
UDK 808.62-441

Originalni znanstveni rad

Primljeno: 11. 11. 1991.

Juraj Bakran, Milan Stamenković
Filozofski fakultet, Zagreb

**FORMANTI PRIRODNIH I SINTETIZIRANIH VOKALA
HRVATSKOGA STANDARDNOGA GOVORA**

SAŽETAK

U radu se iznose rezultati (i problemi) akustičke analize prirodno izgovorenih vokala hrvatskoga standardnoga govora te rezultati percepcije sintetiziranih vokala u trodimenzionalnom prostoru $F1 \times F2 \times F3$. Analiza i sinteza provedeni su digitalnom tehnikom. Rezultati percepcije sintetiziranih vokala podudaraju se s rezultatima akustičke analize prirodnih vokala (u slučajevima kad se oni mogu očitati). Dobiveni rezultati podupiru važnost formanata u percepciji vokala, ali upućuju na to da postoje alternativni mehanizmi percepcije vokala, koji se ne temelje nužno na određenom broju i položaju formanata.

UVOD

Prema tradicionalnom shvaćanju, frekvencije formanata vokala njihove su temeljne odrednice i percepcijski relevantni elementi. Jedna od prvih sustavnih analiza vokalskih formanata (za američki engleski) objavljena je u često citiranom radu Petersona i Barneya (1952) *Control Methods used in a Study of Vowels*. Iako se na temelju novijih istraživanja (Delattre et al. 1952; Chistovich et al. 1979; Chistovich 1985; Bladon 1982; Schroeder et al. 1979; Hermansky 1990) autori priklanjaju skeptičnijoj struji o značaju frekvencija pojedinačnih formanata, potreba da se ispuni jedna velika praznina u akustičkom opisu hrvatskoga standardnoga govora (HSG) i dug prema pionirskom radu Petersona i Barneya, osnovni su motivi za ovo istraživanje. Budući da na jedno od vrlo čestih pitanja: koje su frekvencije formanata vokala hrvatskoga standardnoga govora, nije bilo moguće odgovoriti, jer do sada nema objavljenog sistematskog istraživanja akustičkih karakteristika vokalskog sustava HSG, nisu se na temelju hrvatskoga vokalskoga sustava mogle razvijati ni teorije percepcije (normalizacije) automatskog prepoznavanja i sinteze vokala.

Osim analize vokalskih formanata prirodnog izgovora, u ovom radu bit će prvi put objavljeni rezultati percepcije sintetiziranih vokala hrvatskoga govora. Budući da analizom prirodnog izgovora nije moguće analizirati utjecaj pojedinih formanata, upotrijebljena je metoda sinteze kojom je to moguće ostvariti.

U analizi percepcije sintetiziranih vokala posebno je promatran utjecaj trećeg formanta na odabir optimalnih frekvencija prvih dvaju formanta.

Akustička analiza prirodnog izgovora otkriva različite izvore varijabiliteta vokalskog spektra (govornici, kontekst....) pa zbog toga nema jednostavnog odgovora na pitanje: koje su akustičke karakteristike vokala nekoga govora. Svaki pokušaj odgovora znači grubo pojednostavljenje. Analizom originalnog izgovora moguće je opisati vokalski sustav u izgovoru jednog konkretnog govornika ili izračunati srednju vrijednost nekog definirano uzorka. U ovom radu iznose se rezultati analize parametara vokala u izgovoru muških, ženskih i djece, govornika standardnoga hrvatskoga govora i rezultati percepcije sintetiziranih vokala prema procjeni izvornih govornika hrvatskoga govora.

Formanti se definiraju kao rezonantne frekvencije vokalskog trakta. Ako se formanti analiziraju tradicionalnom sonografskom metodom, prešutno se zanemaruje činjenice da ocrтана istaknuta mjesta u spektru vokala ne predstavljaju izravno rezonantne frekvencije vokalskog trakta, zapravo zanemaruje se utjecaj spektra pobude (laringalni ton) i radijacijske karakteristike usta (točkasti izvor zvuka). Drugim riječima, ono što se na sonografskim prikazima izravno proglašava formantima, nisu formanti. Prema opće prihvaćenoj "pobuda - filter" (source - filter) teoriji produkcije govora ono što se na spektrogramu vidi, zapravo su umnošci spektra pobude, frekvencijske karakteristike vokalskog trakta i radijacijske karakteristike prijelaza od usta do mikrofona (Fant 1960; Flanagan 1965). Uz to, pri očitavanju takvih "pojednostavljenih" formanata ne postoje nedvosmisleno definirani kriteriji određivanja njihovih frekvencija. Frekvencije formanata određuju se vizualnom impresijom o mogućim centrima akustičke mase na temelju "uske" ili "široke" analize. Prema širokoj analizi to je centar zacrnjenja (Peterson

i Barney 1952) a prema uskoj analizi to je zamišljeni lokalni maksimum grupe harmonika. Ova druga tehnika zapravo predstavlja amplitudom ponderiranu srednju vrijednost frekvencija grupiranih harmonika i naziva se optičko gladenje (fiting by eye; Peterson i Barney 1952; Dunn 1961; Pols et al. 1973; Specter 1990).

Teorijske frekvencije formanata mogu se izraziti samo na razini analognih (Fant 1960) ili digitalnih modela (Markel i Gray 1982) produkcije govora. Međutim, u većini slučajeva rezonancijske frekvencije vokalskog trakta imaju najveći utjecaj na konačan oblik spektra (naročito kod muških govornika). Zbog toga su frekvencije formanata dobivene inspekcijom spektrograma blizu teoretskih vrijednosti. To se ne može tvrditi za očitane amplitude formanata, jer nisu poznate amplitude pojedinih faktora.

AKUSTIČKA ANALIZA

Govorni uzorci

Govornici su čitali listu od 10 jednosložnih riječi u kojoj su se vokali dva puta pojavili pod dugim akcentom. Kako po uzoru na Petersona i Barneya (1952) u hrvatskom nije bilo moguće zadovoljiti uvjet istog konsonantskog okruženja, odabrane su riječi u kojima se pretpostavljala minimalan koartikulacijski utjecaj: *pir, bit, pet, red, pas, pag, bog, hod, but, puk*. Uz to snimljene su riječi: *prvi* i *vrlo*, da bi se analizirao izgovor "neutralnog vokala" oko /r/.

Opisane govorne uzorke izgovaralo je 45 govornika, od kojih je bilo 17 muškaraca (20 - 80 godina), 7 žena (20 - 54 godine) i 21 dijete (7 - 11 godina). Nisu analizirani govornici u dobi od 12 do 19 godina zbog pubertetskih transformacija. Podjela na odrasle i djecu nije posebno utemeljena. Nije kontrolirano dijalektalno porijeklo govornika, ali je analiziran samo onaj izgovor koji su autori procijenili standardnim. Kako su za svakog govornika analizirana dva uzorka (segmenti vokala iz dviju riječi), ukupan broj analiziranih segmenata jest za muške 34, za ženske govornike 14 i za djecu 42. Zbog razloga opisanih u poglavlju o teškoćama očitavanja, matrica podataka ne sadrži maksimalan broj.

Metoda analize

Analogne snimke govornog materijala digitalizirane su 12-bitnim A/D konvertorom uz frekvenciju uzimanja uzoraka 10 kHz i prethodno frekvencijsko ograničavanje do 4.5 kHz niskopropusnim filtrom 60 dB/oct. Digitalna obrada izvršena je na osobnom računalu uz primjenu programskog sustava za analizu akustičkoga govornoga signala AGOS (Stamenković et al. 1990). Iz centralnog dijela vokala izdvojen je segment trajanja 20 ms i frekvencijski analiziran FFT postupkom uz primjenu Hammingovog prozora. Budući da je FFT analiza provedena sa 512 spektralnih točaka, to je frekvencijska rezolucija analize oko 10 Hz.

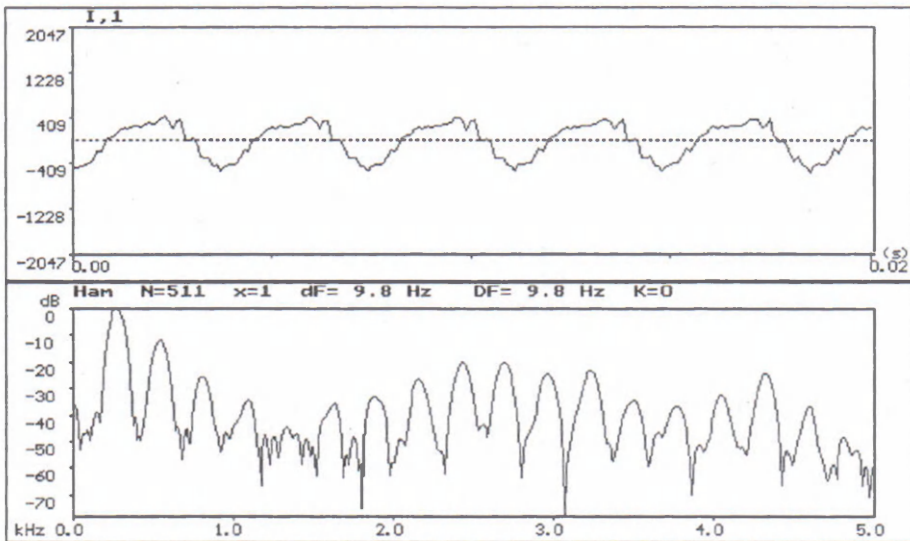
Problemi očitavanja formanata

Uspjeh očitavanja frekvencija formanata ovisi o načinu analize (prikaza). Mi smo se odlučili za samo jednu sekciju od 20 ms iz središnjeg dijela trajanja vokala i na taj način nismo imali kontrolu nad spektrom u čitavom trajanju vokala. Tako

smo, međutim, dobili precizne odnose amplituda pojedinih harmonika. Nadalje, budući da ne postoje nedvosmisleni kriteriji o tome što se smatra "istaknutim mjestom u spektru", uspješnost analize ovisi o samokritičnosti (strogosti) koju određuje eksperimentator.

Problemi očitavanja mogu se svesti na nekoliko kategorija:

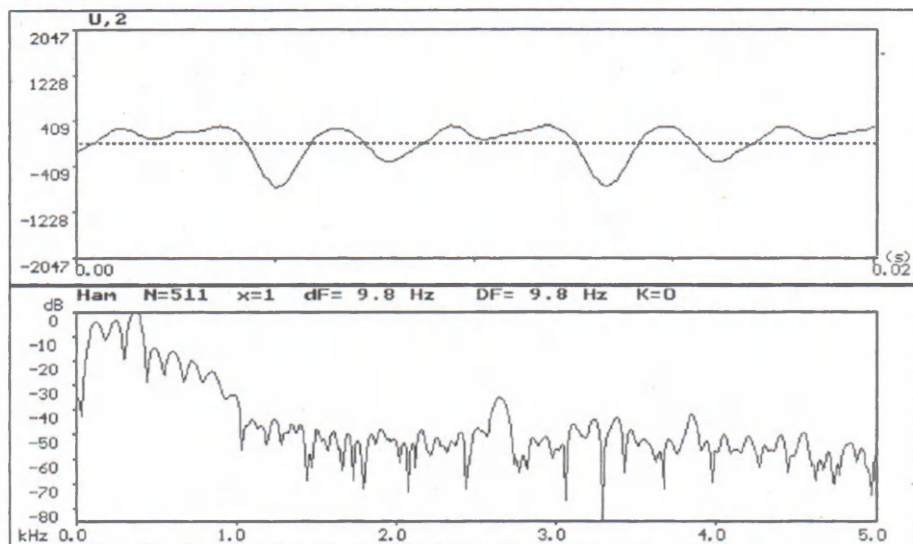
1. Amplitude harmonika padaju linearno počevši od F_0 , pa nije moguće identificirati maksimum, zapravo, on bi jedino mogao biti jednak F_0 , što nije očekivana vrijednost frekvencije formanta. Ovaj se problem uglavnom odnosi na očitavanje F_1 .



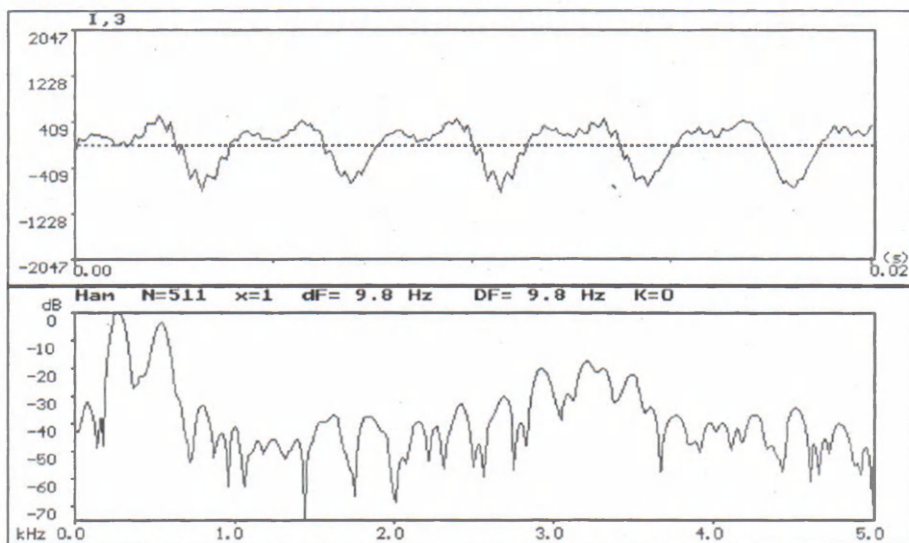
sl. 1 padajući harmonici

2. Nema istaknutih mjesta u spektru. Na mjestu frekvencijske skale na kojem se očekuje formant ne vidi se lokalni maksimum, tj. nema istaknutog harmonika oko kojeg bi se moglo procjenjivati ponderiranu srednju vrijednost. Jedan od razloga malom broju uspješnih očitavanja F_2 i F_3 vokala /u/ (i kod muškaraca i kod žena) jest prigušenje visokih frekvencija tako da formanti malih amplituda ulaze u zonu šuma.

3. Neočekivane frekvencije formanta. Istaknuta mjesta u spektru koja po vizualnom dojmu izgledaju kao formanti imaju frekvenciju neočekivanu za izgovoreni vokal. Takva pojava može biti manifestacija individualnih karakteristika govornika koje ne utječu na percepciju vokalske boje (Klatt i Klatt 1990). Na slici 3 prikazan je neočekivan položaj F_2 od 1750 Hz, jer je izgovoren /i/ za koji F_2 treba biti oko 3000 Hz.

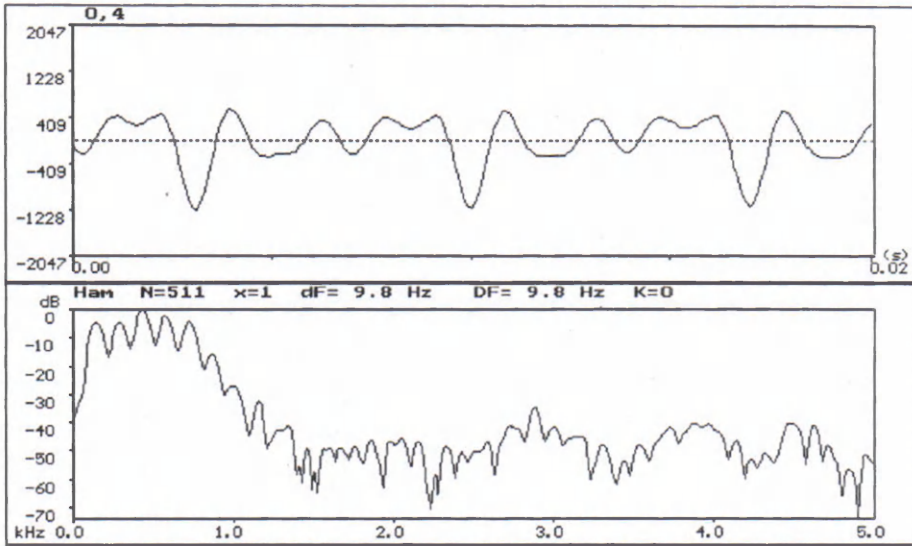


sl. 2 spektar bez istaknutih mjesta



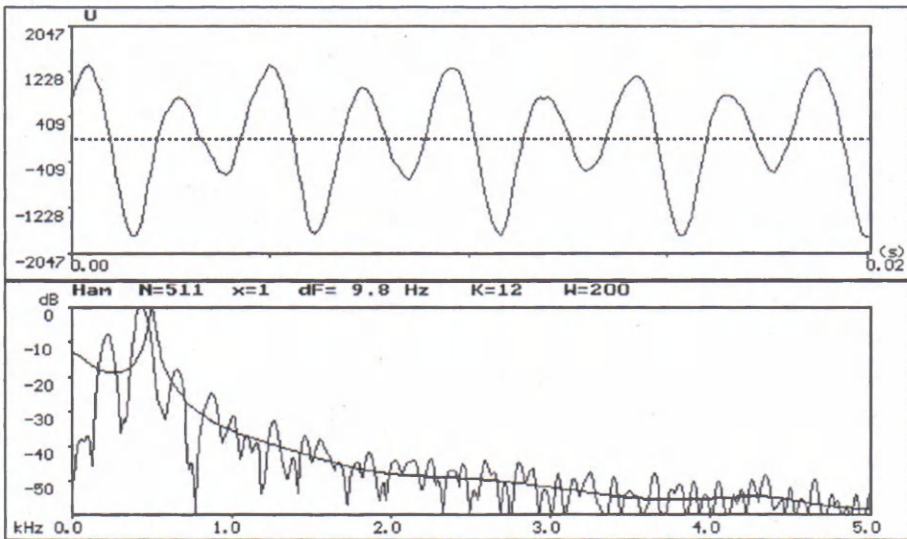
sl. 3 neočekivani formanti

4. Frekvencijski bliski formanti često se stapaju u masu s jednim vrhom. To se odnosi na F1 i F2 stražnjih vokala te na F2 i F3 prednjih vokala. U takvim slučajevima moguće je očitati samo centralnu frekvenciju čime se dobiva centar mase (COG, Dellatre et al. 1952; Chistovich et al. 1979), no time nismo razlučili frekvencije formantata.



sl. 4 spojeni formanti

5. Amplitude viših formanata stražnjih vokala toliko su male da se ne može odrediti "istaknuto mjesto u spektru" ili su ispod razine šuma okoline (sl. 5).



sl. 5 viši formanti nisu istaknuti

6. Visoka frekvencija osnovnog tona (žene i djeca) smanjuje točnost očitavanja, a kod artikulacijski visokih vokala nalazi se u zoni F1 pa nije moguće

identificirati formantski vrh. Jednako je tako kod bliskih formanata spriječno razdvajanje.

Uglavnom zbog opisanih teškoća, nisu kod svih uzoraka očitani svi parametri. Samo je frekvencija laringalnog tona očitana u svim uzorcima.

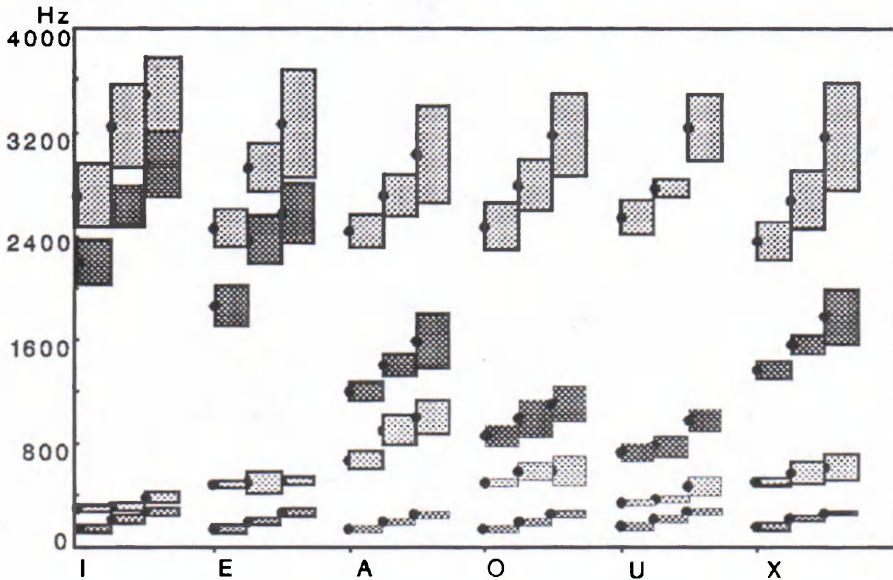
REZULTATI

Zbog bolje preglednosti, donosimo posebno tablicu samo sa srednjim vrijednostima očitanih frekvencija za F0, F1, F2, F3, posebno muških govornika, ženskih govornika i djece. Broj očitavanja na temelju kojih je izračunata svaka srednja vrijednost naveden je u tablici 4. Simbol "x" označava neutralni, nefonološki vokalski element, koji se, kad nema vokalskog okruženja, dodaje da bi se omogućila artikulacija "r" (u riječima "vrlo" i "prvi"). Brojevi u tablici 1 unatoč običaju, namjerno nisu još više zaokruživani prema vrijednostima koje bi se lakše pamtile zato da im se ne bi pridavala pretjerana važnost.

tab. 1 Srednje vrijednosti F0, F1, F2, F3

| | I | E | A | O | U | X |
|-------------------------|------|------|------|------|------|------|
| <i>muški govornici</i> | | | | | | |
| F0 | 135 | 131 | 131 | 132 | 139 | 138 |
| F1 | 282 | 471 | 664 | 482 | 324 | 489 |
| F2 | 2192 | 1849 | 1183 | 850 | 717 | 1342 |
| F3 | 2713 | 2456 | 2433 | 2472 | 2544 | 2353 |
| <i>ženski govornici</i> | | | | | | |
| F0 | 200 | 185 | 186 | 191 | 206 | 220 |
| F1 | 302 | 493 | 884 | 576 | 353 | 553 |
| F2 | 2623 | 2360 | 1393 | 980 | 758 | 1554 |
| F3 | 3246 | 2930 | 2709 | 2776 | 2764 | 2662 |
| <i>djeca</i> | | | | | | |
| F0 | 261 | 253 | 244 | 250 | 263 | 247 |
| F1 | 375 | 500 | 984 | 585 | 463 | 603 |
| F2 | 3033 | 2569 | 1581 | 1095 | 962 | 1761 |
| F3 | 3487 | 3255 | 3024 | 3173 | 3224 | 3151 |

Na slici 6 prikazane su srednje vrijednosti frekvencija (podaci iz tablice 1) kao točke u sredini stupića. Visina stupića predstavlja širinu pojasa od dvije standardne devijacije (podaci u tablici 2). Stupići u grupi po tri predstavljaju različite kategorije govornika (muški, ženski, djeca), a različito ispunjeni stupići (zbog bolje preglednosti) predstavljaju različite formante. Na apscisi prikaza označeni su vokali. Frekvencijska skala (ordinata) je linearna, pa zbog toga viši formanti imaju veće raspršenje.



sl. 6 Srednje vrijednosti i standardne devijacije F0, F1, F2, F3
 tab. 2 Standardne devijacije frekvencija F0, F1, F2, F3

| | I | E | A | O | U | X |
|-------------------------|-------|-------|-------|-------|-------|-------|
| <i>muški govornici</i> | | | | | | |
| F0 | 23.6 | 24.4 | 22.4 | 24.2 | 24.9 | 26.3 |
| F1 | 35.3 | 33.6 | 67.6 | 32.8 | 32.3 | 34.6 |
| F2 | 170.7 | 163.5 | 68.9 | 88.5 | 72.2 | 72.2 |
| F3 | 250.3 | 148.4 | 128.8 | 191.8 | 130.0 | 150.3 |
| <i>ženski govornici</i> | | | | | | |
| F0 | 24.8 | 30.6 | 26.9 | 29.7 | 26.5 | 17.6 |
| F1 | 29.1 | 88.7 | 111.9 | 69.8 | 27.1 | 91.9 |
| F2 | 155.5 | 182.6 | 89.8 | 147.2 | 83.0 | 76.0 |
| F3 | 321.5 | 179.9 | 162.2 | 204.8 | 69.3 | 211.6 |
| <i>djeca</i> | | | | | | |
| F0 | 26.0 | 26.0 | 24.4 | 25.6 | 22.9 | 19.5 |
| F1 | 47.9 | 24.0 | 133.5 | 103.5 | 73.6 | 101.5 |
| F2 | 340.4 | 229.0 | 214.1 | 146.9 | 87.6 | 209.2 |
| F3 | 281.2 | 408.7 | 375.3 | 321.1 | 254.3 | 408.9 |

U tablici 3 prikazani su intenziteti spektralnih vrhova a ne intenziteti formantata, kako je u uvodu istaknuto, jer je to produkt najmanje triju faktora kojima ne znamo veličine. Vrijednosti u tablici imaju deskriptivnu funkciju, a mogu biti podatak za sintetiziranje.

Napomena: Amplitude formantata izražene su u decibelima. Maksimalan je broj 0,0, sve ostale vrijednosti su negativnog predznaka.

Tablica 4 donosi podatke o uspješnosti očitavanja frekvencija formantata. Samo je F0 u svim uzorcima očitano bez izuzetka. Iz tablice se može izračunati da je F1 uspješno očitano u 79% uzoraka, F2 u 74% uzoraka, a F3 u 72% uzoraka. Pri tom, najuspješnije se očitavaju uzorci muških govornika (91%), zatim ženskih, (70%) a najlošije uzorci izgovora djece (65)%. Formanti različitih vokala ne očitavaju se jednako uspješno. Ukupno, formanti /i/ očitani su u 75% slučajeva, kod /e/ u 87%, kod /a/ u 90%, kod /o/ u 70% a kod /u/ u samo 46% uzoraka.

Naravno da se uz srednje vrijednosti i st. devijacije izračunate na temelju manje od 30 elemenata ne može zaključivati o statističkim vrijednostima populacije pa takvi podaci imaju samo deskriptivnu ulogu u odnosu prema analiziranom korpusu. Uz ovako slabu uspješnost očitavanja formantata ženskih i dječjih govornika trebalo bi udeseterostručiti početni korpus da bi se dobio dovoljan broj uspješno očitanih vrijednosti.

U slučajevima kad formante nije bilo moguće zasebno očitati (problem očitavanja, tablica 4), istaknuto mjesto promatrano je kao jedan formant. To je najčešće kod dječjih glasova. Kao što se iz tablice 5 vidi, tako očitane frekvencije nalaze se između frekvencija formantata koji su se mogli očitati.

tab. 3 Intenziteti spektralnih vrhova

| | I | | E | | A | | O | | U | | X | |
|-------------------------|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|
| | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s |
| <i>muški govornici</i> | | | | | | | | | | | | |
| A1 | 0.0 | 0.0 | 0.2 | 1.2 | 0.4 | 1.5 | 0.0 | 0.0 | 0.3 | 0.8 | 0.3 | 1.2 |
| A2 | 19.1 | 6.9 | 14.9 | 5.3 | 8.0 | 4.8 | 12.7 | 6.6 | 13.7 | 4.7 | 16.7 | 3.8 |
| A3 | 23.6 | 6.5 | 19.9 | 4.9 | 24.8 | 5.7 | 34.0 | 6.6 | 39.2 | 6.9 | 29.5 | 5.8 |
| <i>ženski govornici</i> | | | | | | | | | | | | |
| A1 | 0.0 | 0.0 | 0.3 | 0.9 | 0.4 | 1.3 | 0.0 | 0.0 | 0.8 | 1.6 | 0.0 | 0.0 |
| A2 | 23.7 | 6.0 | 14.7 | 7.2 | 10.5 | 6.3 | 5.7 | 2.5 | 14.0 | 5.6 | 20.0 | 5.5 |
| A3 | 26.1 | 4.7 | 20.2 | 6.2 | 21.5 | 3.0 | 33.1 | 5.3 | 42.0 | 1.4 | 25.3 | 6.6 |
| <i>djeca</i> | | | | | | | | | | | | |
| A1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 0.8 | 0.2 | 0.8 | 0.0 | 0.0 | 1.0 | 2.6 |
| A2 | 22.4 | 5.5 | 19.2 | 6.5 | 5.9 | 4.0 | 9.1 | 3.5 | 22.5 | 2.0 | 16.7 | 6.1 |
| A3 | 26.2 | 7.8 | 21.9 | 6.0 | 22.3 | 4.9 | 37.0 | 8.6 | 42.2 | 6.9 | 28.2 | 6.8 |

tab.4 Tablica broja uspješnih očitavanja za svaki formant

| | I | | E | | A | | O | | U | | X | |
|-------------------------|----|-----|----|-----|----|-----|----|-----|----|-----|----|-----|
| | n | % | n | % | n | % | n | % | n | % | n | % |
| <i>muški govornici</i> | | | | | | | | | | | | |
| F0 | 34 | 100 | 34 | 100 | 34 | 100 | 34 | 100 | 34 | 100 | 28 | 100 |
| F1 | 34 | 100 | 33 | 97 | 33 | 97 | 32 | 94 | 34 | 100 | 28 | 100 |
| F2 | 34 | 100 | 34 | 100 | 34 | 100 | 26 | 76 | 22 | 65 | 26 | 93 |
| F3 | 23 | 68 | 29 | 85 | 33 | 97 | 33 | 97 | 24 | 71 | 27 | 96 |
| <i>ženski govornici</i> | | | | | | | | | | | | |
| F0 | 14 | 100 | 14 | 100 | 14 | 100 | 14 | 100 | 14 | 100 | 12 | 100 |
| F1 | 11 | 78 | 12 | 86 | 14 | 100 | 10 | 71 | 9 | 64 | 10 | 83 |
| F2 | 13 | 93 | 14 | 100 | 13 | 93 | 3 | 21 | 3 | 21 | 11 | 92 |
| F3 | 9 | 64 | 11 | 78 | 13 | 93 | 14 | 100 | 2 | 14 | 10 | 83 |
| <i>djeca</i> | | | | | | | | | | | | |
| F0 | 42 | 100 | 42 | 100 | 42 | 100 | 42 | 100 | 42 | 100 | 40 | 100 |
| F1 | 18 | 43 | 37 | 88 | 30 | 71 | 20 | 47 | 14 | 33 | 33 | 82 |
| F2 | 27 | 64 | 38 | 90 | 31 | 74 | 19 | 45 | 8 | 19 | 39 | 97 |
| F3 | 28 | 67 | 28 | 67 | 38 | 90 | 35 | 83 | 12 | 28 | 32 | 80 |

tab. 5 Središnje frekvencije nerazlučivih formanata

| | I (11) | E (5) | A (8) | O (16) | U (19) |
|-------|----------|---------|----------|--------|--------|
| F1=F2 | | | 1052 126 | 640 92 | 468 64 |
| F2=F3 | 3294 182 | 2898 28 | | | |

Uz vokal označen je broj mjerenja. Prvi broj označava srednju vrijednost, a drugi standardnu devijaciju. Podaci se odnose samo na dječji izgovor (razlog je visoka F0).

PERCEPCIJA SINTETIZIRANIH VOKALA

Druga metoda kojom smo definirali frekvencije formanata HSG je slušnom prosudbom boje sintetiziranih izoliranih vokala (analysis by synthesis) kojima su sistematski mijenjane frekvencije prvih triju formanta uz konstantne ostale parametre, uključujući varijaciju frekvencije osnovnog tona i intenzitetski oblik. U prirodnom izgovoru i najvještijem govorniku nije moguće po volji mijenjati frekvencije pojedinih formanata i time kontrolirano određivati boju vokala, pa je utjecaj pojedinih formanata na percepciju vokalske boje moguće promatrati samo s pomoću sintetiziranih stimulusa.

Opis sintetiziranih stimulusa

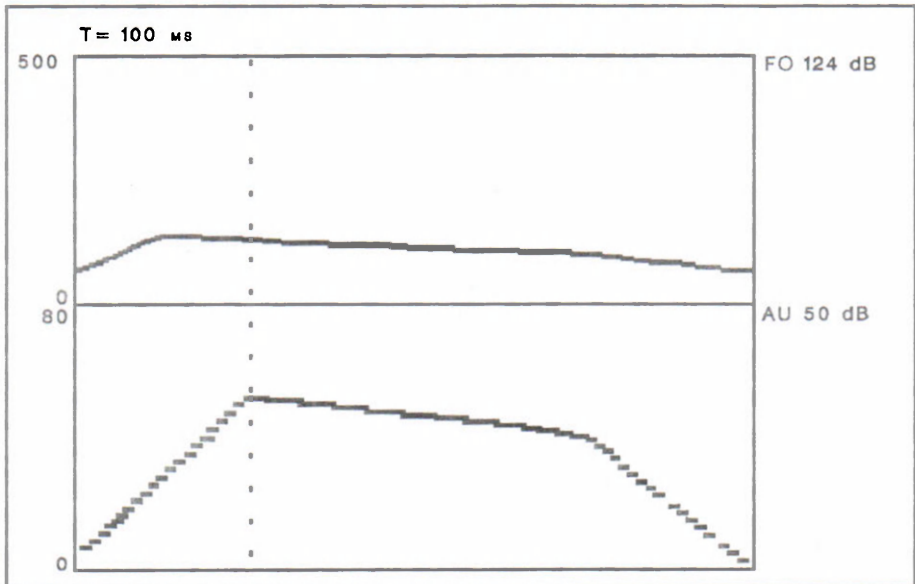
S pomoću Klattovog sintetizatora (Klatt 1980) kreiran je niz vokala. Intenzitetska funkcija (AV), tok frekvencije osnovnog tona (F0) i trajanje (270 ms) određeni su prema analizi prirodnog izgovora, tako da odgovaraju izolirano izgovorenim glasovima. Tok frekvencije osnovnog tona i intenzitetski oblik sintetiziranih vokala prikazani su na slici 7.

Frekvencije formantata sintetizirane su bez variranja u trajanju, bez tranzijenata, a prema frekvencijama:

| | F1 | F2 | F3 |
|----------------|----------|----------|-----------|
| dijapazon (Hz) | 200-1100 | 300-2500 | 1900-3300 |
| stepenice (Hz) | 50 | 100 | 200 |

Kompletna matrica ne sadrži kombinacije u kojima se pojedini formanti preklapaju.

I sve ostale vrijednosti parametara Klattovog sintetizatora (osim navedenih promjena) zadržane su nepromijenjene. Te su vrijednosti iznesene u tablici 6 za polazni "neutralni" vokal:



sl.7 Dijagram F0 i intenzitet sintetiziranih vokala

tab. 6

| | | | | | | | | |
|-----|---|------|-----|---|------|-----|---|-------|
| AV | 2 | 0 | A3 | 1 | 0 | B4 | 0 | 250 |
| AF | 1 | 0 | A4 | 1 | 0 | F5 | 1 | 3750 |
| AH | 1 | 0 | A5 | 1 | 0 | B5 | 0 | 200 |
| AVS | 1 | 0 | A6 | 1 | 0 | F6 | 0 | 4900 |
| F0 | 2 | 0 | AB | 1 | 0 | B6 | 0 | 1000 |
| F1 | 1 | 500 | B1 | 1 | 50 | FNP | 0 | 250 |
| F2 | 1 | 1500 | B2 | 1 | 70 | BNP | 0 | 100 |
| F3 | 1 | 2500 | B3 | 1 | 110 | BNZ | 0 | 100 |
| F4 | 1 | 3500 | SW | 0 | 0 | BGS | 0 | 200 |
| FNZ | 1 | 250 | FGP | 0 | 0 | SR | 0 | 10000 |
| AN | 0 | 0 | BGP | 0 | 100 | NWS | 0 | 50 |
| A1 | 0 | 0 | FGZ | 0 | 1500 | G | 0 | 40 |
| A2 | 1 | 0 | BGZ | 0 | 6000 | NFC | 0 | 5 |

Kratice varijable (parametra) I, IV i VII stupac Status varijable (II, V i VIII stupac) 0 - konstanta, 1 - promjenjiv, 2 - specificiran grafičkim editorom svakih 5 milisekundi. Značenje kratica navedeno je prema Klatt (1980).

Izlučivanje optimalnih stimulusa

U prvoj fazi iteracije prema optimalnim kombinacijama formanata skupina od 60 studenata slušala je sve stimuluse nastojeći ih svrstati u jednu od pet vokalskih kategorija.

Za drugu fazu, na temelju rezultata slušanja kompletnog korpusa, popis stimulusa ograničen je samo na one koji su u 90 posto slučajeva svrstani u jednu od 5 vokalskih kategorija. U ovoj fazi 40 studenata fonetike slušali su stimuluse grupirane u kategorije i procjenjivali udaljenost od, prema njihovoj predodžbi, idealnog predstavnika kategorije.

Za treću fazu iteracije, iz svake F3 ravnine (F1xF2) izlučeno je 9 najviše rangiranih stimulusa iz prethodne faze. Oni su podijeljeni u skupine po 3, a 10 fonetičara (vještijih slušača) iz svake skupine izdvojilo je najboljeg predstavnika. Postupak je ponovljen s najboljima iz svake skupine i na taj način dobili smo za svaku F3 matricu po jednu kombinaciju F1/F2 (od svakog od 10 ispitanika). Srednje vrijednosti zabilježene su u tablici 7.

U posljednjoj je fazi približavanja procjeni optimalnih kombinacija na sličan način od 8 najboljih predstavnika za svaki F3 diskriminacijom slušanjem u paro-

vima izlučen ukupno optimalan predstavnik za svaku vokalsku kategoriju (individualno za svakog ispitanika). Originalni podaci ove faze upisani su u tablici 10.

Upute ispitanicima

Uputa ispitanicima (u dijelu svrstavanja kompletnog korpusa u kategorije, bila je da svaki stimulus svrstaju svakako u jednu od pet vokalskih kategorija.

U dijelu grupiranog slušanja (samo isti vokali u grupi) uputa je bila da brojevima od 0 do 5 označe, prema subjektivnoj procijeni, sličnost stimulusa s onim što smatraju idealnim predstavnikom vokalske kategorije koja se procjenjuje.

U posljednjem dijelu izlučivanja ispitanici su od tri stimulusa prezentirana u grupi trebali izdvojiti najbolji. Pojedine grupe mogle su se ponovno slušati i prosuđivati po želji ispitanika.

REZULTATI (sinteza)

U tablici 7 prikazani su rezultati odabira optimalnih vrijednosti F1 i F2 za svaki od F3 u rasponu od 1900 do 3300 Hz.

tab. 7 Srednje vrijednosti F1 i F2 optimalnih sintetiziranih vokala

| | I3 | 1900 | | 2100 | | 2300 | | 2500 | | 2700 | | 2900 | | 3100 | | 3300 | |
|---|----|-----------|----|-----------|----|-----------|----|-----------|----|-----------|----|-----------|----|-----------|----|-----------|----|
| | | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s |
| I | F1 | | | | | 261 | 11 | 261 | 22 | 239 | 22 | 233 | 25 | 222 | 26 | 239 | 22 |
| | F2 | | | | | 2144 | 72 | 2177 | 99 | 2355 | 53 | 2200 | 86 | 2233 | 71 | 2200 | 86 |
| E | F1 | | | 467 | 25 | 483 | 25 | 467 | 25 | 472 | 26 | 472 | 26 | 483 | 25 | 489 | 22 |
| | F2 | | | 1855 | 53 | 1922 | 44 | 2011 | 60 | 2011 | 67 | 2066 | 86 | 1977 | 67 | 2200 | 99 |
| A | F1 | 828 | 36 | 822 | 26 | 838 | 35 | 828 | 26 | 811 | 22 | 872 | 26 | 833 | 25 | 861 | 22 |
| | F2 | 1189 | 60 | 1167 | 50 | 1167 | 50 | 1144 | 53 | 1178 | 67 | 1156 | 73 | 1156 | 53 | 1167 | 71 |
| O | F1 | 481 | 26 | 487 | 23 | 481 | 26 | 488 | 23 | 481 | 26 | 456 | 18 | 468 | 26 | 475 | 27 |
| | F2 | 737 | 52 | 750 | 53 | 725 | 46 | 812 | 35 | 787 | 35 | 725 | 46 | 725 | 46 | 712 | 35 |
| U | F1 | 269 | 37 | 275 | 27 | 250 | 27 | 250 | 27 | 250 | 27 | 231 | 26 | 231 | 26 | 238 | 23 |
| | F2 | 638 | 52 | 650 | 53 | 625 | 46 | 600 | 76 | 575 | 71 | 563 | 52 | 613 | 64 | 588 | 83 |

Na pitanje kako F3 djeluje na odabir F1 i F2 nije moguće jednostavno odgovoriti. Analiza varijance za svaki pojedini vokal i F1/F2 pokazuje da na neke formante F3 ima utjecaj, a na neke nema:

tab. 8 Značenje utjecaja F3 na F1 i F2

| | I | E | A | O | U |
|----|-------|-------|-------|-------|-------|
| F1 | <0.01 | | <0.01 | | <0.05 |
| F2 | <0.01 | <0.01 | | <0.01 | |

tab. 9 Usporedba srednjih vrijednosti odabira sintetiziranih frekvencija i rezultata analize muških govornika:

| | I | | E | | A | | O | | U | |
|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|
| | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s | \bar{x} | s |
| F1s | 245 | | 480 | | 815 | | 470 | | 260 | |
| F1a | 282 | 35 | 471 | 34 | 664 | 68 | 482 | 33 | 324 | 32 |
| F2s | 2270 | | 1990 | | 1140 | | 750 | | 600 | |
| F2a | 2192 | 171 | 1849 | 163 | 1183 | 69 | 850 | 88 | 717 | 72 |
| F3s | 2760 | | 2500 | | 2500 | | 2540 | | 2360 | |
| F3a | 2713 | 250 | 2456 | 148 | 2433 | 129 | 2472 | 192 | 2544 | 130 |

S F1s, F2s i F3s označene su srednje vrijednosti frekvencija formanta iz tablice 10, a s F1a, F2a i F3a, srednje vrijednosti rezultata analize muških govornika iz tablice 1.

Konačnu tablicu namjerno donosimo bez ikakva usrednjavanja, tako da ostaju sačuvani podaci o kombinacijama F1/F2/F3 koje nose isti rezultat - odabir konkretne vokalske kategorije.

U eksperimentima za sintetiziranim dvoformantskim vokalima ustanovljeno je da je takav model sinteze F2 prednjih vokala sistematski viši, nego u analiziranom korpusu prirodnih vokala. To je razumljivo, jer treba kompenzirati manjak utjecaja bliskog trećeg formanta i viših formanta na vokalsku boju.

Iz usporedbe rezultata percepcije sintetiziranih vokala i rezultata analize prirodnog izgovora muških govornika vidimo da je razlika među frekvencijama oko jedne standardne devijacije. Uočili smo da su ispitanici najveću nedoumicu u odlučivanju pokazali slušajući varijante vokala /u/.

tab. 10 Konkretni optimalni vokali za svakog od 10 ispitanika: Brojevi od 1 do 0 predstavljaju šifre pojedinih ispitanika.

| <i>optimalni F3</i> | | | | | |
|---------------------|----------|----------|----------|----------|----------|
| Hz | I | E | A | O | U |
| 2900 | 24579 | | | 78 | |
| 2700 | 168 | 6 | 8 | 345 | 3 |
| 2500 | 30 | 12345890 | 12346790 | 0 | 190 |
| 2300 | | 7 | 5 | 129 | 468 |
| 2100 | | | | 6 | 27 |
| <i>optimalni F2</i> | | | | | |
| Hz | I | E | A | O | U |
| 2400 | 10 | | | | |
| 2300 | 2468 | | | | |
| 2200 | 379 | | | | |
| 2100 | 5 | 5 | | | |
| 2000 | | 1236890 | | | |
| 1900 | | 47 | | | |
| 1200 | | | 2490 | | |
| 1100 | | | 135678 | | |
| 800 | | | | 12340 | |
| 700 | | | | 56789 | 20 |
| 600 | | | | | 134678 |
| 500 | | | | | 59 |
| <i>optimalni F1</i> | | | | | |
| Hz | I | E | A | O | U |
| 850 | | | 260 | | |
| 800 | | | 1345789 | | |
| 500 | | 124670 | | 3460 | |
| 450 | | 3589 | | 125789 | |
| 300 | 3 | | | | 37 |
| 250 | 1456780 | | | | 12456890 |
| 200 | 29 | | | | |

DISKUSIJA I ZAKLJUČAK

Nismo očekivali tako visok postotak nemogućnosti da se očitavaju formanti. Najmanje uspješni rezultati su u očitavanju dječjih i ženskih vokala, a najviše zbog visoke frekvencije laringalnog tona, odnosno zbog, široko razmaknutih harmonika. To je jedan od razloga da su u većini radova koji analiziraju frekvencije formanata ispitanici samo muškarci.

Postavlja se logično pitanje: ako akustičkom analizom nije moguće razlučiti frekvencije formanata (zbog bilo kojega od opisanih razloga), djeluju li u takvim slučajevima formanti u percepciji kao zasebni elementi? Vrlo vjerojatno ne. Prema tome skloni smo pridružiti se onim istraživačima koji u takvim slučajevima promatraju jednu spektralnu masu s jednim centrom gravitacije - COG (Chistovich et al. 1979).

Za naše rezultate znatno je odgovorna odabrana metoda: analiza samo jednog segmenta od 20 ms, uska analiza s frekvencijskom rezolucijom od 10 Hz, vizualno određivanje centra formantske mase i nedefiniran stupanj samokritičnosti istraživača koji analizira. Kriterij prema kojemu se neko istaknuto mjesto u spektru proglašava formantom i prema kojem se bilježi njegova frekvencija, ne može se strogo definirati. Oslanja se na iskustvo, znanje o mogućem položaju formanta za neki vokal. Na taj se način djelomice izbjegava zamka da se individualne karakteristike govornika, koje postoje i u frekvencijskom rasponu prvih triju formanata, a također nalikuju na formante, proglaše formantima. Budući da se u objavljenim radovima uglavnom analiziraju muški govornici, greška u analizi nije toliko očita, pa se neki usuđuju tu metodu proglašavati objektivnom (Pols et al. 1973).

Usrednjavanje frekvencija pojedinih formanata možda je posve kriv put, jer se temelji na pretpostavci da formanti imaju svaki zasebno neku perceptivnu funkciju. Ostaje otvoreno pitanje što to iz akustičkog signala, koji u analizi pokazuju takvo raspršenje, u percepciji ostaje konstanta - neka vokalska kategorija?

Pristup sintezi koji smo ovdje imali - djelomično imitiranje prirodnog izgovora, (za definiranje intonacije i intenzitetske funkcije) ispitanicima je, s pomoću iluzije da se radi o prirodnom govoru, omogućavao nesintetički - govorni način slušanja. Kako je sinteza formantskih frekvencija podrazumijevala kompletan trodimenzionalni prostor ($F1 \times F2 \times F2$ - sa zadanim stepenicama) nikakve formantske kombinacije nisu bile prejudicirane. Zbog toga relativna podudarnost rezultata analize muških govornika i procjene sintetskih stimulusa (slična $F0$) čudi i ujedno reafirmira teorije o formantima kao ključnim odrednicama vokalskog prostora. Takav zaključak stoji uz ograničenje da su vokali sintetizirani troformantskim modelom. Naravno, nije izvjesno kakvi bi rezultati percepcije bili uz drugačiji model sinteze (npr. jednoformantski ili dvoformantski model), odnosno kako bismo opisali rezultate akustičke analize da nismo inzistirali na traženju prvih triju formanata. S obzirom na to da u mnogim slučajevima u akustičkoj analizi prirodnih vokala nije bilo moguće identificirati troformantski model, nameće se zaključak da postoje alternativni mehanizmi percepcije vokala koji se ne temelje nužno na definiranom broju i položaju formanata.

Ostaje otvoreno pitanje pronalazjenja tih alternativnih modela percepcije vokala i utvrđivanja uvjeta pod kojima pojedini model dominira percepcijom.

REFERENCIJE

Bakran, J. (1989). *Djelovanje naglasaka i dužine na frekvencije formanata*. Govor, VI, br.2, 1-12.

Bladon, A.e(1982). *Arguments against formants in the auditory representation of speech* U: R. Carlson i B. Granstrom (cda) *The Representation of Speech in the Peripheral Auditory System*.

Chistovich, L.A., Sheikin, R.L. i Lublinskaja, V.V. (1979). *Centres of Gravity: and Spectral Peaks as the Determinants of Vowel Quality*. U: B. Lindblom i S. Ohman (ur), *Frontiers of Speech Communication Research*. Academic Press, London.

Chistovich, L.A. (1985). *Central auditory processing of peripheral vowel spectra*. JASA, vol. 77, str. 789-805.

Delattre, P., Liberman, A.M., Cooper, F.S., Gertsman, L.J. (1952). *An experimental study of the acoustic determinants of vowel color*. Word 8, 195-210.

Dunn, H.K. (1961). *Methods of Measuring Vowel Formant Bandwidths*. JASA, No 12, 1737-1746.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton, The Hague.

Flanagan, J.L. (1965). *Speech Analysis Synthesis and Perception*. Academic Press, New York.

Hermansky Hynek (1990). *Perceptual linear predictive (PLP) analysis of speech*. JASA, vol. 87, str. 1738-1752.

Klatt, D.H. (1980). *Software for cascade/parallel synthesizer*. J. Acoust. Soc. Am. 67, 971-975.

Klatt, D.H. (1987). *Review of text-to-speech conversion for English*. J. Acoust. Soc. Am. 82, 737-793.

Klatt, D.H., L.C. Klatt (1990). *Analysis, synthesis and perception of voice quality variations among female and male talkers*. J. Acoust. Soc. Am. 87, 820 - 857

Markel, J.D. (1972). *Digital Inverse Filtering - A New Tool for Formant Trajectory Estimation*. IEEE Transactions on Audio Electroacoustics, vol AU-20, 129-137.

Markel, J.D. i Gray, A.H. (1982). *Linear Prediction of Speech*. Springer-Verlag, Berlin.

Pavlovic, C.V., M. Rossi, R. Espesser (1990). *Use of magnitude estimation technique for assessing the performance of text-to-speech synthesis systems*. J. Acoust. Soc. Am. 87, 373-383.

Peterson, G.E. i H.L. Barney (1952). *Control methods used in study of vowels*. J. Acoust. Soc. Am. 24, 175 - 184, (reprint u D.B. Fry, *Acoustic Phonetics*, Cambridge University Press, 1976).

Pols, I.C.W., H.R.C. Tromp, R. Plomp (1973). *Frequency analysis of Dutch Vowels from 50 male speakers*. J. Acoust. Soc. Am. 53, 1093-1101.

Schroeder, M.R., Atal, B.S. i Hall, J.L. (1979). *Objective Measure of Certain Speech Signal Perception*. U: (eds) B. Lindblom i S. Ohman, *Frontiers of Speech Communication Research*, Academic Press, London, 217 - 229.

Speeter-Beddor, P. (1990). *The influence of spectral prominence on perceived vowel quality*. JASA, vol. 87, str. 2684-2704.

Stamenković, M., J. Bakran, M. Miletić, P. Tancig (1990). *AGOS - programski sistem za analizu govornog signala*. Zbornik savjetovanja: Informatička tehnologija u primijenjenoj lingvistici, Zagreb, 1990. 17 - 22.

Juraj Bakran, Milan Stamenković
Faculty of Philosophy, Zagreb

*THE FORMANTS OF NATURAL AND SYNTHESIZED VOWELS
IN CROATIAN STANDARD SPEECH*

Summary

The paper is concerned with the results (and problems) of the acoustic analysis of naturally pronounced vowels of Croatian standard speech as the results of the perception of synthesized vowels in three - dimensional space $F1 \times F2 \times F3$. Both the analysis and the synthesis have been carried out by means of digital technique. The results of the perception of synthesized vowels correspond to the results of the acoustic analysis of natural vowels (in cases where they can be read). The results that have been obtained underline the importance of formants in vowel perception but they also point to the existence of alternative mechanisms for the perception of the vowels that are not necessarily based on a definite number and a definite position of formants.