

# GOVOR V KOMUNIKACIJI MED STROJEM IN ČLOVEKOM

DAVOR MILJAN IN JURIJ ŠILC

UDK: 681.3:007

INSTITUT „JOŽEF STEFAN“  
LJUBLJANA

Podan je zgodovinski razvoj in današnje stanje na področju analognih in digitalnih govornih sistemov. Narejen je pregled nekaterih najbolj pogosto uporabljenih metod za sintezo govora iz skupine kodirnikov/dekodirnikov signalne krivulje (PCM, DPCM, ADPCM, DM in CVSD metoda) in iz skupine vokoderjev kot so LPC, formantna in kanalna metoda.

Machine - Man Communication by Voice Historical evolution and state of art of analog and digital speech systems is given. We have reviewed some most used methods for speech synthesis from group of waveform coders (PCM, ADPCM DM and CVSD method) and vocoders (LPC, formant and channel method).

## Uvod

Govor je najnaravnejši način človekovega komuniciranja, zato je razumljiva njegova starodavna želja, da bi strojem, ki jih je zgradil, dodal sposobnost govorne komunikacije. V našem stoletju, dobi telekomunikacij, je ideja o govorni komunikaciji med človekom in strojem pridobila na pomembnosti. V industrijsko razvitem svetu se je začelo obširno teoretično in praktično delo na področju govornih komunikacij, ki je kmalu rodilo prve sadove. Nerazvita tehnologija je bila bistvena omejitev pri doseganju boljših praktičnih rezultatov. Po letu 1978, ko je firma Texas Instruments dala na tržišče prvo ceneno LSI-vezje "Speak & Spell", se je začelo novo obdobje v razvoju govornih komunikacij stroj človek. V naslednjih letih se pojavljajo številni večji in manjši proizvajalci, ki ponujajo nova LSI-vezja, komplete na tiskanih ploščah, programske pakete za različne računalnike in sisteme za generiranje in razpoznavanje govora. Odkrivajo se nova in nova področja v profesionalni in v osebni uporabi govorne komunikacije s strojem, zlasti po razširitvi uporabe mikro in osebnih računalnikov.

Govorno komunikacijo med človekom in strojem delimo na tri osnovna področja: digitalne govorne sisteme, sisteme za razpoznavanje govora in sisteme za razpoznavanje govorca. V tem prispevku se bomo omejili na govorno komunikacijo v smeri stroj človek.

Nekatere osnovne prednosti in pomanjkljivosti govorne komunikacije stroj človek so razvidne iz tabele 1 [1].

### Prednosti

- omogočena komunikacija po telefonu
- uporabnikov pogled ni zaseden
- zelo učinkovito pri informacijah o alarmih ali obvestilih

### Pomanjkljivosti

- govor lahko moti uporabnikovo okolico
- učinki se zmanjšajo v hrupni okolici

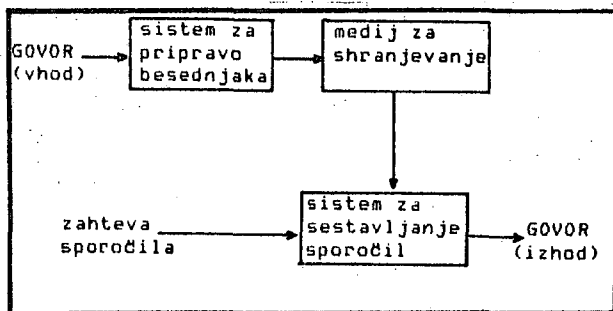
Tabela 1.

## Analogni govorni sistemi

Tisodletja staro željo človeka, da bi izdelal govoreči neživi objekt, so poizkušali uresničiti že stari Grki in Rimljani [2]. Najprej, seveda v verske in mistične namene. Uporabljali so preprosto tehniko in tehnologijo; skrito napeljene cevke in parabolično grajene oboke, ki so prevajali, ojačevali in usmerjali glas skrbno skritega sartrnika.

Razvoj tehnologije je prinesel nove možnosti. Prvi začetki govornih avtomatov segajo v trideseta leta našega stoletja. Najprej so bili to analogni sistemi, pri katerih je besednjak predstavljala zbirka analogno posnetih besed naravnega govora [3] (Slika 1). Medij, na katerem je shranjen besednjak, je bil fotografski ali magnetni film, ki je

bil v obliki traku pritrjen na vrtečem se valju. Takšna, bolj ali manj izboljšana je večina še danes delujočih analognih govornih avtomatov. Od leta 1930 so se govorni avtomati uporabljali v telefoniji za posredovanje podatkov o točnem času, vremenu, pri klicu izklopljene telefonske številke in podobno ter pri različnih javnih avtomatih.



Slika 1: Analogni govorni sistem.

Izboljšave govornega avtomata so potekale v smeri hitrejšega dostopa do posameznih besed ter optimalnega števila shranjenih besed za čim večje število sestavljenih fraz. Hitrost dostopa se povečuje z večjim številom paralelnih trakov s čitalnimi glavami. Trakovi so razdeljeni na polsekundne segmente tako, da so shranjene besede časovno omejene, daljše besede pa se razdelijo na nekaj segmentov. Očitno je, da se z manjšo vrtilno hitrostjo valja povečuje količina zapisanih besed. Ker se valj vrti relativno počasi, morajo biti nekatere besede zaradi hitrejšega dostopa večkrat zapisane. Opisane izboljšave zahtevajo dodatno natančno krmiljenje večjega števila čitalnih glav po določenem vrstnem redu in časovni shemi. To nalogo opravljajo ustrezni avtomati ali v novejšem času računalniki.

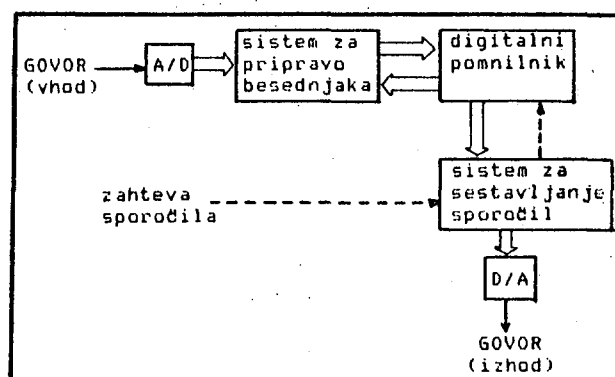
Opisani analogni govorni avtomati se še vedno uporabljajo v primerih, ko besednjak sestavlja manjše število besed in ko se fraze sestavljajo po dovolj enostavnih shemah. Takšni sistemi imajo veliko omejitev. Omenimo naj precejšnjo prilagodljivost sistema, veliko natančnost pri izdelavi in obratovanju mehanskega sistema za reprodukcijo zvoka ter dolgotrajno spreminjanje besednjaka. Še opaznejša je sliha kvaliteta reproduciranih stavkov in fraz, ki jo pri takšnih sistemih le s težavo dosežemo. Fraze sestavljajo poljubne, ločeno posnete besede, zaradi česar nastajajo razlike v nivojih med koncem ene in začetkom naslednje reproducirane besede. Pojavljajo se problemi dinamike in naglas v fraze in podobni pojavi, ki negativno vplivajo na kvaliteto reprodukcije. Takšne napake je mogoče odpraviti z dolgotrajnim, pazljivim snemanjem besednjaka, za kar je potrebna ustrezna oprema. Opisane pomanjkljivosti in omejitve, ki tem sistemom ne omogočajo razvoja v smeri prilagodljivega, bogatega in naravno zvenčega govora, so botrovale razvoju digitalnih govornih sistemov.

### Digitalni govorni sistemi

Postopek, po katerem so zgrajeni digitalni govorni sistemi prikazuje Slika 2. Očitna je velika podobnost z analognim sistemom, ki ga prikazuje Slika 1. Vhodna in izhodna stopnja

digitalnega sistema sta A/D in D/A pretvornika, osrednji del pa je popolnoma digitalen. Digitaliziran naravni govor (digitalno zapisan besednjak) se hrani v digitalnem pomnilniku. Naloga sistema za sestavljanje sporočil je izbiranje digitalno zapisanih enot govora (glas, zlog, beseda, fraza) iz pomnilnika in generiranje poljubnih novih sestavljenih enot (beseda, fraza, stavek).

Postopek digitalnega shranjevanja in generiranja govora, kot je prikazan v splošnem diagramu, dopušča zelo različne postopke načrtovanja digitalnega govornega sistema. Ključni faktor je v načinu digitalnega shranjevanja govornih enot, ki sestavljajo besednjak. Izbira kodirnega postopka, zelo upliva na kvaliteto generiranega govora in velikost digitalnega pomnilnika, ki je potreben za shranjevanje besednjaka.



Slika 2: Digitalni govorni sistem.

Obstajata dva osnovna načina digitalne predstavitve naravnega govora. Prvič, neposredno kodiranje govorne valne oblike in drugič, analiza govorne krivulje, s katero se dobijo parametri, ki opisujejo analiziran govor in so v splošnem parametri za krmiljenje modela govornega trakta.

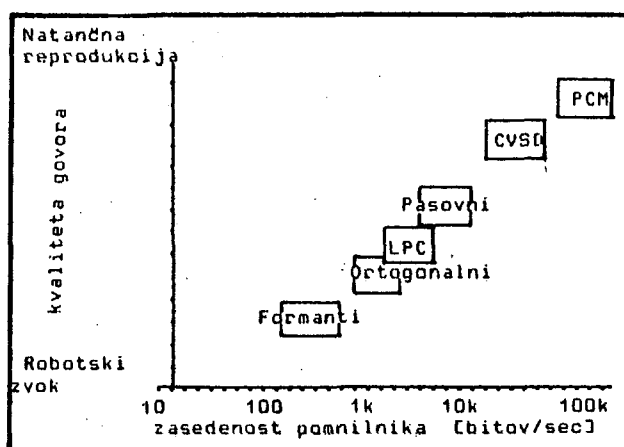
Objektivni kriteriji za vrednotenje digitalnih govornih sistemov še niso institucionalno standardizirani in rabijo le kot smernice načrtovalcem takšnih sistemov. V [32] so podani nekateri merljivi parametri, ki temeljijo na različnih S/N (signal to noise) razmerjih. Ocenjevanje kvalitete govornih naprav je v veliki meri subjektivno in se izvaja na številnih poskusih v obliki poslušanja strojno generiranega govora. Pri tem so v pomoč množice "standardnih" stavkov, ki so tako izbrani, da vsebujejo vse foneme nekega jezika v raznih kombinacijah. Pomembni so trije kriteriji: razumljivost in kvaliteta produciranega govora ter cena sistema.

Elementarni pogoj je razumljivost, ki pa mora biti pri velikem številu aplikacij več kot samo razumljivost. Kvaliteta govora ni v neposredni zvezi z razumljivostjo, je zgolj subjektivna ocena, ki pravi, da je visoko kvaliteten govor najbolj podoben naravnemu govoru, nekvaliteten pa je metalni - robotski govor. V [27] so definirane štiri stopnje kvalitete govornih sistemov:

- radiodifuzna, pri kateri dosežemo frekvenčno območje med 0 Hz in 7 KHz ob minimalnih popačitvah;
- telefonska, v katero sodijo sistemi, ki imajo frekvenčno območje med 200 - 3200 Hz, SNR ≥ 30 dB ter popačitev signala ≤ 2-3 %;
- komunikacijska, ki se ohranja visoko razumljivost ob občutno slabši kvaliteti in večjih popačitvah;

- sintetična, ki jo dobimo pri t.i. vokoderjih ko govorni signal izgublja naravni zvok in deluje metalno-robotsko.

Ko doseže sistem zadovoljivo stopnjo razumljivosti, preostane načrtovalcem še izbira kompromisne rešitve glede cene in kvalitete govora, ob upoštevanju tipa aplikacije. Tako je na primer, pri bančnih avtomatih zaželjen zelo kvaliteten, naravnemu govoru podoben zvok, za razliko od govornih sistemov pri video in igralnih avtomatih, ko metalni zvok ne moti, ali je celo zaželjen.



Slika 3: Kvaliteta govora v odvisnosti od pretoka informacije.

V splošnem lahko rečemo, da je cena govornega sistema sorazmerna ceni pomnilnika. Različni postopki kodiranja ali parametriziranja govora narekujejo različne velikosti pomnilnikov in različne kvalitete govora. Razmerja med temi parametri prikazuje diagram na sliki 3. [1].

### Analiza in sinteza govora

Že prej sta bila omenjena dva osnovna načina delovanja analizatorjev in sintetizatorjev govora. Prvega imenujemo "kodiranje signalne krivulje" in kot že naziv pove, težimo pri tem načinu k čimbolj natančni reprodukciji govornega signala. Neobčutljivost takšnih sistemov na številne govorčeve lastosti in na vpliv šuma ter manjša kompleksnost izvedbe je dosežena z večjim pretokom podatkov (bit rate). Kodirniki signalne krivulje se lahko optimizirajo glede pretoka podatkov ter se ob upoštevanju statističnih rezultatov pri obdelavi govornih vzorcev, še bolj prilagodijo karakteristikam govornega signala.

Drugi način kodiranja temelji na skopem opisu parametrov govora, ki izhajajo iz ugotovitev o tem, kako govor nastaja. Takšne govorne sisteme imenujemo vokoderji (voice coder). Predpostavka je, da je mehanizem nastajanja govora razdeljen na dva ločena dela: izvor zvoka in govorni trakt, ki oblikuje zvok v glas. Nadalje, da izvor zvoka oddaja periodičen zven in šum, ki ga ustvarja zračni tok ter, da govorni trakt lahko simuliramo s spremenljivimi filtri. S številnimi meritvami lahko dosežemo optimalne vrednosti kpnilnih parametrov in s tem tudi dobro kvaliteto govora, kar v praksi ni vedno izvedljivo. Vokoderji običajno generirajo manj kvaliteten (sintetični) govor ob minimalnem pretoku informacij (pod 4.8 Kbitov/s).

Področje med kodirniki signalne krivulje in vokoderji predstavlja prostor za raziskovanje in implementiranje kombiniranih postopkov za generiranje govora z izkoriščanjem dobrih lastnosti obeh omenjenih načinov.

### KODIRNIKI SIGNALNE KRIVULJE

Metode za kodiranje signalne krivulje lahko delimo tudi na časovne in frekvenčne metode. Pri časovnih metodah je govorni signal obravnavan kot enojni signal na celem frekvenčnem območju in se postopki kodiranja razlikujejo glede na stopnjo prilagajanja signalu. Frekvenčne metode delijo signal na nekaj frekvenčnih komponent, ki jih nato ločeno kodirajo. Frekvenčne metode imajo to dobro lastnost, da je število bitov, ki je na razpolago za kodiranje ene frekvenčne komponente deljivo med komponentami glede na potrebe. Tako se frekvenčna komponenta, ki ni zastopana (ali je zastopana pod določeno mejo) sploh ne kodira.

### Časovne metode

Neposredno digitaliziranje in shranjevanje govornega signala je logičen prehod iz analognega na digitalen sistem. Govorni signal se obravnava kot vhodni napetostno spremenljiv signal. Sistem (v našem primeru računalnik) posname govorno krivuljo na ta način, da periodično vzorči napetosti signala s pomočjo A/D pretvornika. Postopek, ki poljubni vrednosti vhodnega signala priredi najbližjo diskretno vrednost imenujemo kvantizacija. Vhodni signal je navzgor frekvenčno omejen z Nyquistovo frekvenco  $f_N$ . Za zagotovitev sprejemljive kvalitete govora mora biti ta frekvenca med 3 in 4 KHz. Hitrost vzorčenja mora biti najmanj  $2 \times f_N$  vzorcev na sekundo (Nyquistova hitrost) [3]. Vsak vzorec ima standardno dolžino 8 bitov, torej  $2^8$  možnih vrednosti s katerimi diskretno opisuje amplitudo vhodnega signala. Opisan postopek imenujemo impulzna modulacija PCM (pulse-code modulation).

Pri kodirnih postopkih se kot merilo, uporablja podatek o pretoku informacije (bit rate). Ta, pri sistemih za prenos podatkov opisuje zasedenost prenosne linije, pri sistemih za shranjevanje podatkov pa opisuje potrebno velikost pomnilnika in je pri PCM metodi enak  $2 \times f_N \times B$ . Zmanjševanje te vrednosti je možno z ožanjem frekvenčnega območja vhodnega signala (govora) in zmanjšanjem števila bitov za posamezni vzorec. Frekvenčno območje ne sme biti pod 3 KHz, ker bi bila s tem močno znižana kvaliteta govora, če je vzorec krajši kot 11 bitov pa se zmanjša učinek kvantizacije. Pri mejnih vrednostih obeh faktorjev znaša pretok informacije okrog 70 Kbitov/s. Primer uporabe PCM metode pri British Telecom govornem sistemu za telefonska obvestila kaže na obsežnost uporabljenega pomnilnika [1]. Besednjak 250-tih besed zaseda približno 6 Mbitov pomnilnika pri pretoku informacije 100 Kbitov/s. Pri tem je potrebno poudariti, da je reproduciran govor kristalno čist, teoretično kvaliteten tako kot digitalni posnetek govora, katerega karakteristike so splošno znane.

Velikost pomnilnika in s tem cena se poizkuša zmanjšati z novimi metodami ali z modifikacijami že znanih metod. Pretok informacije se zmanjša z zgoščanjem podatkov. PCM govorni signal vsebuje veliko odvečne informacije, posebno v področjih, kjer se

krivulja govora počasi spreminja. Zgoščevanje podatkov dosežemo z različnimi postopki kvantizacije govornega signala, ki se med seboj razlikujejo po doseženi stopnji predikcije: časovno spremenljiv korak kvantizacije, ki spremlja časovno spreminjanje vhodnega signala, logaritemska porazdelitev diskretnih stanj (LOG-PCM) ter razne oblike diferencialne kvantizacije (delta modulacija DM, prilagodljiva delta modulacija ADM, diferencialna impulzna modulacija DPCM in prilagodljiva diferencialna impulzna modulacija ADPCM).

Če se pri vzorčenju z Nyquistovo frekvenco shranjujejo podatki, ki pomenijo razliko med amplitudami dveh zaporednih vzorcev govorimo o diferencialni PCM (DPCM).

Izboljšava PCM je tudi prilagodljiva DPCM (ADPCM), ki omogoča dobro kvaliteto govora pri manjšem pretoku informacije. Bistvo metode je dinamično spreminjanje postopka kvantiziranja, ki sledi dinamiki analognega govornega signala. Pri tem se število bitov pri vzorčenju zmanjša iz 12 pri PCM na 3-4 bite. Postopek spreminjanja kvantizacije je prilagojen spremembam govornih krivulj in ni uporaben za ostale vrste signalov. Tudi metoda ADPCM je integrirana v mikrovezjih različnih proizvajalcev [7].

Pri delta modulaciji, kot posebni obliki diferencialne kvantizacije, se analogni vhodni signal vzorči in preoblikuje v enobitno informacijo. Ta opisuje smer spremembe nivoja signala med dvema vzorcema (prvi odvod). Pri reprodukciji se shranjeni vzorci integrirajo, kar sestavi ustrezen približek vhodnega signala [4].

Kvantizacijska napaka je definirana kot razlika med vzorcem izvirnega signala in signala, ki nastane po kvantizaciji in je pri delta modulaciji višja kot pri diferencialni PCM zaradi enobitne informacije ob vzorčenju. Ta vrednost se pri delta modulaciji zmanjšuje z višanjem frekvence vzorčenja, ki v tem primeru ni omejena z Nyquistovo frekvenco. Z dodajanjem integratorja pri kvantizatorju, se lahko spreminja ritem vzorčenja s čimer se doseže boljše spremljanje vhodnega signala. To metodo imenujemo prilagodljiva delta modulacija (adaptive delta modulation ADM).

Z dodatnim integratorjem, množilnikom in nekaj logike dobimo boljšo metodo, ki se imenuje delta modulacija s kontinuirano spremenljivim naklonom (continuously variable slope delta modulation CVSD). Pri tej metodi se zniža pretok informacije na 16 Kbitov/s. Na tržišču obstajajo CVSD koderji / dekoderji v integriranem vezju raznih proizvajalcev kot so: MC3417 (Motorola) [5], FX-209 (Consumer Microcircuits of America) [6] in HC-55516 (Harris Semiconductor).

Metoda zgoščanja govornega signala uporablja kombinacijo opisanih metod z novimi ugotovitvami, ki so izšle iz analize govornih signalov in posebnosti našega dojemanja zvoka. Govorni signal je razdeljen na nekaj karakterističnih frekvenčnih področjih (pitch periods). Pri kodiranju izbira računalnik med krivuljami takšne, ki so simetrične na sredino (zvočne) in shranjuje samo informacije o eni polovici krivulje. Nato poišče krivulje z majhnimi amplitudami v prvi in zadnji četrtini posameznega frekvenčnega območja in te dele nadomesti s "tišino" (half period zeroing). Potopek se konča s kodiranjem vhodnega signala s prilagodljivo delta modulacijo.

Sinteza se začne s tihim področjem, nadaljuje se s četrtino periode govora, ki je shranjen v pomnilniku kot rezultat delta modulacije, ter z reprodukcijo istih podatkov v obratnem vrstnem redu. Zadnja četrtina pa je zopet tiho področje. Narava govornega signala dopušča, brez opazne popačitve zvoka, tri ali štirikratno zapovrstno ponovitev shranjenega izhodnega signala za nezveneče glasove, ali izhodno osemkratno ponovitev za zveneče glasove. Število ponovitev signala, čas trajanja vmesnih tihih področji, izbiro med možko ali žensko višino govora in ostala navodila posredujejo sintetizatorju ukazne besede. Z opisanim postopkom se zmanjša pretok informacije na 1 Kbit/s pri zadovoljivi kvaliteti govora.

### Frekvenčne metode

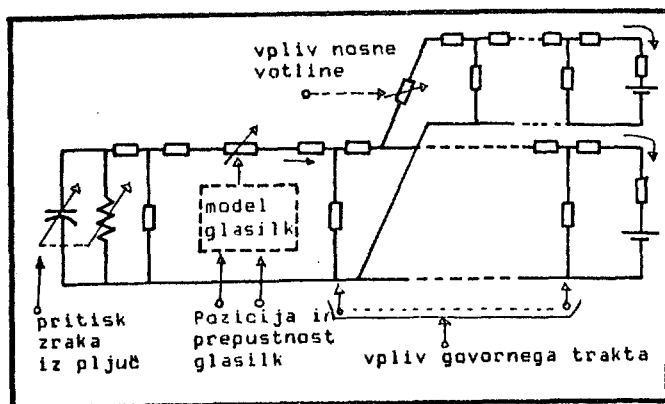
Tako kot pri časovnih metodah, tudi frekvenčne metode se medseboj razlikujejo po kompleksnosti izvedbe in stopnji predikcije. Iz množice različnih metod bomo omenili le dve, pasovno kodiranje in transformacijsko kodiranje.

Pri pasovnih kodirnikih je frekvenčno območje govornega signala, s pomočjo množice pasovnih filtrov, razdeljeno običajno na štiri do osem frekvenčnih pasov. Proces kodiranja se potem nadaljuje ločeno za vsak pas, po eni od zgoraj opisanih metod (npr. APCM). Ob dekodiranju digitalnega zapisa se dekodirani signali v posameznih pasovih združijo v enojni govorni signal. Dobra lastnost opisane metode je v tem, da kvantizacijske napake ostajajo znotraj posameznih pasov in se ne prenašajo (prištevajo) napakam v ostalih pasovih. Obenem ta metoda omogoča zmanjševanje kvantizacijske napake z uporabo različnih kvantizacijskih korakov v posameznih pasovih.

Bolj kompleksna je metoda prilagodljivega transformacijskega kodiranja. Valovna oblika govornega signala se segmentira; nato se nad segmentom signala izvede ena od diskretnih časovno-frekvenčnih transformacij (za govorni signal je najprimernejša diskretna cos transformacija), katere rezultat je množica transformacijskih koeficientov. Koeficienti se potem ločeno kvantizirajo. Rekonstruiranje govornega signala poteka v obratnem vrstnem redu, razumljivo z uporabo inverzne transformacije. Boljše rezultati se dosežejo z upeljavo dodatne informacije o dinamiki govora znotraj posameznega segmenta. Ta, t.i. "stranska informacija", ki predstavlja spektralni nivo znotraj posameznega segmenta, omogoča prilagodljivost opisane metode.

### VOKODERJI

Do sedaj omenjene metode sodijo v skupino metod, ki temeljijo na analizi valovne oblike govornega signala in pri katerih ni pomembno nastajanje temveč dojemanje zvoka. Obstajajo pa tudi metode, katerih osnova so elektronski modeli posameznih delov govornega trakta (grlo, nosni kanali, glasilki, ustna votlina itd ...). Glas nastaja s filtriranjem zvonečih (generator čistega zvona) in nezvenečih (generator šuma) zvokov. Prenosne funkcije filtrov ponazarjajo delovanje posameznih delov govornega trakta, ki so udeleženi pri nastanku glasu (Slika 4).



Slika 4: Digitalni model govornega trakta.

Vokoderji se med seboj razlikujejo po obliki parametrov s katerimi je opisan govorni trakt. Obstajajo različni tipi parametrov kot so: kratkotrajni amplitudni spekter govornega signala na posameznih frekvencah (kanalni vokoderji); koeficienti linearne predikcije, ki opisujejo spektralno ovojnico (LPC vokoderji); podatki o frekvenci ob glavnih resonančnih frekvencah govornega trakta (formantni vokoderji); vzorci kratkotrajnih avtokorelacijskih funkcij govornega signala (avtokorelacijski vokoderji); koeficienti množice ortonormiranih funkcij, ki aproksimirajo govorno krivuljo (ortogonalno funkcijski vokoderji) in številni drugi tipi.

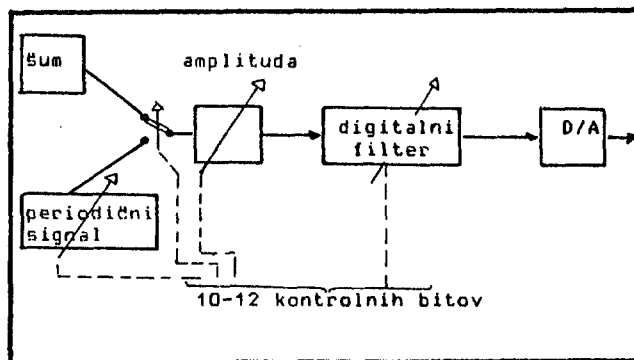
Opisali bomo dva od naštetih tipov. Prvega (LPC vokoder) uvrščamo med časovne, tako kot avtokorelacijske in ortogonalno funkcijske, za razliko od formantnih in kanalnih, ki sodijo med frekvenčne vokoderje.

#### LPC vokoderji

Osnovna ideja linearne predikcije je predpostavka, da če ugotovimo, koliko informacije vsebujejo o  $i$ -tem vzorcu, predhodni vzorci, pričakujemo, da bomo lahko z njimi do določene mere napovedali  $i$ -ti vzorec. Govor se analizira tako, da se analogni govorni signal najprej vzorči s hitrostjo med 8 in 10 KHz. Iz analize vzorcev se določijo vrednosti krmilnih parametrov digitalnega filtra, določa se ali je zvok zvenek ali ni in osnovna frekvenca ter amplituda govora.

Število krmilnih parametrov digitalnega filtra je odvisno od željene kvalitete govora in znaša med 10 pri slabši kvaliteti do 12 pri dobri kvaliteti reproduciranega govora. Pretok informacije je pri tej metodi med 3000 in 1000 bitov/s, kar se doseže z dodatnim krčenjem shranjenih podatkov. V primerih, ko se zahteva samo razumljivost govornega sistema pade pretok informacij med 1000 in 20 bitov/s.

Sinteza govora z LPC metodo je razdeljena na dva osnovna dela: krmiljenje vhodnega signala in krmiljenje digitalnega filtra (Slika 5). Vhodni signal (zvok) ima dva vira. Generator psevdo naključnega šuma omogoča nastajanje nezvenelih glasov. Generator čistega zvena (pravilni periodični signal) omogoča nastajanje zvenelih glasov npr. samoglasnikov. Govor nastane tako, da se vhodni signal (šum ali periodični signal določene frekvence) določene amplitude oblikuje skozi  $N$ -polni LPC digitalni filter, ki predstavlja model dloveškega govornega sistema.



Slika 5: LPC generator govora.

#### Formantni vokoderji

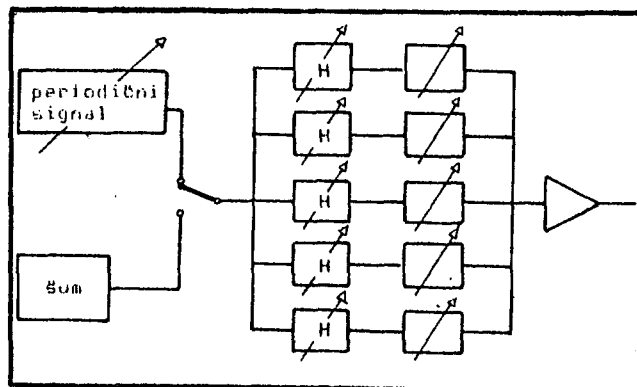
Formantni generatorji govora so po materialni opremlitvi zelo podobni LPC sintetizatorjem (Slika 6). Razlika je v tem, da se pri tej metodi uporabljajo podatki o poziciji središča in pasovnih širin formantnih frekvenc, ki so rezultat predhodne analize govora. Formanti so področja z visoko govorno energijo [3], ki so prisotna na različnih frekvencah govornega signala in so v bistvu resonančne frekvence govornega trakta.

F0	50 ... 250 Hz
F1	200 ... 900 Hz
F2	600 ... 2500 Hz
F3	1500 ... 3000 Hz
F4	3000 ... 4000 Hz
F5	.....
šum	3000 ... >8000 Hz

Tabela 2: Področja formantnih frekvenc za moški glas.

Običajno so tri ali štiri formantne frekvence, ki se gibljejo za moški glas med vrednostmi, ki so prikazane v tabeli 2. Formantne frekvence ženskega glasu so približno 25% višje. Pasovne širine posameznih formantnih frekvenc se določajo algoritmično pri sintezi, ker v splošnem, natančno določanje pasovnih širin ne upliva direktno na kvaliteto generiranega govora.

Opisana metoda omogoča še manjši pretok informacije, ki se giblje med 600 in 800 bitov/s za kvaliteten govor.



Slika 6: Formantni generator govora.

## Zaključek

Posebno področje pri govorni komunikaciji med strojem in človekom predstavljajo sistemi za pretvorbo teksta v govor, ki se pogosto uporabljajo pri govorečih terminalih ali pisalnih strojih in so v veliko pomoč vidno prizadetim osebam. Pri teh sistemih predstavlja posebnost dovolj bogata banka znanja o prevajanju pisanih besed v govor [14, 17]. Običajno se vhodni tekst z računalnikom prevede v foneme ali zloge, ki se potem s pomočjo ene od zgoraj omenjenih metod oblikuje v govor.

V zadnjih letih se pri generiranju govora uporabljajo splošni, ali posebnost za govor zgrajeni, hitri signalni procesorji. Zaradi hitrosti procesiranja omogočajo izvajanje algoritmov s katerimi simulirajo posamezne dele analognih vezij, ki gradijo LPC ali formantne digitalne filtre [18, 19, 20].

Tako LPC kot formantne generatorje govora lahko že nekaj let kupimo na tržišču v obliki integriranih vezij, kompletov integriranih vezij na tiskani plošči ali govornih sistemov. Lahko naštejemo nekaj proizvajalcev: American Microsystems (LPC 1400 bitov/s), General Instruments (LPC vezja in plošče), Centigram (PWC metoda podobna LPC, 4800 bitov/s), EG&G Reticon (signal procesorji), E-Systems (LPC, 2400 bitov/s), Mimic Electronics (vezja in plošče), ITT Semiconductors, Matsushita, National Semiconductor, Nippon Electronic, Sharp, Telesensory Systems, Texas Instruments, Toshiba, Triangle Digital Services (plošče, 650-6500 bitov/s), Votrax, Yahara (govorni sistemi) in mnogi drugi.

Področja uporabe opisanih govornih sistemov so različna. Izbira metode in postopka je odvisna od zahtevane kvalitete in cene izdelka. Znani so govorni sistemi za računalniško generiranje oziroma različnih list, za telefonično sporočanje različnih podatkov in informacij o telefonskih naročnikih in vremenskih podatkih ter podatkih o cenah proizvodov ter za sporočanje podatkov o letalskih poletih. Poleg velikih sistemov obstajajo in še nastajajo neštete manjše aplikacije ob privatnih računalnikih, za alarmna sporočila, za navodila pri uporabi različnih proizvodov in podobno.

V prispevku smo podali pregled zgodovine in sodobnega stanja na področju računalniške sinteze govornega naravnega jezika.

Glede na aktualnost in uporabnost področja (pri tem naj zlasti omenimo generiranje govora iz odprtega teksta/sporočil/podatkov) in glede na dejstvo, da bo moral vsak (vsaj majhen) narod sam poskrbeti za osnovne raziskave - brez katerih ni mogoča kvalitetna aplikacija današnjih tehničnih možnosti - na področju morfoloških značilnosti svojega jezika, je smiselno napovedati naslednji prispevek s sledečo vsebino:

- pregled dosedanjega dela in doseženih rezultatov pri nas na področju matematičnih in tehničnih osnov (materialna in programska oprema) na področju sinteze govora;
- opis pristopa k osnovnim raziskavam posebnosti slovenščine, ki jih moramo obvladati, če naj naši sintetizatorji govorijo slovensko in ne ameriško slovenščino;
- pristop k ustvarjanju ustreznega materialnega in programskega okolja za osnovne in aplikativne raziskave sinteze slovenščine.

## Literatura

- [1] Voice Input/Output sistem and devices, Electronic Engineering, May 1982, page 76 - 101
- [2] James L. Flanagan: Computers that Talk and Listen: Man - Machine Communication by Voice, Proceedings of the IEEE, April 1967, page 405 - 415
- [3] Lawrence R. Rabiner, Ronald W. Schafer: Digital Techniques for Computer Voice Response Implementations and Applications, Proceedings of the IEEE, April 1967, page 416 - 432
- [4] Raymond Steele: Chip delta modulators revive designer's interest, Electronics, October 13, 1977, page 86 - 93
- [5] Earle West: IC trio simplifies speech synthesis, Electronic Design, October 28, 1982, page 175 - 178
- [6] FX-209 Analogue / Digital Converter, Product Information, Consumer Microcircuits Limited
- [7] Steve Garcia: Use ADPCM for Highly Intelligible Speech Synthesis, Byte, June 1983, page 35 - 49
- [8] Cecil H. Coker: A Model of Articulatory Dynamics and Control, Proceedings of the IEEE, April 1967, page 452 - 474
- [9] Computing Power Boosts Speech Quality, Digital Design, May 1982, page 44 - 50
- [10] Gene Helms & Steve Petersen: Portable speech development system creates linear predictive codes, Electronics, September 8, 1982, page 151 - 156
- [11] Speech synthesis: device and applications, Electronic Engineering, January 1981, page 41 - 57
- [12] Roger J. Godina: Voice input output: a special report, Electronics, April 21, 1983, page 126 - 143
- [13] R. Pickvace: Speech synthesis; the new frontiers, Electronic Engineering, Juli 1980, page 37 - 52
- [14] Kathryn Fons and Tim Gargaliano: Articulate Automata: An Overview of Voice Synthesis, Byte, February 1981, page 164 - 187,
- [15] Noriko Umeda: Linguistic Rules for Text - to - Speech Syntheses, Proceedings of the IEEE, April 1967, page 443 - 451
- [16] A. Yatagai: Limited and unlimited vocabulary speech synthesis systems, Electronic Engineering, November 1980, page 31 - 41
- [17] Jonathan Allen: Synthesis of Speech from Unrestricted Text, Proceedings of the IEEE, April 1967, page 433 - 442
- [18] Edward Bruckert, Martin Minow and Walter Tetschner: Three-tiered software and VLSI aid developmental system to read text aloud, Electronics, April 21, 1983, page 133 - 138
- [19] M.E. Hoff and Matt Townsend: Single-chip NMOS microcomputer processes signals in real time, Electronics, March 1, 1979, page 105 - 110
- [20] M.E. Hoff and Wallace Li: Software makes a big talker out of the 2920 microcomputer, Electronics, January 31, 1980, page 102 - 107
- [21] Veljko Zgaga: Sinteza govora u mikroracunarskim sistemima, Zbornik Informatica 77, Bled 1977, 2 212
- [22] James C. Anderson: An Extremely Low-Cost Computer Voice Response System, Byte, February 1981, page 36 - 43

\* Zaradi pomanjkanja domačih prispevkov o tej temi je podan obsežnejši spisak tujih prispevkov, ki obravnavajo področje sinteze govora.

- [23] Tomihiro Matsumura: Future Microprocessor Trends, Information Processing 83, R.E.A.Mason(ed.) page 213 - 217
- [24] Bruce LeBoss: Speech I/O is making itself heard, Electronics, May 22, 1980, page 95 - 105
- [25] Michael W. Hutchins and Lee Dusek : How Vocabulary is Generated Determines Speech Quality, Computer Design, February 1984, page 89 - 96
- [26] Single silicon chip synthesizes speech in \$50 learning aid, Electronics, June 22, 1978
- [27] James L. Flanagan, Manfred R. Schroder, Bishnu S. Atal, Ronald E. Crochiere, Nuggehally S. Jayant, Jose M. Tribolet : Speech Coding, IEEE Transaction on Communications Vol. COM-27, No. 4, April 1979, page 710 - 735
- [28] R.E.Crochiere, S.A.Webber and J.L.Flanagan Digital Coding of Speech in Sub-bands, The Bell System Technical Journal, Vol 55 No 8, October 1976, page 1069 - 1085
- [29] D.J.Goodman, B.J.McDermot and L.H.Nakatani: Subjective Evaluation of PCM Coded speech, The Bell System Technical Journal, Vol 55 No 8, October 1976, page 1087 - 1109
- [30] N.S.Jayant : Pitch-Adaptive DPCM Coding of Speech With Two-Bit Quantization and Fixed Spectrum Prediction, The Bell System Technical Journal, Vol 56, No 3, March 1977, page 439 - 454
- [31] L.R.Rabiner, C.E.Schmidt and B.S.Atal : Evaluation of Statistical Approach to Voiced - Unvoiced - Silence Analysis for Telephone-Quality Speech, The Bell System Technical Journal, Vol 56, No 3, March 77 page 455 - 482
- [32] D.J.Goodman, C.Scagliola, R.E.Crochiere, L.R.Rabiner and J.Goodman : Objective and Subjective Performance of Tandem Connections of Waveform Coders with an LPC Vocoder, The Bell System Technical Journal, Vol 58, No 3, March 1979, page 601 - 629.
- [33] J.J.Dubnowski and R.E.Crochiere: Variable Rate Coding of Speech, The Bell System Technical Journal, Vol 58, No 3, March 79 page 577 - 600
- [34] N.S.Jayant and S.W.Christensen : Adaptive Aperture Coding for Speech Waveforms - I, The Bell System Technical Journal, Vol 58 No 7, September 1979, page 1631 - 1645
- [35] J.L.Flanagan, J.D.Johnston, J.W.Upton : Digital Voice Storage in a Microprocessor, IEEE Transac. on communicat. Vol COM-30, No 2, February 1982, page 336 - 345

# informatica 85

Vabilo k sodelovanju

Call for Papers

**Posvetovanje in seminarji Informatica '85**  
Nova Gorica, 24. - 27. september 1985

**Posvetovanje**  
18. jugoslovansko mednarodno posvetovanje za računalniško tehnologijo in uporabo  
Nova Gorica, 24. - 27. september 1985

**Seminarji**  
Izbrana poglavja iz računalniške tehnologije in uporabe

**Razstava**  
Razstava računalniške tehnologije, uporabe, literature in drugih računalniških naprav, z mednarodno udležbo

**Roki**  
1. april 1985 Zadnji rok za sprejem formularja s prijavo in 2 izvodov razširjenega povzetka  
15. julij 1985 Zadnji rok za sprejem končnega teksta prispevka

**Symposium and Seminars Informatica '85**  
Nova Gorica, September 24th - 27th, 1985

**Conference**  
18th Yugoslav International Conference on Computer Technology and Usage

**Seminars**  
Selected Topics in Computer Technology and Usage  
Nova Gorica, September 24th - 27th, 1985

**Exhibition**  
Exhibition of Computer Technology, Usage, Literature and Other Computer Equipment with International Participation  
Nova Gorica, September 24th - 27th, 1985

**Deadlines**  
April 1, 1985 Submission of the application form and 2 copies of the extended summary  
July 15, 1985 Submission of the full text of contribution