

EESTIKEELSE LOOMULIKU KÕNESÜNTEESI POOLE

Meelis Mihkla. Kõne ajalise struktuuri modelleerimine eestikeelsele tekst-kõne sünteesile. Modelling the Temporal Structure of Speech for the Estonian Text-to-Speech Synthesis. Dissertationes linguisticae Universitatis Tartuensis 8. Tartu: Tartu Ülikooli Kirjastus, 2007. 176 lk.

Vaadeldava raamatu põhjal kaitses Meelis Mihkla 15. jaanuaril 2008 Tartu Ülikoolis edukalt doktoritöö. Raamat koosneb kahest osast. Esimeses tutvustab M. Mihkla üldise ja eestikeelse kõnesünteesi alaseid seniseid uurimusi ja oma töö eesmärgi, materjali, meetodeid ja tulemusi. Lõppkokkuvõttes kirjeldab ta ka tulevikuplaane. Teine osa sisaldab kaheksa publikatsiooni, millest viie puhul on M. Mihkla ainus autor, kolme puhul peaautor.

M. Mihkla annab hea ülevaate kõnesünteesi strateegiatest ja tekst-kõnesünteesi ajalise struktuuri mudelitest. Eestikeelse kõne struktuuri ja modelleerimise peatükist ilmneb muu hulgas, et aastatel 1997–2002 loodud eestikeelses tekst-kõne-sünteesis pauside ja piiripikenduste väärtusi ei modelleeritud, vaid need lisati sünteetilisse kõnesse konstantsete suurustena. Eesti keele ajalist struktuuri on aga vaja modelleerida, et jõuda võimalikult loomulikuna kõlava kõnesünteesini, mis on ka M. Mihkla uurimuse eesmärk.

Kõnematerjalina on tarvitatud erinevat liiki tekste: ilukirjandust, uudiseid ja eesti keele foneetilise andmebaasi materjale. Tekstid on lugenud ja salvestanud 27 mees- ja naissoost diktorit. Varasemates foneetika ja kõnesünteesi uurimistöödes tarvitati materjalina nn laboratiivset kõnet, s.t üksikuid sõnu või üksikuid lühikesi lauseid. Nende tulemused ei ole aga kooskõlas loomuliku seotud kõne ja pikemate loetud tekstide

prosoodiaga. M. Mihkla uurimus annab kõnesünteesi jaoks palju uusi ja vajalikke andmeid. Tekib siiski küsimus: miks on materjalina tarvitatud nii palju võrdlemisi lühikesi kõnelõike? Kas ei oleks paremad pikemad tekstid nagu näiteks just need kaks töös kasutatud raadiodiktorite loetud teksti? Pikemate lõikude korral oleks rohkem võimalusi kõrvutada mitmesuguseid nähtusi sama kõneleja puhul, teiseks oleks suurem hulk lõigulõpu pause ja pikendusi, mida üksteisega võrrelda.

Suur tähtsus on M. Mihkla töös statistilistel meetoditel. Kõik kasutatud mudelid on andnud statistiliselt olulisi andmeid, kirjeldades valdavalt osa pauside ja häälikute variatiivsusest. Statistika kasutamine aitab ära hoida suuri vigu ja võimaldab välja tuua ka väikeseid, kuid olulisi erinevusi.

Tabelis (lk 40) on esitatud statistilised hinnangud sisendite kohta, mis on olulised häälikute kestuste modelleerimisel. Tabelist nähtub muu hulgas, et nii foneemi enda kui ka seda ümbritsevate foneemide kestused on olulised, kuid kõnetakti valde on väiksema tähtsusega. M. Mihkla arvates võib see sõltuda asjaolust, et väldet ei saa esitada lineaarse tunnuseks, sest siin on tähtis osa häälikukestuste silpidevahelisel suhtel. Sel puhul võib esiteks vastu väita, et vältetaju puhul on rõhulise silbi tuuma ja kooda kestusel väga palju suurem tähtsus kui ühelgi järgmise silbi häälikul, sellepärast ei saa ka kestuste suhe eriti oluline olla. Teiseks võivad ümbritsevad häälikud vältetajumisele mõju avaldada ka siis, kui nad asetsevad väljaspool kõnetakti või isegi sõna piiri.¹ Mis on oluline taju puhul,

¹ H. Traunmüller, D. Krull, The Effect of Local Speaking Rate on the Perception of Quantity in Estonian. – *Phonetica* 2003, nr 60, lk 204–205, tabel 8, 13.

võib seda olla ka kõnesünteesi modelleerimisel. (Ülalnimetatud kestuste suhe on muidugi oluline veldete kirjeldamisel.)

M. Mihkla töö sihiks on saavutada tehiskõne, mis tunduks inimkõrvale loomulik. Selleks peavad häälikute ja pauside kestused ja nende jaotus ning põhitooni väärtused olema võimalikult sarnased loomuliku kõne vastavate omadustega. M. Mihkla on statistiliste meetodite abil prognoosinud eriti pause, lõpupikendusi ja nende vahekorda, s.t kõne prosoodilise liigenduse ning teksti süntaktilise liigenduse suhet. Iseäranis huvitav on, et tal on õnnestunud eristada eri liiki pause, olgugi et pauside variatiivsus on väga suur. Kahjuks ei olnud see võimalik pausieelsete pikenduste puhul, kuid neil on arvata-vasti eesti keeles väiksem tähtsus. Siin kerkib küsimus: kui sihiks on saavutada inimkõrvale loomulikuna tunduv kõne, mis ei erine märgatavalt inimeste kõnest, miks ei ole siis tehtud

rohkem tajuteste? Nende abil saaks teada, mis tundub loomulik, mis mitte. Akustika ja taju suhe ei ole lihtne, ainult akustiliste andmete põhjal on raske öelda, mis võiks kuulajatele loomulik tunduda.

Meelis Mihkla on saavutanud väga häid tulemusi ka häälikukestuste prognoosimisel. Tähtis on samuti see, et kõnesünteesi puhul on arvesse võetud ka morfoloogiat.

Töö kokkuvõttes on ära märgitud edasised uurimissuunad. On rõhutatud tajutestide vajadust, eriti silmas pidades neid, kes on tekst-kõne-süntesaatori peamised kasutajad: pimedad ja nägemispuudega inimesed. Peale selle on kavas kõne prosoodia teiste külgede – põhitooni ja intensiivsuse – statistiline modelleerimine. Selleks annab vaadeldav töö hea aluse.

DIANA KRULL