
UDK 801.41:159.9

159.932:801.4

801.41:534

Izvorni znanstveni rad

Prihvaćeno 25.05.1994.

Juraj Bakran,
Filozofski fakultet, Zagreb

*FUNKCIJA VOKALSKIH TRANZIJENATA U
PERCEPCIJI SINTETIČKIH OKLUZIVA¹*

SAŽETAK

Vokalski tranzijent nakon okluzije predstavlja akustički znak za raspoznavanje, odnosno za međusobno razlikovanje okluziva. Sintetizirani su CV slogovi sa zvučnim okluzivima i hrvatskim vokalima tako da su sustavno mijenjane početne frekvencije prvih triju formanata. Šum eksplozije nije uključen. Testiranje slušanjem pokazalo je da treći formant djeluje na raspoznavanje mjesta artikulacije, odnosno na međusobno razlikovanje zvučnih okluziva ali da, suprotno nekim tvrdnjama, taj utjecaj nije presudan.

Ključne riječi: vokalski tranzijenti, okluzivi, percepcija okluziva, akustički znak, hrvatski jezik

(1) Računarsku opremu osigurao je FELMET iz Zagreba.

UVOD

U govornoj komunikaciji akustički kanal nedvojbeno je dominantan. Zbog toga je upravo akustička fonetika pozvana da pridonese razumijevanju percepcije (i produkcije) govora. Jedan od temeljnih problema pritom je odrediti akustičke znakove (acoustic cues) odgovorne za prepoznavanje fonemskih predodžaba. Među konsonantima najistraživanija je kategorija okluziva (*p, t, k, b, d, g*). Mnoštvo akustičkih znakova relevantnih za percepciju okluziva, koji se u spominju u literaturi, moguće je podijeliti na one koji okluzive odvajaju kao glasovnu kategoriju (razlikuju ih od frikativa, afrikata itd), zatim na one koji ih dijele na zvučne i bezzvučne te one koji tako raspolovljenu kategoriju dijele na pojedine članove. Ova posljednja grupa akustičkih znakova naziva se najčešće prema artikulacijskoj terminologiji pa dijeli pojedine podskupove prema "mjestu artikulacije" na bilabijalne, alveolarne i palatalne. U akustičkom odvijanju okluziva ovi akustički znakovi smješteni su (i bez obzira na kontekst) izvan glavnine vremena koje okluzivi zapremaju. Bezvučni se okluzivi (*p, t, k*) međusobno, kao i zvučni (*b, d, g*) unutar svoje skupine, razlikuju prema akustičkim znakovima smještenim na granicama sa susjednim glasovima. O relativnoj važnosti "ulaznog" i "izlaznog" tranzijenta i šuma eksplozije pisali smo na drugom mjestu (vidi Bakran et al. 1992).

Iz dosadašnjih radova u vezi s percepcijom okluziva poznato je da su tranzijenti vokala nakon eksplozije dovoljan akustički znak za identifikaciju mjesta artikulacije (Stevens and Blumstein 1978, Harris et al. 1958, Sharf and Hemeyer 1972, Blumstein and Stevens 1980; Bakran et al. 1992). Zanimljivo je da tranzijenti vokala prije okluzije, ako se radi o intervokalskom okluzivu, predstavljaju efikasniji akustički znak od tranzijenata nakon okluzije (Bakran et al. 1992). Dugo se vjerovalo da su prva dva formanta dovoljna za međusobno razlikovanje glasova govora (u prvom redu vokala) dok je, što se konsonanata tiče, ustanovljen termin "lokus" za početnu frekvenciju tranzijenta drugog formanta (Delattre et al. 1955) te "hub" (Potter et al. 1947) sugerirajući na taj način da je to dovoljan akustički znak. Delattre je predložio frekvencije lokusa drugog formanta za međusobno razlikovanje */b, d, g/*, odnosno za identifikaciju mjesta artikulacije: za *b/* 700 Hz, za *d/* 1800 i za *g/* 3000 Hz uz ograničenje da je za *g/* u kontekstu sa */u/* 700 Hz. U spomenutim radovima nije bilo govora o djelovanju trećeg formanta. Harris et al. (1958) i Datschewit (1989) smatraju da tranzijent trećeg formanta ima važnu ulogu za međusobno razlikovanje okluziva.

U pokušaju da se Klattovim sintetizatorom (Klatt 1980) oblikuju hrvatski okluzivi, postavio se problem određivanja odgovarajućih početnih frekvencija prvog, drugog i trećeg formanta vokalskih tranzijenata nakon eksplozije. Akustičkom analizom izgovora muških govornika dobivene su početne frekvencije tranzijenata hrvatskih okluziva, koje primijenjene kao parametri sintetizatora nisu postigle zadovoljavajući stupanj raspoznavanja. Zbog toga smo sustavnim mijenjanjem početnih vrijednosti frekvencija formantata sintetizirali niz okluziva i potom provjerili (testiranjem izvornih govornika) koje su vrijednosti za raspoznavanje okluziva najvjerojatnije. Na taj način ujedno ponavljamo Datschewitov pokus i odgovaramo na pitanje o funkcionalnosti trećeg formanta u percepciji okluziva.

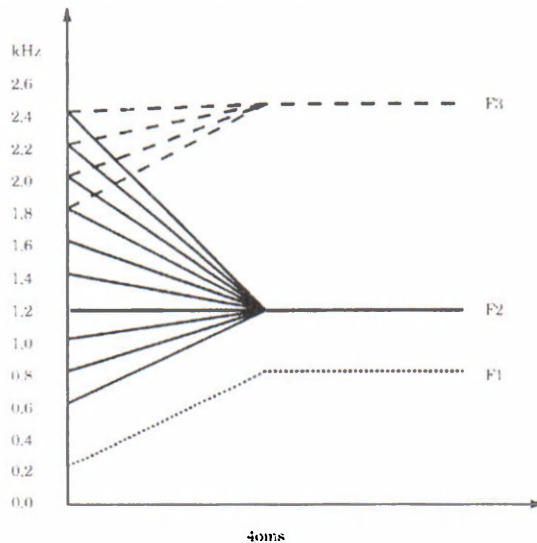
POKUS

Stimulusi

Sintetizirani su CV slogovi s konsonantima /b, d, g/ i s pet hrvatskih vokala /i, e, a, o, u/. Datscheweit (1989) je radio samo sa /i, a, u/. Intonativni oblik, te trajanje ugodeni su tako da su, što je više moguće, nalik na prirodno izgovorene slogove muškog govornika s dugosilaznim naglaskom. Trajanje stimulusa bilo je 400 milisekundi. Frekvencije formanata vokala (target) namještene su prema Bakran i Stamenković (1990), slično kao što su Harris et al. (1958) upotrijebili za engleski jezik rezultate analize Petersona i Barneya. Istina je da se vokale može uspješno sintetizirati s relativno velikim dijapazonom frekvencije trećeg formanta, a da se ne ugrozi njihov identitet. Da bismo se riješili ove nove, premda moguće varijable, i frekvencija trećeg formanta fiksirana je prema izmjerenim vrijednostima prirodnog izgovora. Upotrijebljene vrijednosti navedene su u tablici 1.

Tablica 1. Frekvencije (kHz) sintetiziranih vokala prema Bakran i Stamenković (1990)

	i	e	a	o	u
F1	0.25	0.48	0.8	0.47	0.26
F2	2.2	1.8	1.2	0.8	0.6
F3	2.7	2.5	2.5	2.5	2.3



Slika 1. Shema frekvencija tranzijentata formanata uz vokal 'a'

Tranzijenti trajanja 40 milisekundi oblikovani su tako da su se sustavno mijenjale početne frekvencije formanata. Početna frekvencija prvog formanta bila je 250 Hz i nije se mijenjala jer je prije zaključeno da ona nema bitnu funkciju u razlikovanju unutar kategorije zvučnih okluziva. Početna frekvencija drugog formanta mijenjanla se u granicama od 600 do 2400 Hz u skokovima po 200 Hz. Početna frekvencija trećeg formanta bila je 1800, 2000, 2200 i 2400 Hz. Frekvencija drugog formanta nikad nije bila viša od frekvencije trećeg formanta.

Ispitanici

Dvadeset i pet izvornih govornika slušalo je sintetizirane slogove, koji su emitirani udobnom glasnoćom i slučajnim redoslijedom. Njihov je zadatak bio za svaki stimulus označiti kojim je konsonantom, /b/, /d/ ili /g/, slog započeo. Odgovor je morao biti upisan bez obzira na stupanj nesigurnosti. Ovaj se postupak razlikuje od Datscheweitova i po načinu kreiranja stimulusa i po zadatku koji je postavljen ispitanicima.

REZULTATI

Rezultati testova slušanja prikazani su u tablicama 2-(i,e,a,o) i dijagramima na slici 2. Prvi stupac predstavlja početnu vrijednost tranzijenta drugog formanta (apscisa dijagrama). Sljedeća tri stupca obilježena sa 1800 (početna frekvencija trećeg formanta) predstavljaju postotak identifikacija /b, d, g/ uz početak tranzijenta trećeg formanta na 1800 Hz. Sljedeće tri kolone odnose se na postotak identifikacije /b, d, g/ uz početnu frekvenciju tranzijenta trećeg formanta 2000 Hz itd.

Tablica 2-i. Postotak identificiranih /b, d, g/ uz vokal /i/.
Podaci se odnose na dijagrame
i-1800, i-2000, i-2200, i-2400.

F2	1800			2000			2200			2400		
	b	d	g	b	d	g	b	d	g	b	d	g
600	92	0	7	78	0	14	85	0	14	92	0	7
800	92	0	7	92	0	7	92	0	7	85	0	14
1000	100	0	0	100	0	0	100	0	0	100	0	0
1200	100	0	0	92	7	0	92	0	7	100	0	0
1400	100	0	0	100	0	0	100	0	0	100	0	0
1600	100	0	0	92	0	7	100	0	0	78	21	0
1800	71	14	14	100	0	0	85	0	14	92	7	0
2000				64	0	35	78	0	21	78	0	21
2200							42	0	57	35	7	57
2400										21	0	78

*Tablica 2-e. Postotak identificiranih /b, d, g/ uz vokal /e/.
Podaci se odnose na dijagrame
e-1800, e-2000, e-2200, e-2400.*

F2	1800			2000			2200			2400		
	b	d	g	b	d	g	b	d	g	b	d	g
600	100	0	0	100	0	0	95	4	0	100	0	0
800	100	0	0	95	4	0	100	0	0	100	0	0
1000	100	0	0	95	0	4	100	0	0	100	0	0
1200	100	0	0	100	0	0	100	0	0	85	14	0
1400	90	0	9	85	14	0	90	9	0	71	28	0
1600	57	4	38	61	4	33	61	28	9	47	52	0
1800	14	23	61	57	9	33	42	19	38	33	38	28
2000				4	9	85	14	4	80	14	14	71
2200							0	0	100	0	0	100
2400										0	0	95

*Tablica 2-a. Postotak identificiranih /b, d, g/ uz vokal /a/.
Podaci se odnose na dijagrame
a-1800, a-2000, a-2200, a-2400.*

F2	1800			2000			2200			2400		
	b	d	g	b	d	g	b	d	g	b	d	g
600	86	0	13	86	0	13	82	0	17	82	0	17
800	82	8	13	86	0	13	82	0	17	95	0	4
1000	86	0	13	91	0	8	86	0	13	91	0	8
1200	73	0	26	73	0	26	86	0	13	78	21	0
1400	30	0	69	43	8	47	33	17	43	17	73	8
1600	0	52	47	4	26	69	4	56	39	0	95	4
1800	0	30	69	0	26	73	4	47	47	4	91	4
2000				4	8	86	0	26	73	0	95	4
2200							0	26	69	0	56	43
2400										0	34	65

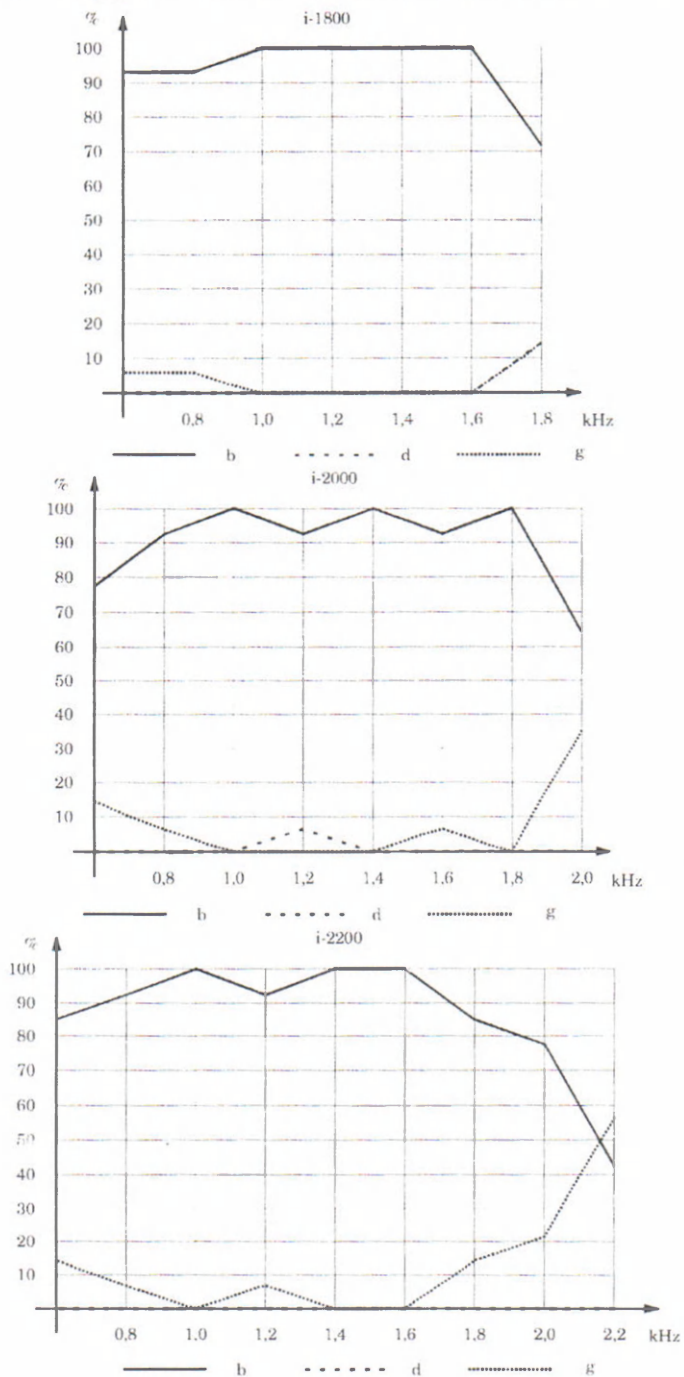
Tablica 2-o. Postotak identificiranih /b, d, g/ uz vokal /o/.
Podaci se odnose na dijagrame
o-1800, o-2000, o-2200, o-2400.

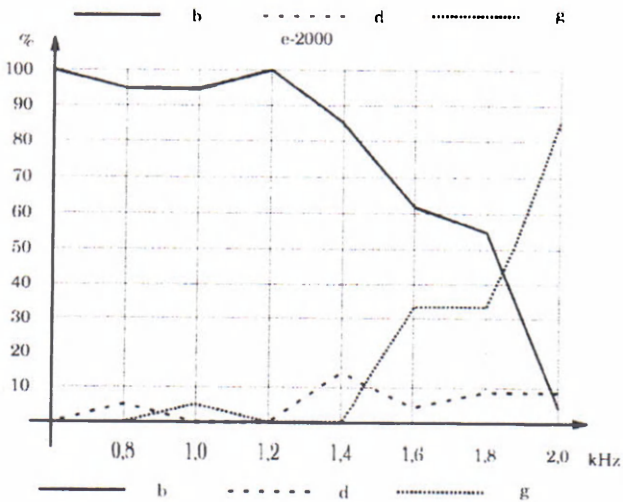
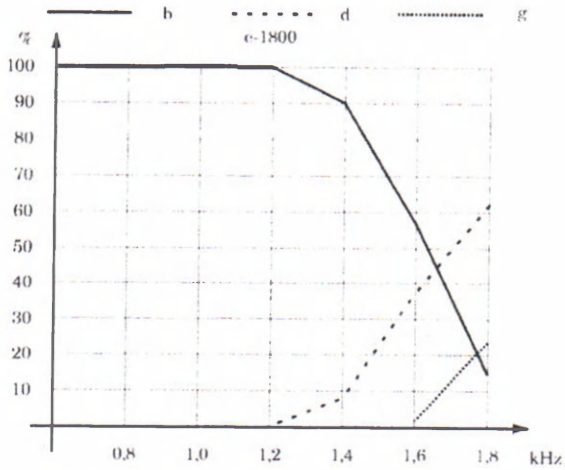
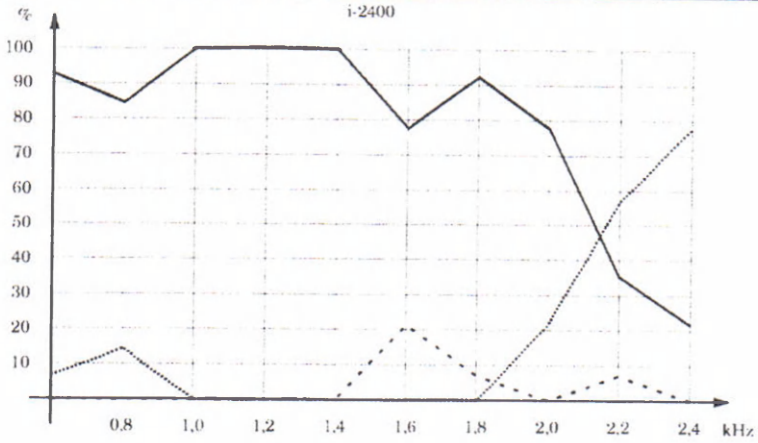
F2	1800			2000			2200			2400		
	b	d	g	b	d	g	b	d	g	b	d	g
600	100	0	0	33	57	9	14	85	0	100	0	0
800	95	4	0	95	0	0	100	0	0	100	0	0
1000	95	0	4	90	4	4	100	0	0	100	0	0
1200	85	9	4	42	47	4	33	52	14	42	52	4
1400	28	61	9	23	71	4	9	85	4	9	90	0
1600	0	100	0	4	95	0	0	100	0	4	95	0
1800	0	100	0	0	100	0	0	100	0	4	95	0
2000				0	100	0	0	100	0	0	95	4
2200							0	100	0	0	100	0
2400										0	100	0

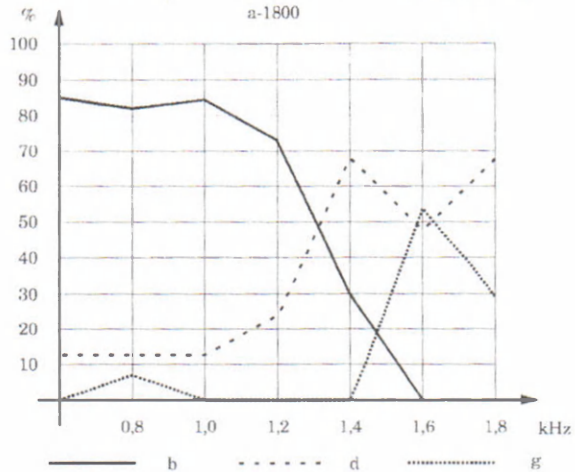
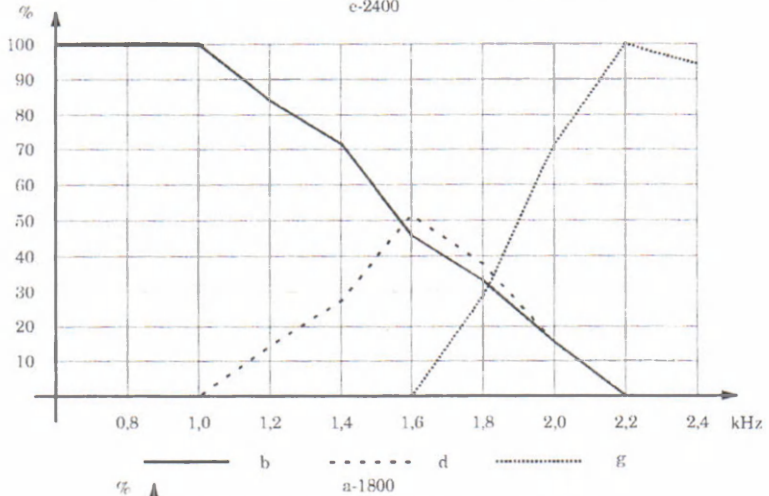
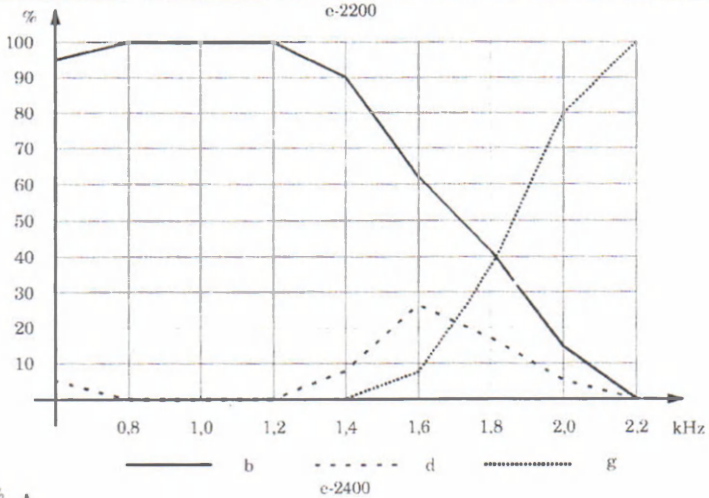
Tablica 2-u. Postotak identificiranih /b, d, g/ uz vokal /u/.
Podaci se odnose na dijagrame
u-1800, u-2000, u-2200, u-2400.

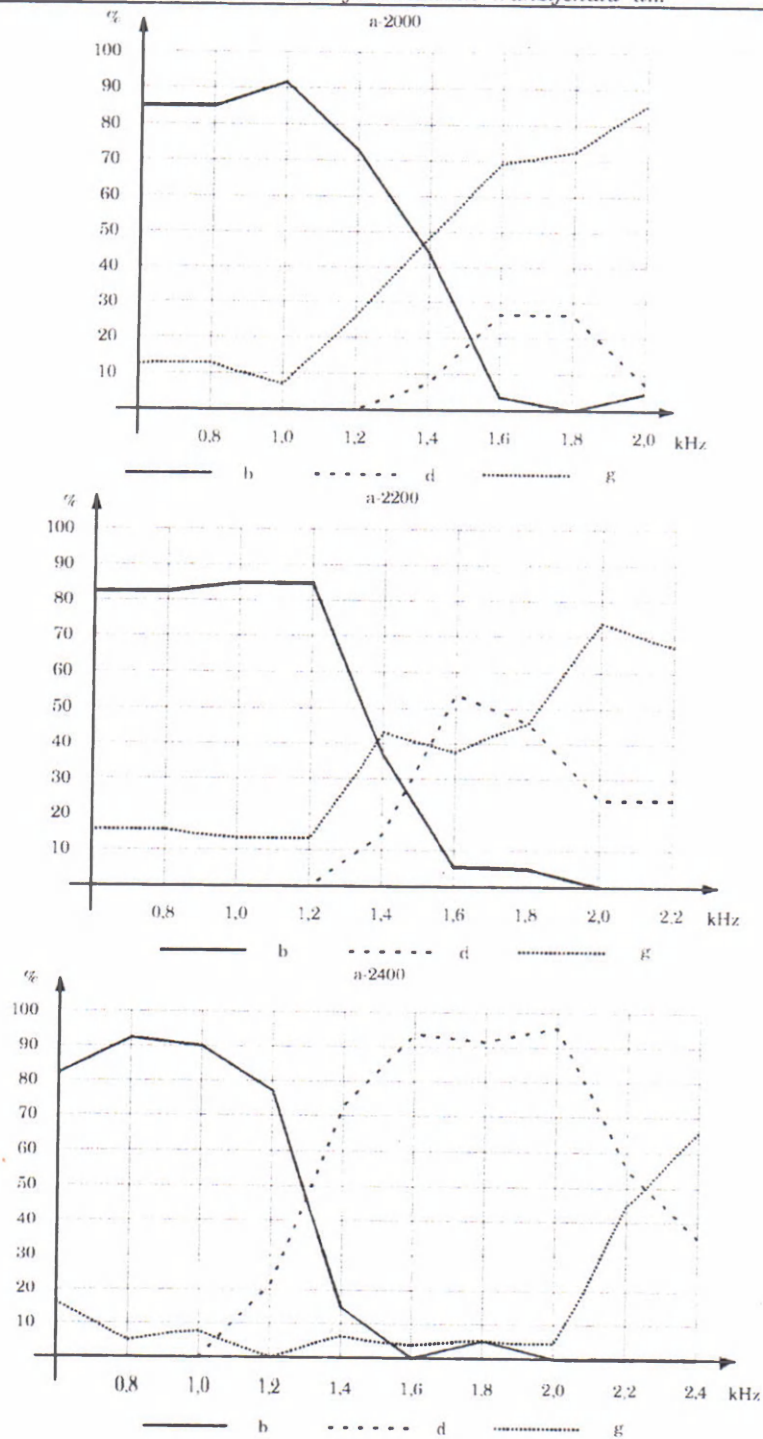
F2	1800			2000			2200			2400		
	b	d	g	b	d	g	b	d	g	b	d	g
600	100	0	0	95	4	0	100	0	0	100	0	0
800	90	0	9	95	0	4	95	0	0	90	0	9
1000	85	0	14	90	0	9	85	0	14	95	0	4
1200	95	0	4	85	4	9	90	4	4	90	4	4
1400	66	28	4	85	9	4	85	14	0	76	23	0
1600	61	38	0	66	33	0	71	28	0	38	61	0
1800	33	61	4	28	71	0	14	85	0	14	85	0
2000				19	80	0	19	80	0	14	85	0
2200							14	85	0	14	85	0
2400										0	100	0

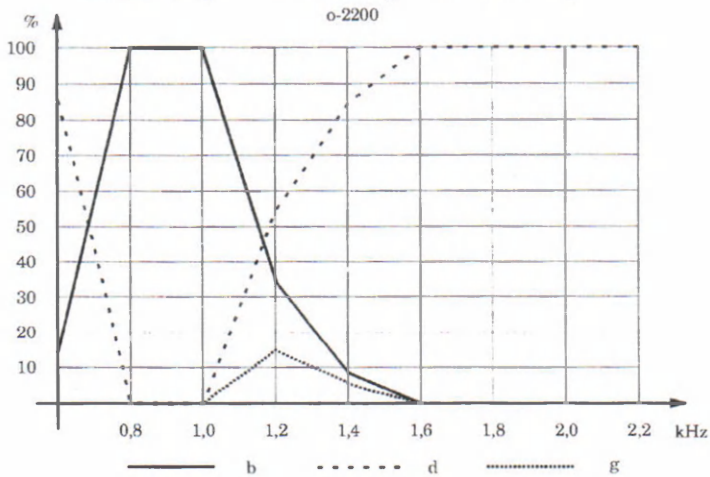
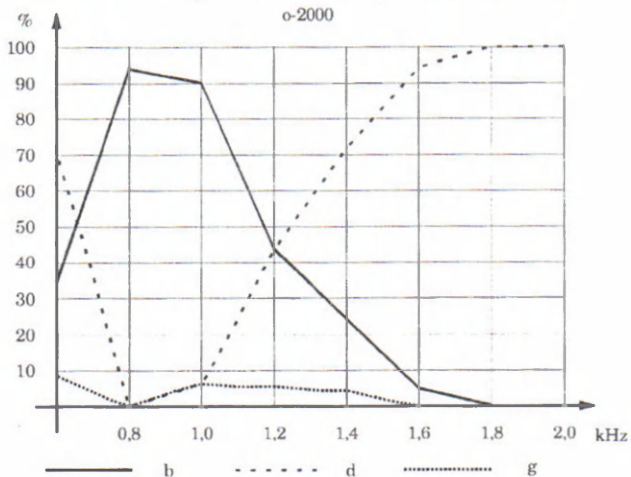
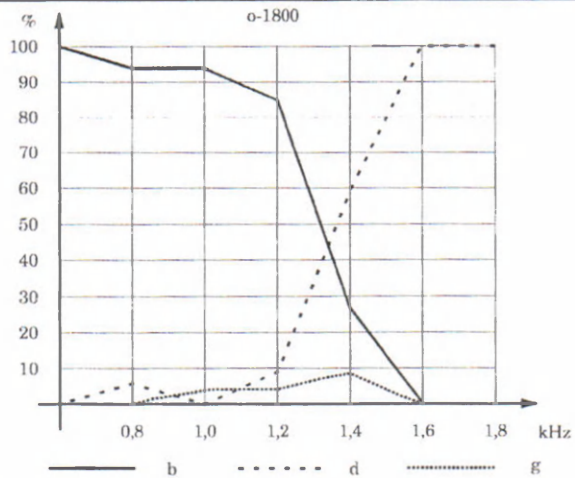
Slika 2. Dijagrami rezultata testova slušanja (podaci iz tablice 2.)

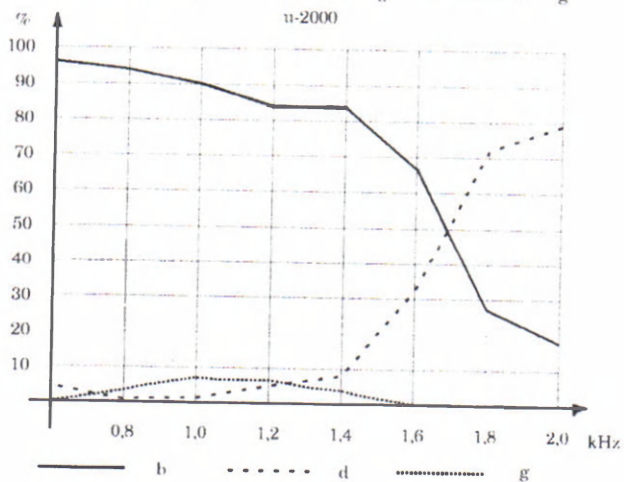
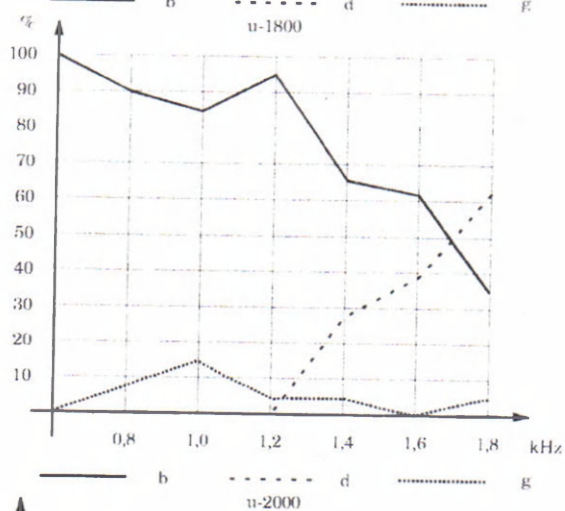
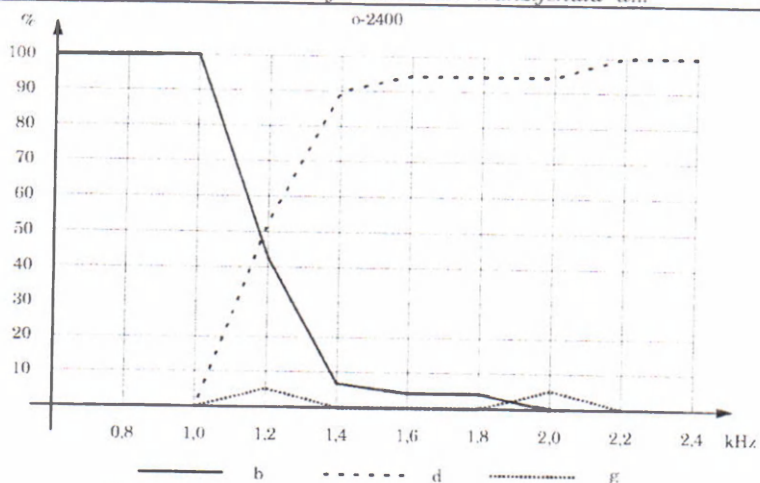


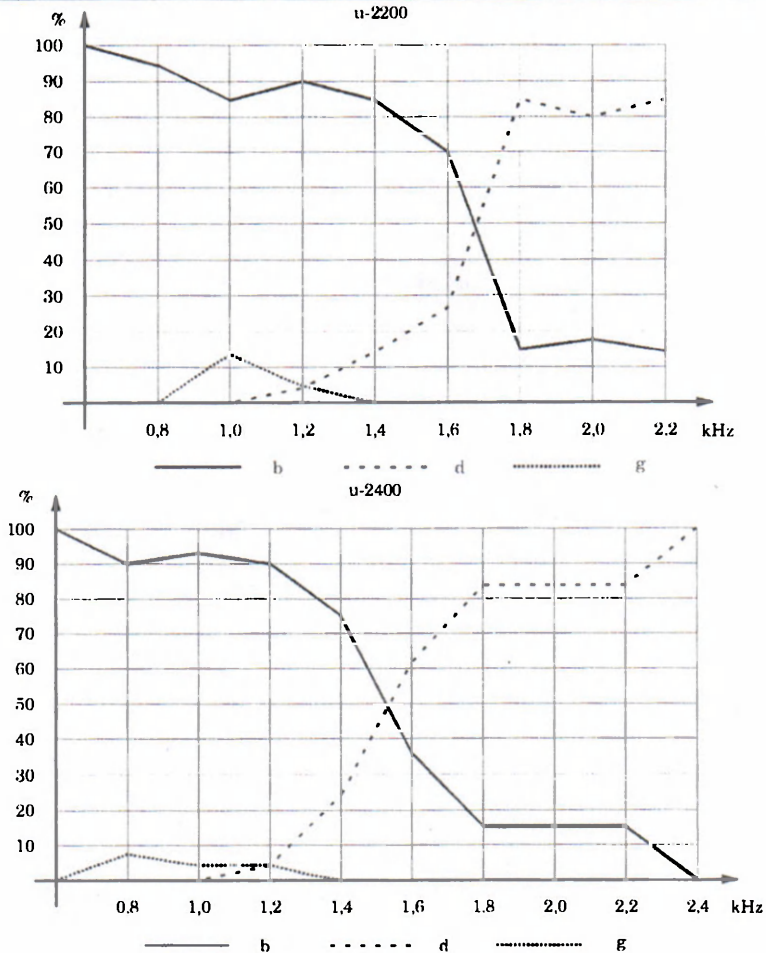












Svaki dijagram prikazuje postotak identifikacije /b, d, g/ s jednom od početnih frekvencija tranzijenata trećeg formanta. Po četiri dijagrama odnose se na isti vokalski kontekst. Puna crta prikazuje postotak odgovora /b/, crtkana se odnosi na /d/, a točkasta na /g/.

Tablica 3. Raspon početnih frekvencija drugog formanta koje maksimalno određuju pojedino mjesto artikulacije

	i	e	a	o	u
b	<1.4	<1.0	<1.0	<1.0	<1.0
d	1.6	1.6	1.6-2.0	<1.4	<1.8
g	2.4	2.4	2.4	1.0	0.8

Vokalski kontekst jasno utječe na identifikaciju /b, d, g/, na način koji su opisali drugi istraživači. U kontekstu /i/ slabo se identificira /d/ neovisno o kombinaciji upotrijebljenih početnih frekvencija formantskih

tranzijenata. Mjesto artikulacije /g/ identificira se u više od 50 posto slučajeva ako je F3 2200 i 2400 Hz, a F2 viši od 2200 Hz. U kontekstu sa /e/ okluziv /d/ identificira se do 50 posto samo kad je F3 2400 Hz i F2 1600 Hz. Identifikacija /g/ u porastu je s višim F3. U kontekstu /a/ sva tri okluziva dostižu znatan stupanj identifikacije uz određenu kombinaciju početnih frekvencija formanata. Kontekst /o/ loše djeluje na identifikaciju /g/, ali se zato /d/ i /b/ dobro identificiraju u različitim zonama F2.

Sjecište krivulja identifikacije /b/ i /d/ leži u području oko 1200 Hz i ne ovisi o položaju trećega formanta. U kontekstu sa /u/, slično kao sa /o/, nijedna od predložениh kombinacija početnih frekvencija formanata nije omogućila identifikaciju /g/. U svakom vokalskom kontekstu /b/ dostiže stupanj identifikacije od 100 posto. Variranje F3 ne utječe na identifikaciju /b/.

Glas /d/ razmjerno se dobro raspoznaje samo u kontekstu sa /a, o, u/. Najizraženiji utjecaj F3 uočava se u kontekstu sa /a/. Viši F3 povoljnije djeluje na prepoznavanje /d/. Prema našim rezultatima identifikaciju /g/ možemo promatrati samo u kontekstu sa /i, e, a/. Najuspješnije identifikacija glasa /g/ postiže se sa F2=F3=2400.

RASPRAVA I ZAKLJUČAK

Iako ne u svim detaljima, rezultati ovih pokusa slažu se s drugim studijama. Datscheweit (1989) tvrdi da za perceptivno odvajanje /d/ i /g/ u kontekstu sa /a/ F3 ima presudnu važnost. Naši rezultati pokazuju da je za njihovo razdvajanje dovoljna promjena F2. Ipak, utjecaj početne frekvencije F3 jasno se uočava u raspoznavanju /g/. Optimalno razdvajanje među kategorijama postiže se sa F3=2400 Hz.

Iz rezultata opisanih pokusa nije moguće izdvojiti pojedinačne vrijednosti koje bi predstavljale frekvencije lokusa kao što su objavili Delattre et al. (1955). Ipak, naši rezultati ne proturječe Delattreovima. Predložene početne frekvencije formantskih tranzijenata ne postižu zadovoljavajući stupanj raspoznavanja /d/ u kontekstu s /i/. Jednako tako ni jedna od upotrebljenih kombinacija frekvencija formanata nije dovela do dobre razumljivosti /g/ u kontekstu sa /u/.

Možemo reći da početna motivacija za ovaj rad – traženje optimalnih početnih frekvencija tranzijenata za sintetiziranje okluziva – nije zadovoljena. To je neočekivan ishod s obzirom na relativno širok dijapazon u kojem su varirane frekvencije formanata. Treba istaknuti da identifikacija okluziva (uz određenu kombinaciju formanata) bitno ovisi o vokalskom kontekstu i ne može se prigovoriti da su odabrane frekvencije izvan opsega kojim bi se postigao veći stupanj identifikacije. To je svakako još jedan dokaz o nepostojanju nepromjenjivih akustičkih znakova (invariant acoustic cues) ili zastupamo posve krivo stajalište kad na opisani način promatramo frekvencije formanata. Jedan od vjerojatnih uzroka niskom stupnju identifikacije jest i nepotpunost slogova stimulusa. To što nedostaje šum eksplozije vjerojatno je najviše djelovalo na lošu identifikaciju /g/, jer on koartikulacijsku prilagodljivost djelomice kompenzira prepoznatljivim šumom. Uz to, kod pripreme stimulusa mijenjale su se samo frekvencije formanata na opisan način, a ostale vrijednosti parametara sintetizatora bile su nepromjenjive, možda neadekvetne.

REFERENCIJE

- Allen, J., M.S. Hannicutt, D.H. Klatt** (1987). *From text to speech*. Cambridge University Press.
- Bakran, J., D. Horga, M. Stamenković** (1992). *Percepcija mjesta artikulacije bezvučnih okluziva*. Govor, VIII/IX, 1991/1992, 31-48.
- Bakran, J., M. Stamenković** (1990). *Formanti prirodnih i sintetiziranih vokala hrvatskoga standardnoga jezika*. Govor VII, 119-138.
- Blumstein, S.E., K.N. Stevens** (1980). *Perceptual invariance and onset spectra for stop consonants in different vowel environments*. J.A.S.A. 67, 648-662.
- Datscheweit, W.** (1989). *Quantitative measurement of the influence of acoustic cues on the perception of voiced plosives*. Proceedings of EUROSPEECH '89, vol.1. 59-62.
- Delattre, P.C., A.M. Liberman and F.S. Cooper** (1955). *Acoustic loci and transitional cues for consonants*. J.A.S.A. 27, 769-773.
- Harris, K.S., H.S. Hoffman, A.M. Liberman, P.C. Delattre, F.S. Cooper** (1958). *Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants*. J.A.S.A. 30, 122-126.
- Klatt, D.H.** (1980). *Software for cascade/parallel formant synthesizer*. J. Acoust. Soc. Am. 67, 971-995.
- Potter, R.A., G.A. Kopp, H.C. Green** (1947). *Visible speech*. Van Nostrand.
- Sharf, D.J. and T. Hemeyer** (1972). *Identification of Place of Consonant Articulation from Vowel Formant Transitions*. J.A.S.A. 51. 652-658.
- Stevens, K.N., S.E. Blumstein** (1978). *Invariant cues for place of articulation in stop consonants*. J.A.S.A. 64, 1358-1368.

Juraj Bakran
Faculty of Philosophy, Zagreb

*FUNCTION OF VOWEL TRANSIENTS IN THE
PERCEPTION OF SYNTHETIC OCCLUSIVES*

SUMMARY

The vowel transient after the occlusion represents the acoustic cue for recognition or for the inter-recognition of the occlusives. CV syllables with voiced occlusives and Croatian vowels have been synthesised so that the beginning frequencies of the first three formants have been systematically changed. The noise of the explosion is not included. Testing of the hearing has proved that the third formant is active in the recognising of the point of articulation, or that it is important for the interrecognition of voiced occlusives, and that, contrary to the beliefs of some, the impact does not appear to be of crucial importance.

Key words: *vowel transients, occlusives, perception of occlusives, acoustic cue, Croatian language*
