words and unidentified character strings. The online environment of the phonetic database of foreign names enables addition of pronunciation equivalents and checking the pronunciation of foreign names. The synthesis interface enables the user to choose between different synthetic voices and to modify the speech rate for the voice selected. Those two interfaces cooperate to convert text files into synthetically voiced audio files. The hub of the system is input text analysis and the processing of abbreviations, digits and special characters. Perception tests have been used to specify how to best reflect the text structure (paragraphs, titles and direct speech) in synthesised speech. The text-to-speech system Vox Populi is available as a public service and Iselugeja is a free additional feature to the application Elisa Raamat. Thus, the devices for listening to audio books and for mediating the digital archives of libraries are obviously taking us to an era where being read to in interactive style is becoming an everyday service to library readers.

Meelis Mihkla (b. 1955), PhD, Institute of the Estonian Language, Senior Researcher, meelis@eki.ee

Indrek Hein (b. 1963), Institute of the Estonian Language, Senior Software Developer, indrek.hein@eki.ee

Andrus Hiiepuu (b. 1966), Elisa Estonia, Head of Private Customer Division and Member of the Board, andrus.hiiepuu@elisa.ee

Indrek Kiissel (b. 1968), Institute of the Estonian Language, Software Developer, indrek@eki.ee

Raivo Ruusalepp (b. 1970), MA, National Library of Estonia, Director of Development; Tallinn University, School of Digital Technologies, Lecturer of Records Management, raivo.ruusalepp@nlib.ee

Urmas Sinisalu (b. 1974), National Library of Estonia, IT Architect, urmas.sinisalu@nlib.ee