



MARKO MUMMI

Aastaid on räägitud,
et kõnetuvastus
on tulevikutehnoloogia.
Kui kaugemale tulevikku
oleme tänaseks jõudnud?
Mida inimene ja arvuti
on seda teed käies
teada saanud?

EINAR MEISTER, TANEL ALUMÄE

KUIDAS arvuti kuulab ja kõneleb

Kõneleva ja kõnet mõistva masina loomiseks peab üht-teist teadma kõnesignaali kohta.

Kõnesignaali olemus

Kõnesignaali võib käsitleda kolmest aspektist. Esiteks läbi kõneorganite tegevuse. Teiseks kõneüksuste akustiliste omaduste alusel. Kolmandaks – häälelaineiga edastatavate hääldusüksuste kuuldelise eristamise ning äratundmise kaudu. Kõigi nende aspektidega tegeleb keeleteaduse haru, mille nimi on foneetika – vastavalt siis artikulaatorne, akustiline ja pertseptiivne foneetika.

Kõnetehnoloogia lähtub kõnesignaalist kui akustilisest, elektriliselt registreeritavast signaalist. Kõnesignaali sünteesiks ja tuvastuseks on siiski vaja uurida akustilise signaali omaduste seoseid nii artikulatsioonile kui ka tajumehhanismidega.

Erinevalt tekstist ei ole kõne kirjeldatav üksikute tähemärkide jadana. Reeglina vastab tähemärgile kindel kõnesegment – häälik, mis on defineeritud kui väikseim kuuldeliselt eristatav artikulaatorsete ja akustiliste omadustega määratletav kõneüksus. Samas sõltuvad hääliku omadused selle lähimatest naabritest. Näiteks kaashäälik /l/ sõnades *kulu* ja *heli* on üsna erineva kõlaga, /m/ sõnas *kalm* on heliline, kuid sõnas *lehm* helitu. Nähtust, mis mõjutab hääliku omadusi sõltuvalt selle ees ja/või järel asuvatest häälikutest, nimetatakse koartikulatsiooniks. Just koartikulatsioon liidab häälikud keeruliseks akustiliseks tervikuks, mille puhul üleminek häälikult häälikule sisaldab sageli rohkem infot kui hääliku muutumatu osa. Kõnesignaali jagamine kindlapiirilisteks üksusteks on seetõttu üsnagi tinglik.

Hääliku vasteks tajutasandil on foneem. See on häälikuetalon – abstraktsioon, mis tegelikus häälduses realiseerub igal üksikjuhul mõnevõrra iseguse kvaliteediga kõnesegmendina (ehk häälikuna). Sõltuvalt eesmärgist kasutatakse kõnesignaali eri omaduste kirjeldamiseks ka suuremaid üksusi, nagu näiteks difoon (kahe hääliku ühend), trifoon (kolme hääliku ühend), silp, rõhutakt, sõna, fraas, lause jm.

Kõnesignaali akustilisi omadusi mõjutavad ka kõnelejust tingitud tegurid, nagu kõneorganite füsioloogia, kõneleja tervislik ja emotsionaalne seisund, tarbitud alkohol ja narkootilised ained jm. Samuti mõjutavad seda ümbritsev keskkond, näiteks taustmüra või -kõne

Just koartikulatsioon liidab häälikud keeruliseks akustiliseks tervikuks, mille puhul üleminek häälikult häälikule sisaldab sageli rohkem infot kui hääliku muutumatu osa.

ning ruumi akustilised omadused, aga ka sidekanali, näiteks mikrofoni omadused, nagu sagedusriba piirangud, mürad jne.

Seega on inimkõne äärmiselt keerukas signaal, mille analüüs, süntees ja tuvastamine on raske ülesanne.

Kõneanalüüs

Kõneanalüüsi eesmärk on leida kõnesignaalist mitu tunnust, mis kirjeldaksid kõneüksuste akustilisi omadusi, aitaksid selgitada kõnetekitamise ja tajumehhanisme ning oleksid kasutatavad tehnoloogilistes süsteemides, näiteks kõne automaatsel sünteesil ja tuvastusel.

Kasutatavad tunnused sõltuvad uurimisülesandest, need võivad olla näiteks segmentaalsed ja suprasegmentaalsed. Segmentaalsed tunnused on mingit

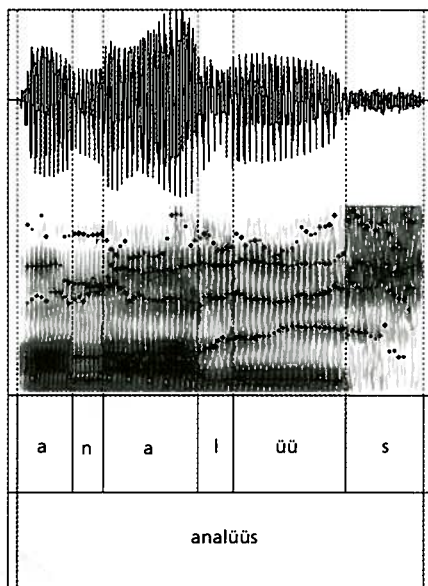
väiksemat kõneüksust – tavaliselt häälikut – iseloomustavad tunnused: hääliku spekter, omakestus, põhitoon jm. Suprasegmentaalsed tunnused, mida sageli nimetatakse ka prosoodilisteks, haaravad mitmeid kõnesegmente (silp, sõna või lause) ja kirjeldavad näiteks kõnetempot, rütmistruktuuri, intonatsiooni.

Kõnesignaali analüüsis rakendatakse mitmesuguseid digitaalset signaalitöötluse meetodeid – spektraalanalüüsi, kepstraalanalüüsi, lineaarprognoosi, formantanalüüsi jpt.

Jooniselt, kus on esitatud sõna „analüüs” helilaine, selle jaotus häälikuteks ja spektrogramm, eristab vilunud silm ka helilaines mitmesuguse iseloomuga häälikuid, kuid oluliselt enam infot annab siiski signaali spektrogrammi analüüs. Spektrogramm on signaali kolmemõõtmeline esitus, kus horisontaaltelje möödik on aeg, vertikaalteljel sagedus ning signaali energiat esitatakse halltoonidena: mida tumedam, seda suurem energia. Spektrogrammilt võime näha, et helienergia ei ole sagedusteljel jaotunud sugugi ühtlaselt, sõltuvalt häälikust on energia kontsentreerunud eri sagedusribadesse. Neid energia lokaalseid maksimume nimetatakse formantideks ja vastavat sagedust – formantsageduseks.

Sisuliselt on formandid kõnetrakti resonantsagedused, mis määravad ära hääliku kvaliteedi, st millist häälikut me

SÕNA „ANALÜÜS” HELILAINE, SELLE JAOTUS HÄÄLIKUTEKS NING SPEKTROGRAMM



Sõna „analüüs” helilaine

Helilaine spektrogramm

Punktijadana on esitatud formantide trajektooriid

Helilaine jaotus häälikuteks

tajume. Traditsiooniliselt piirduakse nelja esimese formandi vaatlusega. Näiteks kui on tegemist meeshäälega, siis on eesti vokaalile /a/ iseloomulikud formantsageduse väärtused järgmised: F1 = 630 Hz, F2 = 950 Hz, F3 = 2500 Hz, F4 = 3100 Hz; vokaalile /i/ vastavad formantsagedused F1 = 260, F2 = 1880, F3 = 2980, F4 = 3400 Hz. Spektrogrammilt näeme ka, et formantide väärtused kõnes ei ole ühe hääliku kestel sugugi püsivad (formantide trajektooriid on esitatud punktjadadena), vaid muutuvad pidevalt.

Formantsagedusi kasutatakse eelkõige siiski kõne eksperimentaalfoneetilistes uuringutes, ent kõnetuvastuse jaoks pole need kuigi sobivad. Kõnetuvastuses rakendatakse valdavalt melkepstri kordajaid – need on kõnespektrit kirjeldavad tunnused, mis leitakse inimkõrva kõnetöötlemehhanisme matkivate signaalitöötlemetoditega.

Kõnesüntees

Tekst-kõne sünteesi eesmärk on teisendada kirjutatud tekst loomuliku kõlaga kõneks. Selleks on vajalikud järgnevad etapid.

- Lingvistiline tekstitöötlus: selle tulemusena teisendatakse kirjutatud tekst hääldustekstiks. Ülesande raskusaste sõltub keelest. See on ilmne, kui võrrelda näiteks inglise ja soome keele kirjapilti ning hääldust.

- Kõneprosoodia modelleerimine: häälikute kestuse ja lausetüübile vastava meloodiakontuuri arvutamine.

- Kõnesignaali tekitamine. Signaali tekitamiseks kasutatava meetodi järgi jagunevad kõnesüntesaatorid artikulatorseteks, formant- ja kompilatiivseteks süntesaatoriteks.

- Artikulatoorne süntees

Artikulatoorne süntees rajaneb kõneproduktiooni füsioloogilisel mudelil ja kõnetraktis hääle tekkimise füüsikalisel kirjeldusel, st selle puhul modelleeritakse inimese kõneorganite tegevust. Artikulatoorseid mudeleid kasutatakse eelkõige kõnelemise mehhanismide uurimisel. Praktiliseks rakenduseks ei ole need kuigi otstarbekad, sest sünteesitava häälelaine arvutamine nõuab suurt arvutusvõimsust ja suvalise teksti süntees reaajas pole sageli võimalik.

- Formantsüntees

Kõnesünteesi formantmudel tugineb kõnesignaali akustilis-foneetilisele kirjeldusele. Elektroonne baasmudel koosneb allikast ja filtrist, kusjuures allikas modelleerib häälekurdude võnkumist ja filter kõnetrakti resonantsagedusi – formante. Nii allika kui filtri omadusi juhib tavaliselt suur hulk keelespetsiifilisi reegleid.

Formantsüntees oli valdav meetod 1970.–1980. aastatel, mil maailmas loodi hulgaliselt rakendusi eri keeltele, sealhulgas ka eestikeelse ja venekeelse kõne süntesaatorid Küberneetika Instituudis. Tänapäeval kasutatakse formantsünteesi mudeleid peamiselt eksperimentaalfoneetilises uurimistöös, aga ka sünteesirakendustes piiratud arvutusvõimsusega seadmetes, näiteks mobiiltelefonides.

- Kompilatiivne süntees

Kompilatiivne süntees rajaneb naturaalkõnest välja lõigatud signaalilõikude, näiteks difoonide, trifoonide, silpide jm sobival ühendamisel. Selleks koostatakse kõnesegmentide andmebaas, millest sünteesi käigus valitakse

Kõnetuvastuse abil teisendab arvuti inimkõne sõnadeks või lauseteks.



MARKO MÜJUM

Arvutile saab selgeks õpetada sõnavara ja lauseehituse, mida mingis valdkonnas kasutatakse. Kui niisugune tekstiarhiiv ja selle lugemisel saadud kõnekorpus on kokku korjatud, võib arvuti kindlal temaatikal uut teksti üles kirjutada suure täpsusega. Labori stuudios käibki parajasti ühe sellise kõnekorpusse salvestust.

sünteesitavale tekstile vastavad segmentid. Need ühendatakse ühtseks lauseks spetsiaalse signaalitötlusalgoritmiga. Tulemuseks on kõrgekvaliteediline sünteeskõne.

Estikeelne kompilatiivmeetodit rakendav kõnesüntesaator töötati Eesti Keele Instituudi, Küberneetika Instituudi ja OÜ Filosoft koostöös välja aastaks 2002. Selles kasutatakse kõnesegmentidena difoone, s.o segmente, mis koosnevad kahest järjestikusest häälikust, nagu konsonant-vokaal, vokaal-konsonant, vokaal-vokaal, konsonant-konsonant, paus-vokaal, paus-konsonant, vokaal-paus, konsonant-paus. Difoonide arv on keeleti erinev, jäädes vahemikku 800–2000. Eesti keele difoonide andmebaas sisaldab ligikaudu 1700 difooni.

Kõnetuvastus

Kõnetuvastuse abil teisendab arvuti inimkõne sõnadeks või lauseteks. Selle tehnoloogia võimaldab luua mitmesuguseid rakendusi, kus kõnetuvastuse väljund võib olla lõpptulemuseks (näiteks dikteerimine) või sisendiks edasisele töötlemisele (näiteks seadmete juhtimine).

Kõnetuvastusteadus on interdistsiplinaarne valdkond, kus kasutatakse arvutiteaduse, signaalitötluse, matemaatika, foneetika ja lingvistika meetodeid. Tänapäeva kõnetuvastusmootorite südameks on statistilised mudelid, mida „treenitakse” suurte kõne- ja tekstikogumite põhjal. Kasutatavad algoritmid ei sõltu keelest ning konkreetse keele kõnetuvastusmootori realiseerimiseks pole praktiliselt vaja põhjalikke keelespetsiifilisi lingvistilisi teadmisi.

• Akustiline mudel

Akustilise mudeli abil modelleeritakse häälikuid. Häälikumudeliga saab modelleerida sõnu ja sõnamudeliga terveid lauseid. Mida see modelleerimine täpsemalt tähendab? Akustilise mudeli põhiomadus on võime öelda suvalise kõnesignaali kohta, kui sarnane on see kõnesignaali „minu” mudelile. Teisisõnu – kui tõenäoliselt see signaal on just „minu” signaal. Akustilise mudeli võimet seda vastavust määrata ei õpetata iga hääliku jaoks eraldi, vaid kõigi häälikumudelite optimaalsed parameetrid arvutatakse välja ühiselt, suurte kõneandmebaaside põhjal.

Kõneandmebaasides on hulk paljude inimeste kõnet koos transkriptsioonidega: iga kõnelõigu kohta on andmebaasis kirjas, mida selles lõigus öeldi.

Selle info abil saab akustilisi mudeleid treenida: seada mudeli parameetrid just sellisteks, et need kõneandmebaasiga kõige paremini sobituksid. Kui andmebaasis on piisav hulk kõnet (reeglina vähemalt kümneid tunde ja sadu kõnelemaid), peaks mudel saama treenimise käigus üldistusvõime ehk sobima ka andmebaasis mitteolevate inimeste kõne tuvastamiseks.

• Keelemudel

Akustilise mudeli abil saab arvutada, kuivõrd üks või teine sõna sarnaneb kasutaja öelduga. See ei ole aga lause rekonstrueerimiseks piisav. Põhjusi on mitmeid: inimesed kipuvad rääkima lohakalt, sõnalõpud muutuvad dünaamiliselt järgmisteks sõnadeks, sõnade vahel tuuakse kuuldavale mitmesuguseid kõhklushäälsusi jne. Samuti on inimeste kõne väga varieeruv: ühe inimese /o/ häälik võib sarnaneda teise inimese /u/ häälikuga jne. Teiseks, isegi kui häälikud ideaalselt tuvastada, on keeruline neid sõnade jadaks teisendada, teadmata midagi keelest, kuna sidusa kõne puhul puuduvad sõnade vahel pausid, mis aitaksid sõnapiire paika panna. Nii näiteks võib häälikute jada /marilähebjalakooli/ tähendada nii lauset „Mari läheb jala kooli” kui ka „Maril äheb ja lagooli”, kuigi viimane ei oma mingit mõtet. Kuidas suudab arvuti seda vahet teha?

Appi tuleb jälle statistika. Kasutades suuri tekstiarhiive ehk korpusi, saame leida, mis sõnad keeles esinevad. See aitab masinal antud näite puhul öelda, et „Mari läheb jala kooli” on ilmselt tõenäolisem vaste öeldule, kuna sõnu „äheb” ja „lagooli” eestikeelsetes teks-

tides tihti ei esine. Tuvastusmootor suudabki tegelikult tuvastada üksnes neid sõnu, mis keelemudeli sõnastikus esinevad, ning loomuliku keele puhul koosneb sõnastik üksnes neist sõnadest, mis esinevad treeningkorpuses piisavalt tihti.

Ainult sõnade loendamine ei aita aga lahendada nn homofoonilisuse probleemi: kas häälikujada /kassaolevalmis/ vastab lausele „Kas sa oled valmis” või „Kassa oled valmis”? Selle kahetimõistetavuse lahendamiseks tuleb uurida, kuidas sõnu omavahel kombineeritakse. Näiteks ilmneb, et sõnakolmik „kas sa oled” on üle saja korra sagedasem kui sõnapaar „kassa oled”. Taoline statistika aitabki arvutil leida paljude akustiliselt sarnaste sõnakombinatsioonide seast loodetavasti selle õige.

Eesti keeles on sõnu väga palju: keele süvastruktuurist mitte midagi teadva arvuti seisukohast on ju ka kõik liitsõnad ning sama sõna käänd- ja pöördvormid unikaalsed sõnad. Seepärast kasutatakse keelemudelis sõnade asemel morfeeme, näiteks sõna „keelemudeliga” lahutatakse kolmeks osaks: „keele”, „mudeli”, „-ga”. Morfeemid liidetakse pärast tuvastamist uuesti sõnadeks, seega toimib see protsess kasutajale märkamatu.

Tuvastusprotsess

Treenitud akustilise mudeli ja keelemudeli ning sõnade hääldussõnastiku abil leiab tuvastusmootor sisendlausele kõige tõenäolisema tuvastusväljundi. Seda protsessi nimetatakse dekodeerimiseks.

Piiratud sõnavaraga tuvastuse puhul on tuvastusprotsessi lihtne ette kujutada. Niisugust tüüpi rakenduse puhul, nagu näiteks mobiiltelefonis kasutatav häälvälimine, arvutatakse sisendsignaali vastavus kõigi telefoniraamatu nimede akustilise mudeli suhtes ning valitakse välja kõige tõenäolisem vaste. Loomuliku ja sidusa kõne tuvastusel on olukord keerulisem, kuid põhimõte on siiski samasugune: kõigi loomuliku keele lausete hulgast tuleb valida välja selline, mis akustilise ja keelemudeli seisukohast on kõige tõenäolisem. Kuna loomulikus keeles on võimalikke lauseid praktiliselt lõpmatu hulk, kasutatakse otsingus keerulisi algoritme, mille abil „lootusetuna” näivad ehk suhteliselt vähetõenäolised harud sellises lausete puus võib varakult kõrvalda, vastasel korral võtaks dekodeerimine lause pikkusega võrreldes sadu

MIS ON MIS ?

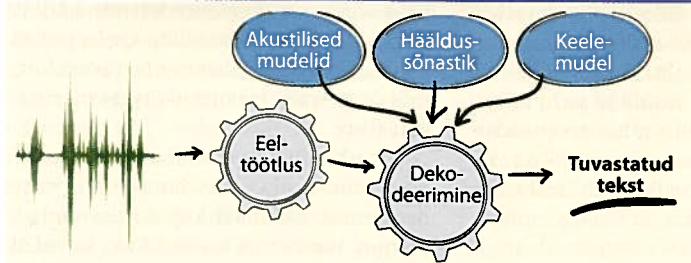
FONEETIKA – keeleteaduse osa, mille uurimisobjekt on inimkõne.

ARTIKULATOORNE FONEETIKA – uurib kõneloomega, s.o artikuleerimise ja foneerimisega seotud protsesse.

AKUSTILINE FONEETIKA – uurib kõneloometegevusega tekitatud häälelained ja selle vahendusel edastatava suulise kõne üksuste akustilisi omadusi

PERTSEPTIIVNE FONEETIKA – uurib häälelained edastatavate hääldusüksuste kuuldelise eristamise ja tajumise (äratundmise) probleeme.

KÕNETUVASTUSSÜSTEEMI STRUKTUUR



TOOMAS PÄÄSUKE

Kõige loomulikum kõnetuvastuse rakendus on kirjade ja muude dokumentide koostamine neid arvutile dikteerides.

või tuhandeid kordi rohkem aega. Sellist otsinguruumi piiramist nimetatakse kärpimiseks (ingl *pruning*) ning kärpimise ulatuse abil saab dekodeerimise kiirust „tuunida”. Paratamatult võib aga juhtuda, et liiga ulatusliku kärpimise korral lõigatakse ära ka mõned niisugused harud, mis hiljem võivad osutada „viljakaiks”, ning seetõttu ei saa seda liiga innukalt rakendada.

Rakendused

Kõige loomulikum kõnetuvastuse rakendus on kirjade ja muude dokumentide koostamine neid arvutile dikteerides. Paljudes keeltes on taoline dikteerimistarkvara juba aastaid küllalt levinud ja töötab päris hästi. Sellist rakendust kasutatakse tihti just spetsiifilistes valdkondades, näiteks radioloogias. Radioloogid diagnoosivad röntgeni-, ultraheli- ja muude piltide põhjal haigusi ja vigastusi. Kuna radioloogi tähelepanu on pööratud uuritavale pildile, on tal tülikas samal ajal klaviatuuri abil diagnoosi sisestada. Kõnetuvastus teeb selle protsessi tõhusamaks.

Dikteerimisele lisaks saab sidusa loomuliku keele tuvastust kasutada paljudes rakendustes, kus kõnetuvastus

kulgeb kasutaja eest varjatult. Näiteks heliarhiivide automaatseks transkribeerimiseks, et võimaldada nende paremat organiseerimist ja indekseerimist. Tänapäeval on paljude raadiote veebilehekülgedel aastatepikkused saadete arhiivid. Ainus võimalus neist midagi otsida on tugineda saate kirjeldusele (kui see üldse olemas on) ning saade lihtsalt läbi kuulata. See on tülikas ja nõuab aega. Palju mugavam oleks ka heliarhiivis otsida infot märksõnade põhjal, nagu otsime mingit sõna tekstidokumentidest.

Kõnetuvastus on ka üks oluline osa sellises mõneti ulmelises rakenduses nagu automaatne kõnest-kõnesse tõlkimine. Sellist rakendust illustreerib joonisel kujutatud olukord, kus jaapani neiu ja eesti noormees suhtlevad tehnoloogia vahendusel teineteise keelt oskamata. See toimub järgmiselt: jaapanikeelne kõnetuvastussüsteem tuvastab jaapani keele ja see tõlgitakse automaatselt eestikeelseks tekstiks, mis loetakse ette eestikeelse süntesaatoriga. Eesti noormehe kõne vahendatakse vastupidise skeemi kohaselt. Samas on võimalik vestluse automaatne tõlge ka mitmesse teise keelde.

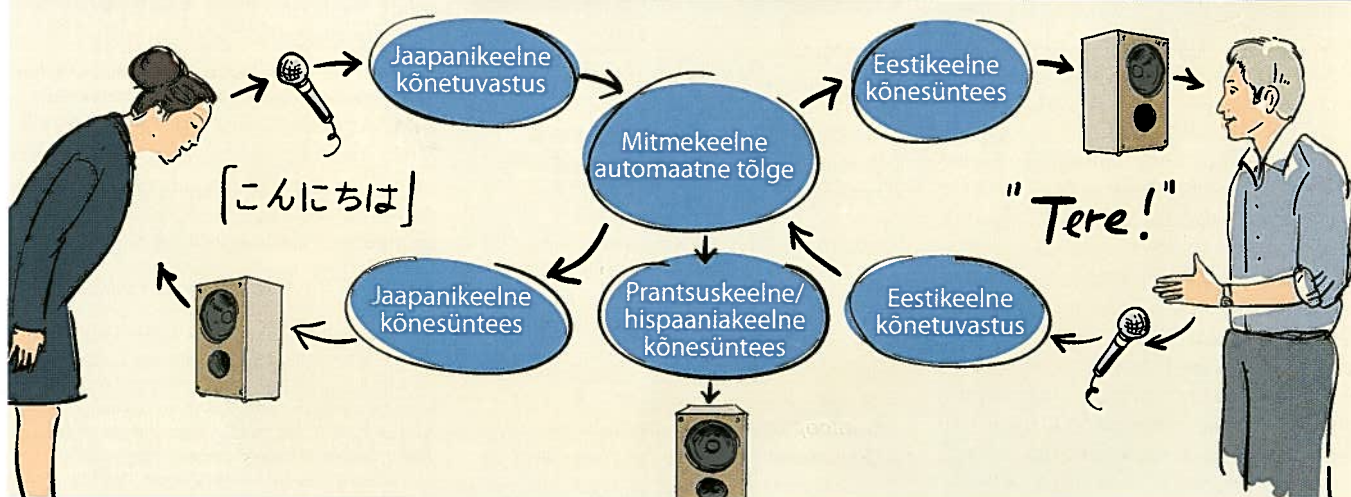
Aastaid on räägitud, et kõnetuvastus on tulevikutehnoloogia ning peagi kaob arvuti klaviatuur ja infot hakatakse sisestama kõne abil. Mingeid märke selle toimumisest on lõpuks märgata: järjest populaarsemaks muutuvad nuti-telefonid, millel klaviatuur puudub, ning näiteks uutel Google'i Android-operatsioonisüsteemiga telefonidel saabki tekstide sisestamiseks kasutada kõnetuvastust. Google on kõnetuvastuse valdkonnas üldse väga aktiivne: ka selle YouTube'i keskkonnas saab tellida videotele automaatselt loodavad subtiitrid. Esialgu toimivad need rakendused küll ainult inglise keele puhul.

Eesti keele tehnoloogiline tugi

Et uued info- ja kommunikatsioonitehnoloogia rakendused oleksid tulevikus kättesaadavad ka emakeelsena, on vaja luua ja arendada eesti keele spetsiifikast lähtuvaid keeletötlusmeetodeid, sest ingliskeelse kõnetuvastuse abil on võimatu tuvastada eestikeelset kõnet.

Keeletehnoloogia on olemuselt valdavalt keespetsiifiline ning seetõttu iga keele puhul ühtemoodi teadmiste-

SUULINE KOMMUNIKATSIOON KÕNETEHNOLOOGIA VAHENDUSEL



TOOMAS PÄÄSUKE



MARKO MUMM

Stuudio mikrofoni ees seisavad Kübl foneetika ja kõnetehnoloogia labori vanemteadur Tanel Alumäe (vasakul) ja labori juhataja, vanemteadur Einar Meister.

Kõneuringud Kübls

Juba mõni aasta pärast Küberneetika Instituudi asutamist alustas toonane automaatikasektori juhataja Eugen Künnap kõnesignaali analüüsi uurimist. Toonane eesmärk oli kasutada inimkõnet automaatsüsteemide juhtimisel. Niisuguse uurimissuuna käivitamiseks sobis Küberneetika Instituut ideaalselt – on ju küberneetika teadusharu, mis uurib juhtimise ja kommunikatsiooni seaduspärasusi nii tehnilistes, looduslikes kui ka sotsiaalsetes süsteemides. Kõne on inimese jaoks kõige loomulikum kommunikatsioonivahend ning selle uurimine ja rakendamine inimese-masina vahelises suhtluses mahub kenasti „küberneetika“ definitsiooni alla.

Teise maailmasõja eelne ja aegne ning järgnenud külma sõja periood olid inimkeele tehnoloogilise töötuse seisukohalt viljakad. 1930. aastatel leiutas Ameerika Ühendriikide firma Bell Laboratories insener Homer Dudley kõne analüüsi-sünteesi süsteemi, mida kasutati edukalt militaarrakendustes. 1940. aastatel loodi samas firmas spektrograaf, mis võimaldas uurida kõnesignaali ajas muutuvat spektrit. Kümnekond aastat hiljem töötati välja mitu kõnesünteesi mudelit ning alustati masintõlkealaseid uuringuid nii Ühendriikides kui Nõukogude Liidus. 1960 avaldas Rootsi teadlane Gunnar Fant kõneproduktiooni akustilise teooria. Need ja teised teaduslikud-tehnoloogilised saavutused, elektronarvutite kasutuselevõtt ning küberneetika rehabiliteerimine Nõukogude Liidus kujundasid ootuse, et kõnesignaal peituvat „salakoodi“ lahtimuukimine on ülesanne, mis lahendatakse kiirelt. Nii arvasid 1960. aastate keskpaiga Nõukogude Liidu juhtivad kõneuurijad, et masinate kõnelema õpetamiseks kulub umbes viis ja kõnetuvastuse loomiseks kümnekond aastat.

Võib arvata, et niisugune optimistlik foon mõjutas ka otsust alustada kõneuringuid Küberneetika Instituudis. Küsida seda tollastelt otsustajatelt kahjuks enam ei saa. Igatahes on tänastel küberneetikust kõneuurijatel põhjust olla tänulik, et see uurimissuund käivitati. Läbi tõusude-mõõnade on kujunenud elujõuline foneetika ja kõnetehnoloogia labor. Uurijate õnneks on kõnesignaal osutunud oluliselt keerukamaks uurimisobjektiks kui 1960. aastatel arvati – kõnesüntees ja -tuvastus sai kasutusküpses alles 20. sajandi viimasel kümnendil, mil keeleteaduse ja infotehnoloogia sümbioosina kujunes välja keeletehnoloogia uurimisvaldkond, mis haarab nii kirjutatud kui ka suulise keele töötlust arvtis. Keeletehnoloogia alamvaldkonda, mis tegeleb suulise keele töötusega, nimetatakse kõnetehnoloogiaks. ●

ja töömahukas ning kallis. Kui suure kõneleajate arvuga keelte keeletehnoloogilist arendust toetab turunõudlus, siis eestikeelne turg on nii väike, et eesti keeletehnoloogia arendus ei saa toimuda turumajanduse reeglite koha-

selt – hinnanguliselt tasub see end majanduslikult ära alates kümnest miljonist kõneleajast. Et eesti keel oleks tehnoloogiliselt jätkusuutlik, on vaja, et keeletehnoloogilist uurimis- ja arendustööd rahastab riik.

Õnneks on seda mõistnud ka Eesti poliitikud ning 2004. aastal valitsuses heaks kiidetud „Eesti keele arendamise strateegia (2004–2010)“ sisaldab ka keeletehnoloogia peatükki. Viimasest lähtuvalt koostati ja käivitati 2006. aastal Eesti Vabariigi Haridus- ja Teadusministeeriumi haldusalas riiklik programm „Eesti keele keeletehnoloogiline tugi (2006–2010)“. Selle ülesanne on uurida ja arendada eesti keele arvutitöötuseks sobivaid tehnoloogilisi lahendusi ning luua uuringuteks ja tehnoloogiaarenduseks vajalikke keeleressursse.

Küberneetika Instituudi foneetika ja kõnetehnoloogia labor osaleb riiklikus programmis kolme projektiga. Nendeks on esiteks „Eestikeelse kõnetuvastuse meetodite uurimine ja arendamine“, kus lisaks teoreetilistele mudelitele käib hetkel töö kahe eespool kirjeldatud rakendusega: kõnetuvastus radio- loogidele ning raadio vestlussaadete automaatne transkribeerimine ja indekseerimine. Teiseks „Kõne analüüs ja variatiivsuse mudelid“, milles uuritakse kõneparameetrite varieeruvust eri kõneleajate ja kõnestiilide korral, sealhulgas ka aktsendiga kõnes. Ja kolmandaks „Kõnekeele ressursid ja kõnetehnoloogia andmebaasid“, kus salvestatakse kõnetehnoloogilisteks ja eksperimentaal-foneetilisteks uuringuteks vajalikke kõnekorpuse, nagu akadeemilised loengud, konverentsiettekanded, raadiouudised, vestlussaadet ja aktsendiga kõne, ning luuakse korpuste haldussüsteem. ●

AUTORITEST

EINAR MEISTER (1957) on Tallinna Tehnikaülikooli Küberneetika Instituudi foneetika ja kõnetehnoloogia labori juhataja, vanemteadur, PhD. Alates aastast 1999 Rahvusvahelise Kõnekommunikatsiooni Assotsiatsiooni (*International Speech Communication Association*) liige. Eesti Vabariigi teaduspreemia 2003 tehnikateaduste alal „Eesti keele tekst-kõne süntees“ koos Meelis Mihkla, Aevo Eegi ja Heiki-Jaan Kaalepiga. Teadustöö põhisuunad: kõnesignaali analüüs, süntees ja tuvastus, kõnelejatuvastus, kõne andmebaasid, eksperimentaal-foneetika, keeletehnoloogia rakendused.

TANEL ALUMÄE (1976) on Tallinna Tehnikaülikooli Küberneetika Instituudi foneetika ja kõnetehnoloogia labori vanemteadur, PhD. Doktorikraadi kaitses 2006. aastal Tallinna Tehnikaülikoolis teemal „Suure sõnavara eestikeelse kõnetuvastuse meetodid“. Peamine uurimisteema jätkuvalt automaatne kõnetuvastus. Täiendanud end muu hulgas Saksamaal Erlangen-Nürnbergi ülikoolis ning järeldoktorantuuris Pariisis ja Helsingis Aalto ülikoolis.