Komputila parolrekono

d-ino KOUTNY Ilona
Lingvistika Katedro de Eötvös Universitato;
Parol-esplora Laboratorio de Budapesta Teknika Universitato

IKU-prelego dum la 75-a UK en Havano 1990

1. Komunikado inter homo kaj masino

Nuntempe la uzo de komputiloj pliĝeneraliĝas preskaŭ en ĉiuj terenoj de la scienco kaj vivo. Tio levas la problemon de natura komunikado inter homo kaj maŝino, do en parolata natura lingvo. La vera dialogo inter homo kaj maŝino inkluzivas komputilajn rekonon kaj generon de parolo, analizon kaj generon de naturlingvaj informoj, logikan "komprenon" de homa demando far komputilo kaj formuladon de respondo perhelpe de la metodoj de artefarita intelekto (fig.1 el Koutny 1988).

Se en la parola komunikada procezo la rolon de ricevanto plenumas komputilo, necesas solvi la taskon de komputila parolrekono. Tio signifas, ke la akustika signalo (parolo) devas esti mapata aŭtomate al simbola reprezentaĵo (skribo). Pri tio ni okupiĝos nun pli detale, sed unue ni vidu, en kiuj terenoj tio utilas!

2. Aplikoj de parolrekono

* Parolo certigas pli naturan rilaton inter homo kaj komputilo (voĉaj ordonoj, dialogaj ekspertaj sistemoj);
* Voĉa direktado de maŝinoj en industria ĉirkaŭaĵo sen mano kaj eventuale sen lumo (regado de robotoj, kvalitkontrolisto kontrolante pecojn parole registras la rezultojn, ktp.);
* Sendiska alvoko de telefonnumero, aŭ informpeto el datumbanko tra telefonlineo aŭ rekte (ekz. operacianta kuracisto petas informojn pri la paciento de la komputilo;
* Aŭtomata skribado de diktata teksto en oficejoj (datumfiksado);
* Malavantaĝuloj povas funkciigi maŝinojn parole;
* Medicinaj, psikologiaj kaj lingvistikaj esploroj;
* Lingvoinstruado.

3. Problemoj de parolrekono

La homa parolo portas *lingvistikajn, socilingvistikajn kaj parolantodependajn informojn.* Tio malfaciligas la rekonon, la malkodigon de la informo. La sama lingvistika informo neniam aperas dufoje en la sama fizika formo. La menciitaj aliĝantaj karakterizoj kaŭzas la t.n. *interpersonan diferencon.* Tio ekzemple malhelpas la parolantosendependan unusencan klasifikon de parolsonoj surbaze de iliaj akustikaj trajtoj kiel la formantoj (ekz. ies /i/ povas tre simili al alies /e/, se oni elprenas ilin el la parolfluo).

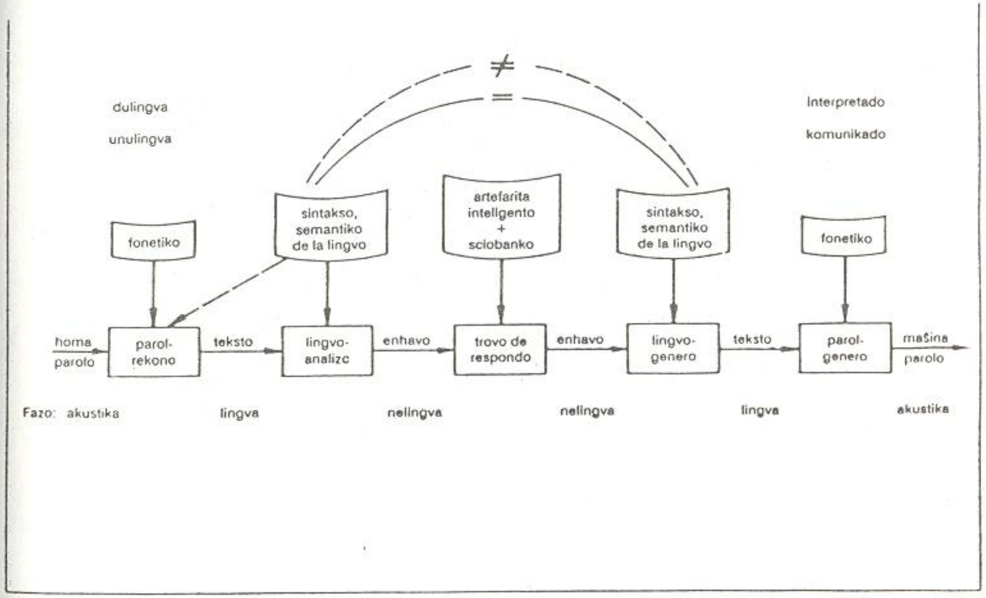


Fig.1. Komunikada procezo inter homo kaj maŝino

La voĉo diferencas laŭ sekso, aĝo kaj la formo de parolorgano Fizike la prononco de la sama frazo diferencas ĉe diversaj persono tamen ni konsideras ilin identaj en lingva nivelo. La voĉo de virin estas pli alta, tiel ankaŭ la virinaj formantoj estas pli altaj, tam la diferenco inter la vira kaj virina voĉoj ne estas konsekvence u el la alia derivebla.

Eĉ la sama persono ne povas prononci ion komplete same, ja ti fizikaj karakterizoj dependas ankaŭ de la san- kaj animstato de parolanto. Tio estas la t.n. *enpersona diferenco*.

Malgraŭ la supre klarigitaj diferencoj ni bone interkompreniĝa eĉ se temas pri telefona interparolo, kie la frekvencobendo estas reduktita al 3400 Hz-oj (la normala homa parolo etendiĝas ĝis almenaŭ ĝis 8000 Hz!). Kiel ni sukcesas kapti la nevarian informon el la varia aperformo? La variojn kompensas la *kunteksto, scio pri la situacio kaj la mondo.* Nuntempe ankoraŭ ne eblas provizi komputilon per tiu ĉi kompleksa scio. *Sciobankoj* povas enteni nur informojn pri specialaj terenoj. Krome mankas la malkovro pri la heŭristika funkciado de la homa cerbo. Komputila sistemo funkcias kutime laŭvice kaj eĉ se paralele, ĝi pritaksas ĉiujn eblojn.

Ni observu la sonogramon de la konata verso *En la mondon venis nova sento...* (fig. 2)! La reala parolfluo estas kontinua, ĝi ne disiĝas al vortoj kaj sonoj, male al la skriba reprezentaĵo. La tasko estas segmenti la parolsignalon kaj mapi ĝin al fonologiaj unuoj, al fonemoj kaj vortoj. La mapadon malhelpas ankaŭ tio, ke eĉ en la sama parolfluo de la sama persono la akustikaj karakterizoj de la unuopaj sonoj ŝanĝiĝas en la funkcio de la ĉirkaŭantaj sonoj (kp *en ve* kaj *se*!) - tio estas la *kunartikulacio*.

Pro la supre menciitaj problemoj la rekono de flua homa parolo per komputilo signifas defion por komputiko. Sed la rekono de aparte, izolite diritaj vortoj estas pli esperiga. Tiaj sistemoj funkcias en la mondo. Ankaŭ la metodoj uzataj ĉi kampe diferencas.

4. Rekono de izolitaj vortoj

La rekono de izolitaj vortoj uzas la teknikon de *formorekono*. La parametroj de la rekonendaj formoj estas storataj - ili estas la referencaj formoj, referencaĵoj, en nia kazo referencaj vortoj. La tasko estas identigi la envenantan formon kun iu el la referencaĵoj, do rekoni ĝin. Tio okazas per formoparigo (A: pattern matching). Tiu ĉi metodo estas vaste uzata kaj ne eluzas la specialecon de lingvaj datumoj, do la rekono estas *lingvosendependa.*

La parolsignalo estas priskribata per sia energispektro kiel funkcio de tempo. La signalo unue trairas analog/diĝitan konvertilon, kie okazas samplado (tipe per 10 KHz). La tempo-funkcio de parolsignalo provizas tro multe da datumoj, pro tio kutime okazas iu transformo por redukti ilin kaj poste sekvas la fazo de identigo. Do la prilaboro de paroldatumoj okazas en la sekvaj paŝoj (vd fig. 3):

1. Redukto de datumoj

Por redukti la paroldatumojn esence du metodoj estas uzataj:

- FFT (Fast Fourier Transformation) filtrado, kiu produktas la energi-distribuon laŭ kanaloj (tipe 10-15 frekvencibendoj) aŭ
- LPC (Linear Predictive Coding).

Post la transformo la tempointervaloj (10-20 ms) de iu parolaĵo estas karakterizataj per 10-15 parametroj.

Figuro 2: Sonogramo de "En la mondon venis nova sento..."

2. Determino de komenco kaj fino

Tre gravas la preciza limigo de parolaĵo. Tio okazas kutime surbaze de la *signalo/bruo rilato*. Oni devas atenti, ke la paŭzo antaŭ plozivoj ne konsideriĝu kiel vortfina paŭzo.

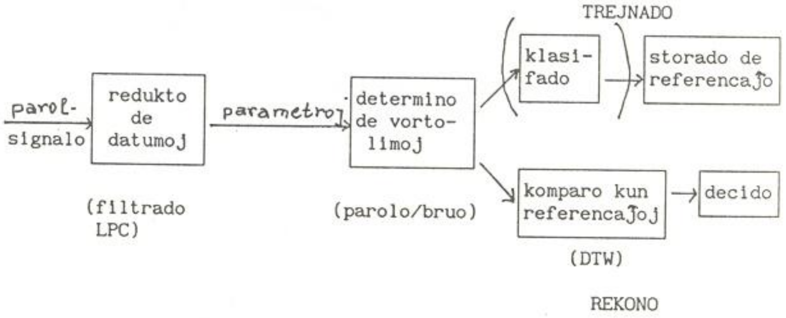
En la trejna fazo por ĉiu vorto la sistemo storas 1-3 prononcojn. En la rekona fazo sekvaj paŝoj estas efektivigataj.

3. Komparo kun referencaĵo

Antaŭ la komparo povas okazi normigo laŭ energio kaj tempo, ja la diversaj prononcoj de la sama vorto ne estas same fortaj kaj longaj. Plej ofte anstataŭ la lineara laŭtempa normigo la metodo de *dinamika tempoaligo* (A: Dinamic Time Warping) estas aplikata. Dum tiu procedo la plej similaj tempointervaloj de la testovorto kaj referenca vorto estas dinamike aligitaj kaj la intervalo-distancoj sumigitaj.

4. Decido

La testovorto estas parigita kun ĉiuj referencaĵoj kaj fine la sistemo elektas kiel rekonitan vorton la *plej proksiman najbaron*, do la referencaĵon kun la plej malgranda distanco de la testovorto. La elektita vorto devas malsuperi iun sojlovaloron; se ne, tiam la testovorto estas rejetita. Tio estas la kazoj de ne permesataj, ne bone prononcitaj aŭ ne bone detektitaj vortoj.



Figuro 3: Rekono de izolitaj vortoj

La plej simpla RIV sistemo estas *parolanto-dependa*, t.e. la sama parolanto devas fari la trejnadon kaj la rekonadon. Tamen pli utila estas la *parolanto-sendependa* sistemo. En tiu ĉi kazo iu ajn povas uzi la sistemon (ce la publika uzo de la sistemo, kiel telefona pridemando de datumbazo).

Ĉe **parolanto-sendependa RIV** la elformo de referencaĵoj estas kompleksa procedo. Reprezentantoj de diversaj voĉkarakteroj devas prononci la vortojn. La vortomodeloj estas elformataj surbaze de diversaj prononcoj; averaĝoj aŭ klasoj reprezentataj per sia centroido estas uzataj kiel referencaĵoj.

En ambaŭ kazoj la rekonendaj vortoj estas de kelkdek ĝis kelkcent. Parolanto-sendependa sistemo funkcias sekure nur malmultaj vortoj. Pluraj sistemoj en la mondo (kiel VoiceScribe de Dragon Systems au Voice Communication de IBM) atingas tre altan rekonprocentajon (super 99 %) kaj tiel ili praktike uzeblas kaj uzatas.

La **rezulto** de la rekono dependas de multaj faktoroj:

- distingebleco de vortoj (similajn vortojn kaj homo kaj maŝino konfuzas, ekz. mano - nano);
- longeco (plifacilas identigi plursilabajn vortojn);
- nombro de vortoj (kutime kun la altiĝo de la nombro de vortoj iom malaltiĝas la fidindenco de la sistemo, sed la reago-tempo pligrandiĝas;
- fonaj kondiĉoj (en brua ĉirkaŭaĵo la rekono estas pli malcerta.

5. Rekono de koneksaj vortoj

En la praktiko ofte necesas la rekono de simplaj frazoj, kies vortprovizo kaj gramatiko estas forte limigita. Ekzemple informpeto pri ekvetur- kaj alvenhoroj de trajnoj aŭ aviadiloj, kaj mendo de biletoj, kie la sekvaj frazoj estas tipaj:

Kiam ekflugas aviadilo al Novjorko?
Mi petas 2 biletojn por la aviadilo al Novjorko je la 16a 50.

Tiel la rekono de koneksaj vortoj povas baziĝi sur la rekono de izolitaj vortoj. La referencaĵoj estas vortoj izolite prononcitaj aŭ prenitaj el flua parolo. Ŝlosilvortoj estas serĉataj per vortopariga metodo kaj konsiderante la sintaksajn limigojn.

6. Rekono de flua parolo

La nombron de rekonendaj vortoj ne eblas senlime altigi, ja ili facile konfuziĝas, la kreo de referencaĵoj estas malagrabla, ilia storado malkonvena kaj la kompartempo multa. Super 1000 vortoj la rekonrezultoj ne estas kontentigaj. Ĉiukaze la normala homa parolo ne disiĝas al vortoj, ĝi estas flua (v. fig. 2), paŭzo aperas nur inter ritmaj unuoj (lingve ili pli malpli koincidas kun sintagmoj), nomataj ankaŭ *prozodiaj vortoj*. Do unuo pli malgranda ol vorto necesas, kiel *fonemo*, *dufonoj* au *silaboj*.

La *segmentadon* al fonemoj malhelpas la kunartikulacio, la modifo de akcento kaj eventuala reduktiĝo de kelkaj sonoj en rapida parolo. Krom la leksika akcento la parolanto povas fokusiĝi ion per akcento. Pro tio la akustika rekono neniam estas perfekta kaj ne sufiĉas por la mapado de parolo al skribo. Necesas pli altnivela analizo de la parolaĵo. La leksika kaj sintaksa analizo povas korekti la erarojn de la akustika rekono. Do la rekono de flua parolo estas *lingvodependa*.

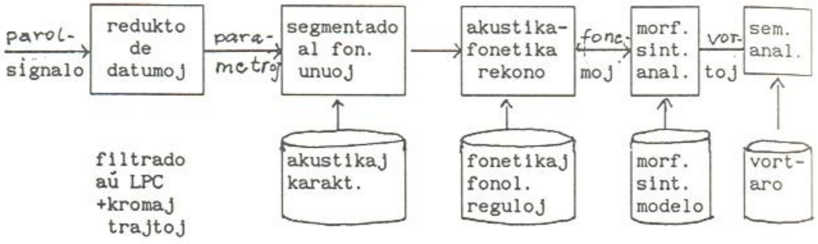
La mapadon malfaciligas ankaŭ la ekzisto de homofonoj, do same prononcataj vortoj, ĉefe en la angla lingvo (ekz. /sʌn/ povas esti skribita kiel *son* au *sun*). Tiel la plena tasko estas:

akustika reprezentaĵo -- fonologia reprezentaĵo -- skriba formo

kiu realigas en pluraj paŝoj (fig. 4):

1. Redukto de datumoj

Same kiel ĉe rekono de izolitaj vortoj tio signifas la mallong-tempan spektron de parolsignalo, filtradon aŭ LPC transformon. Ankaŭ kelkaj aliaj parametroj estas determinataj, kiuj povas esti uzataj ĉe la segmentado (energio, amplitudo, nul-transiroj, formantoj).



Figuro 4: Rekono de flua parolo

2. Segmentado kaj posta akustika-fonetika rekono, la t.n. etikedizo

Tiu ĉi procedo - en la kazo de fonemrekono la aligo de fonemoj al la segmentoj - povas okazi en du etapoj: la unua malfajna (kiu distingas ekz. konsonantajn kaj vokalajn segmentojn) kaj la sekva pli fajna (kiu provas determini la unuopajn sonojn). Oni povas uzi referencaĵojn de spektraj prototipoj.

Fonetikaj kaj fonologiaj reguloj helpas la rekonon: kiuj fonemoj povas kombiniĝi, kiujn alofonojn havas iu fonemo ktp. (ekz. /st/ ne povas aperi en vortokomenca pozicio en la germana, nur /ʃt/, sed en la angla kaj franca inverse; en Eo ambaŭ kombinoj eblas: *stelo*, *ŝtelo*).

3. Morfologia, sintaksa kaj semantika analizoj

Post la akustika-fonetika rekono ekestas vortokandidatoj, kies konsiston la sistemo devas kontroli en la vortaro. La vortaron jam eblas uzi dum la fonetika rekonprocezo por limigi la eblajn sekvajn fonemojn.

La morfologia kaj sintaksa moduloj ekzamenas la estigintan vortoĉenon perhelpe de morfologiaj kaj sintaksaj reguloj, tiel eblas korekti malĝustajn finaĵojn kaj elekti la plej verŝajnan vortoĉenon.

La rekono de la flua natura homa parolo estas esplorata, la funkciantaj sistemoj postulas klaran, distingitan prononcon kaj kutime estas parolanto-dependaj. Se nova parolanto volas uzi la sistemon, li devas per kelkminuta parolo trejni la sistemon.

7. Esperanto kaj parolrekono

1. Rekono de izolitaj vortoj

Kiel ni vidis, la rekono de izolitaj vortoj ne estas vere lingvodependa tasko. Esperanto montras la samajn problemojn, kiel aliaj lingvoj. Ekzistas similaj vortoj, kiuj ne bone distingeblas ĉe la rekono: ekz. *tubo - dubo* aŭ *mano - mamo*. Ankaŭ la similaj finaĵoj kontribuas al tio. Tion konfirmas miaj eksperimentoj en la parolantodependa RIV sistemo VERBIDENT evoluigita en la Budapeŝta Teknika Universitato kaj tiuj de Janot-Giorgetti (Universitato Nancy).

M.-T. Janot Giorgetti (1985) uzis Esperanton en la sistemo MIKROLEA kun 50-200 vortoj kun sufiĉe bonaj rezultoj. Pluraj prononcoj de la sama vorto plibonigis la rezultojn. Ŝi provis apliki la sistemon

en instruado. Tiel krom la bonaj prononcoj ankaŭ la kutimaj prononceraroj de francparolantoj devis esti registritaj (ekz. /glazo/ anstataŭ /glaso/, /kaʒero/ anst. /kajero/).

2. Rekono de flua parolo

Tiu ĉi tasko estas lingvodependa. Ankoraŭ neniu okupiĝis pri la rekono de Esperanto, tamen kelkajn konstatojn ni povas fari. La akustika-fonetika rekono ne povas solvi la taskon, ja Eo havas sufiĉe kompleksan konsonantan sistemon, kiu ebligas la kutimajn konfuzojn (ekz. /s/ - /z/, /ʃ/ - /ʒ/, /k/ - /g/), kaj tre diversaj sonkombinoj eblas (v. /st/ kaj /ʃt/). Maloftas vico de pluraj konsonantoj, la plej ofta silabotipo estas KV. Pro la diversnacieco estas multaj alofonoj en Eo. La vokala sistemo estas 5-membra kaj klare distingebla, tiel la vokaloj povas servi kiel fiksaj punktoj en la rekono.

En la morfologia kaj sintaksa niveloj Eo estas regula, pro tio ĝi havas avantaĝojn. La segmentado al vortoj estas esperiga, ĉar la vortofinoj estas limigitaj.

Bibliografio

F.Fallside, W.A.Woods (red. 1985): Computer Speech Processing. London: Prentice-Hall International

M.-T.Janot-Giorgetti (1985): Parol-rekono kun limigita vortaro, ĝia apliko en la instruado de parolataj lingvoj. En: Koutny I. (red): Perkomputila tekstoprilaboro. Budapest: Scienca Eldona Centro

Koutny I. (1988): Komputila parolgenero. En: Fokuso 1988/4

M.-J.Vigneron (1985): Aŭtomata rekono de la kontinua parolo. En: Koutny I. (red): Perkomputila tekstoprilaboro.