

---

UDK 808.62.41:159.9

159.932:808.62

*Izvorni znanstveni rad*

---

*Prihvaćeno 10.04.1997.*

**Vlasta Erdeljac**  
Filozofski fakultet, Zagreb

## **PREPOZNAVANJE RIJEČI U POVEZANOM GOVORU (NA PRIMJERU HRVATSKOG JEZIKA)**

### **SAŽETAK**

*U pokusima provedenim na hrvatskom jezičnom materijalu pokazalo se da je proučavanje fonetsko-fonološke komponente u prepoznavanju riječi osnovica cijeloga sustava razumijevanja jezika, ali i da je nužno u samo razumijevanje uključiti i podatke s viših jezičnih razina.*

***Ključne riječi:*** prepoznavanje riječi, razumijevanje jezika, hrvatski jezik

---

Prepoznavanje izgovorenih riječi, što je osnovni problem kojim se bavi ovaj rad, jest spona između percepcije govora i kognitivnih procesa viših razina. Svrha istraživanja te pojave bolje je razumijevanje načina na koji slušatelji prevode akustički signal u mentalnu predodžbu pohranjenu u svojoj memoriji te određivanje postupaka koji omogućavaju da se tim predodžbama koristimo za razumijevanje. Istraživanje prepoznavanja izgovorenih riječi nužno je interdisciplinarno: ono zahtijeva znanja s područja psihologije, kognitivne znanosti i lingvistike, a također iskustva fonetičara, audiologa i kompjutorskih inženjera.

Prepoznavanje riječi obično se smatra ranijim procesom unutar razumijevanja jezika u skladu sa zamisli da sustav jezičnog razumijevanja obuhvaća više različitih podprocesa (prvo se odlučuje koje riječi uopće dolaze u obzir za prepoznavanje, odnosno jesu li mogući kandidati za razumijevanje, a nakon toga se slaže ukupno značenje lingvističkih podataka). Podprocesi u razumijevanju jezičnih poruka jesu - prepoznavanje riječi (*word recognition*) shvaćeno kao povezivanje senzornog ulaznog podatka s informacijama o riječima pohranjenim u mozgu čovjeka (*low-level perceptual processing*), određivanje rečenične strukture (*parsing*), konstituiranje značenja (*semantic interpretation*) i konačno ostvarenje poruke koju je govornik namjeravao prenijeti (*pragmatic interpretation*) (Garnham, 1985).

Na temelju dosadašnjih istraživanja izdvojilo se mišljenje da je prepoznavanje izgovorenih riječi mnogo složeniji proces nego li se to čini na prvi pogled. Naime, umjesto jednostavnog preslikavanja između akustičko-fonetskih karakteristika i jedinica u mentalnom leksikonu, pokazalo se da se radi o procesu koji sadrži različite razine sužavanja i praćenja kandidata za prepoznavanje te različite komplicirane strategije korigiranja na temelju podataka dobivenih iz prethodnog i naknadnog konteksta u kojem se konkretna riječ pojavljuje.

Dvije su osnovne skupine problema u kojima mora voditi računa svaka teorija prepoznavanja izgovorene riječi u povezanom iskazu - prva skupina problema odnosi se na segmentaciju povezanog iskaza, odnosno određivanje granica među riječima od kojih se iskaz sastoji, a drugi skup problema predstavljaju fonološke varijacije (osobito na granicama pojedinih riječi).

Cini se, doduše, da osim za usko stručno zainteresirane istraživače, prepoznavanje riječi iz povezanoga govora nije posebno zanimljiv problem. Naime, iskustvo slušatelja pokazuje da oni čuju seriju diskretnih, odvojenih riječi, i to u slijedu jednu iza druge, čim su izgovorene. U teoriji prepoznavanja riječi zamisao o sekvencijalnom procesiranju postoji u različitim pristupima prepoznavanju - eksplicitno (Cole i Jakimik, 1979) ili implicitno (Forster, 1976; Morton, 1969, Marslen-Wilson i Welsh, 1978).

No, iskustvo pokazuje da neke riječi mogu biti prepoznate i prije negoli su do kraja izgovorene, a neke se opet ne mogu prepoznati prije početka sljedeće riječi, čime se (opravdano) ugrožava teza o prepoznavanju kao dosljedno "slijeva-nadesno" provedenom procesu (Grosjean, 1985).

Budući da u valnom obliku govornoga stimulusa najčešće nema fizičkih

znakova koji bi konzistentno i pouzdano označavali granice među riječima, može se zaključiti da je segmentacija kontinuiranoga govora i dalje jedan od glavnih problema teorije prepoznavanja izgovorene riječi (Bond i Ganes, 1980).

Činjenica jest da izolirane riječi, za razliku od onih koje se ostvaruju u povezanom govoru, karakterizira razmjerno stabilan akustički oblik koji proizlazi iz pažljivoga artikuliranja, osobito pred mikrofonom, a može biti posljedica važnosti ili informativne vrijednosti riječi, situacijskoga konteksta, govornikove percepcije okoline itd.

U normalnom razgovornom obliku akustički oblik riječi ni približno nema tako preciznu strukturu. Fonetski segmenti prepoznatljivi u izoliranom izgovoru, u konverzaciji mogu biti izmijenjeni, reducirani ili mogu sasvim izostati. (Navedene tvrdnje mogu se ilustrirati usporedbom spektrograma istih riječi izvađenih iz rečenice i izgovorenih izolirano, npr. Bakran, 1996).

Iako je broj mogućih izgovora pojedine riječi u kontinuiranu govoru velik, istina je ipak da je taj broj konačan pa, premda bi to značilo nužno uspostavljanje većeg broja uzoraka za jednu konkretnu riječ, fonološke varijacije unutar same riječi nisu nepremostiv problem za teorije prepoznavanja riječi koje bi se temeljile na uspoređivanju uzoraka.

Problem je u tome da se fonološke najveće promjene događaju upravo na granicama riječi. Dok unutar jedne riječi jezik ne dopušta teško izgovorljive kombinacije fonema (glasova), u povezanom govoru najčešće upravo segmenti na spojevima dviju riječi znače težak ili neprirodan artikulacijski pokret koji se u govoru nužno automatski pojednostavljuje. Rezultat tih pojednostavljenja su promjene, brisanja ili umetanja fonetskih segmenata na granici riječi.

Poznato je da su leksička svojstva riječi u kontekstu u interakciji s različitim izvorima znanja kojima se ubrzava ili usporava proces prepoznavanja (npr. slušateljevo znanje o svijetu i jezičnim pravilima, lingvističke informacije koje odašilje govornik, poznavanje situacije i teme). Treba istaknuti da točna priroda interakcije između unutrašnjih svojstava riječi (uključujući i njihove govorne karakteristike) i ovih izvora znanja ostaje nedovoljno razjašnjena. Trebalo bi razriješiti nesuglasice u vezi s trenutkom u kojem podaci viših jezičnih razina (*top down* informacije) ulaze u proces prepoznavanja (Forster, 1976; Marslen-Wilson i Welsh, 1978), kao i veličinu udjela neposredno prisutnih akustičko-fonetskih informacija.

Dakle, budući da se temelji na percepciji diskretnih elemenata iz kontinuirano promjenjiva signala, riječ u nekom izričaju prepoznaje se tijekom interakcije ulaznog signala i u skladu s ograničenjima koja proizlaze iz svih raspoloživih izvora znanja (od akustičkih, sintaktičkih, semantičkih i logičkih čimbenika do opće tematske strukture i šireg situacijskog konteksta).

Lingvisti se uglavnom slažu u tome da se jezik može opisati s pomoću podsustava kao što su fonologija, morfologija, sintaksa i semantika. Ti podsustavi pretpostavljaju postojanje pravila i reprezentacija. Među psiholingvistima aktualna je opet rasprava o tome prevodi li se takva strukturalna autonomija i u autonomiju procesiranja.



U skladu s jednim mogućim gledištem, s tzv. hipotezom modularnosti, sustav procesiranja jezika sastavljen je od skupa procesnih modula koji su funkcionalno autonomni u tom smislu da modul za određeni ulazni podatak proizvodi isti izlazni podatak (*output*), neovisno o informacijama koje se obrađuju u drugim modulima (Forster, 1979). Modularne teorije suprotstavljaju se interaktivnim modelima za koje je karakteristično da su informacije iz različitih domena dostupne čitavom sustavu razumijevanja (Elman i McClelland, 1984; Marslen-Wilson i Welsh, 1978; McClelland i Rumelhart, 1981 itd.).

Prepoznavanje izgovorene riječi (Cole i Jakimik 1980; Elman i McClelland, 1984; Marslen-Wilson, 1987; Marslen-Wilson i Tyler 1980; Pisoni i Luce 1987) najčešće se postavlja u odnos prema prepoznavanju pisane riječi (Miller et al. 1954; McClelland, 1974; Becker i Killion, 1977; Vellutino, 1982; Taft, 1982; Garnham, 1985). Mnogi modeli prepoznavanja izgovorene riječi razvili su se prilagodbom modela koji opisuju vizualno prepoznavanje na istoj jezičnoj razini (Forster 1976, 1989; Morton, 1969, 1970). Ipak, u prilagodbi tih modela učinjene su važne preinake koje zahtijeva specifičnost govora kao temporalnog fenomena.

Rezultati pokusa koji su prikazani u ovome radu protumačeni su u skladu s "kohortnom" teorijom W. Marslen-Wilsona (Marslen-Wilson et al. 1978; 1980; 1987). Ta teorija predstavlja jedan od tzv. aktivnih modela izravnog pristupa prepoznavanju riječi koji se u procesu prepoznavanja razlaže na dvije faze: autonomnu i interaktivnu. U prvoj, autonomnoj fazi početna akustičko-fonetska informacija ("ulazne" riječi) aktivira u memoriji sve riječi koje imaju isti početak. Sve riječi aktivirane na temelju početne informacije u riječi predstavljaju kohortu. Aktivacija kohorte autonoman je proces u tom smislu da se članovi kohorte odabiru na temelju samo akustičko-fonetske informacije. U toj fazi modela prepoznavanje riječi potpuno ovisi o podacima koji neposredno pristižu s najniže, fonetsko-akustičke razine (*data-driven* ili *bottom-up* proces).

Temelj ovog pristupa jest u zamisli kohorte, zapravo skupa riječi kandidata za prepoznavanje, koji su definirani zajedničkim akustičko-fonetskim osobinama početne glasovne sekvencije riječi. Smatra se da je riječ prepoznata u trenutku kad se ustanovi različitost jedne riječi u odnosu prema svima ostalima s istom početnom sekvencijom. Teorija predviđa olakšavajuću funkciju konteksta, ali za razliku od nekih drugih modela (npr. logogenskog, Morton, 1979) s pomoću konteksta se eliminiraju neadekvatni kandidati kako bi se smanjila veličina kohorte. Da bi model funkcionirao, unutarnja je struktura riječi formalno opisana kao slijed diskretnih segmenata.

Kad je kohorta jednom aktivirana, svi mogući izvori informacija mogu pomoći izboru primjerene riječi. Ovo se načelo provodi dosljedno, i na taj se način zapravo informacije prikupljaju "slijeva nadesno". U ovoj fazi procesa prepoznavanja riječi mogu sudjelovati i informacije više razine tako da neka riječ može biti uklonjena iz kohorte na temelju semantičko-sintaktičkih informacija, pa je proces prepoznavanja interaktivan. Izolacijom samo jedne

riječi iz kohorte završeno je prepoznavanje riječi.

Marslen-Wilsonova kohortna teorija (iako podliježe stanovitim kritičkim primjedbama) posljednjih godina privlači priličnu pozornost najvjerojatnije zbog razmjerno preciznog opisa procesa prepoznavanja, zbog nove ideje da se početnom stadiju prepoznavanja aktiviraju sve riječi iz mentalnoga leksikona koje imaju zajedničke početne akustičko-fonetske informacije s riječju koja je ulazna jedinica te zbog važnosti koja se pridaje početku riječi.

Pokusi prepoznavanja riječi u povezanom hrvatskom govoru provedeni su tzv. *gating* - metodom ili postupkom vremenskog ograničavanja signala, prema Grosjean (1980). Radi se o modificiranom standardnom zadatku percipcijske identifikacije u kojem je istaknutija vremenska priroda govorenog jezika. U *gating* - paradigmi u svakom je pojedinom pokušaju identifikacije (osim naravno u zadnjem) samo dio signala dostupan slušanju. (Primjerice, u prvom pokušaju emitira se samo prvih 30 milisekundi signala, u sljedećem pokušaju emitira se još dodatnih 30, dakle prvih 60 milisekundi, i na taj način dalje dok se ne iscrpi čitavo trajanje signala. U alternativnoj verziji signal se maskira bukom, a samo se pojedini dijelovi eksponiranju nemaskirani.) Ispitanici imaju zadatak identificirati stimulus nakon svakog eksponiranja i procijeniti pritom stupanj sigurnosti vlastita odgovora.

Prednost je ove metode taj da se njome simultano dobivaju podaci o točki izolacije, točki prepoznavanja i o individualnoj procjeni sigurnosti u točnost odgovora. (Točka izolacije onaj je trenutak u trajanju signala u kojem je količina akustičko-fonetskih informacija dovoljna da se samo jedna riječ izdvoji između raspoloživih riječi - kandidata, a svi se ostali kandidati isključuju, a točka prepoznavanja jest onaj trenutak u trajanju signala kad se skupilo dovoljno informacija da se riječ identificira.)

Ove dvije točke nisu nužno simultane (primjerice, za riječ u odgovarajućem kontekstu točka prepoznavanja može prethoditi točki izolacije). Promatranjem razlika u vremenima izolacije i prepoznavanja riječi mogu se proučavati odnosi sustava prepoznavanja riječi i sintaktičkih i semantičkih procesora.

## POKUSI

Pretpostavka koja povezuje sva tri pokusa koji slijede jest ta da se u procesu (identificiranja i) prepoznavanja govorene riječi kontinuirani akustički signal prima i prepoznaje u slijedu na taj način da se s povećavanjem količine akustičkog materijala povećava i stupanj prepoznatosti pojedine riječi (i dijela riječi). To načelo vrijedi i za leksičku razinu, ali i za razinu fonema.

Druga je pretpostavka da se odstupanje od tog temeljnog načela događa u slučajevima dvosmislenosti na fonetsko-akustičkoj razini - onda kad je moguća različita interpretacija glasovnog materijala, koja se odražava u različitim segmentacijama glasovnoga kontinuuma.

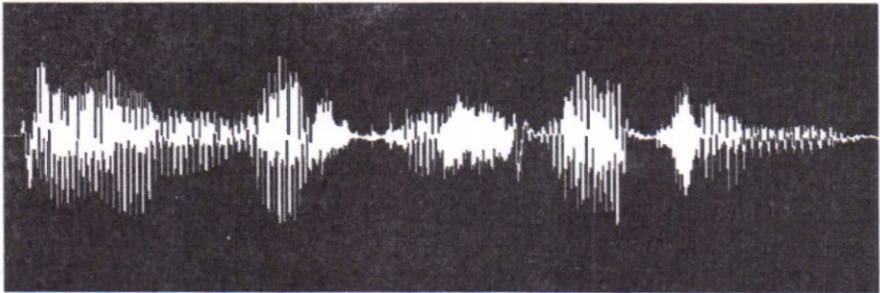
Mjesta koja su problematična za prepoznavanje moguće je djelomice predvidjeti i protumačiti na temelju zvučne slike pojedinih glasovnih segmenata u nizu (npr. slijed vokal-nazal).

### Pokus 1.

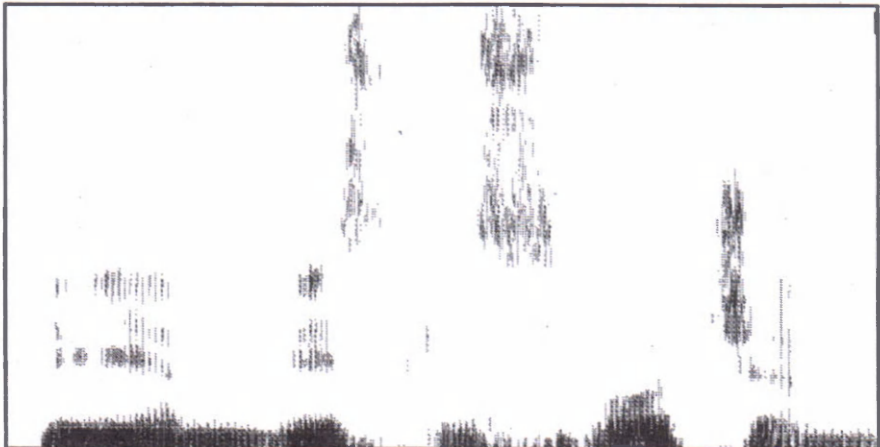
Svrha pokusa jest promatrati vremensku dimenziju prepoznavanja riječi, odnosno utjecaj njezine fonetsko-akustističke organizacije na uspješnost prepoznavanja.

#### Stimulusi

Izričaj "Pijem viski s voćem" izgovoren je muškim glasom, normalnim govnim tempom i snimljen u zvučno izoliranoj prostoriji intonacijom izjavne rečenice. Snimka je digitalizirana 16 bitnim konvertorom s frekvencijom uzimanja uzoraka od 20 kHz i akustički analizirana LSI sustavom (Laughborough Sound Images). Grafički prikaz valnog oblika rečenice prikazan je na slici 1. a spektrografska analiza na slici 2.



Slika 1. Valni oblik rečenice "Pijem viski s voćem."



Slika 2. Spektrogram rečenice "Pijem viski s voćem."

Daljnja priprema stimulusa sastojala se u tome da se izgovorena rečenica pripremi za slušanje prema zahtjevima *gating*-metode. To znači da je izričaj, koji je trajao 1300 milisekundi, podijeljen na vremenski diskretne segmente



od kojih je najkraći vremenski odsječak ili prvi stimulus (1. *gate*) obuhvatio samo prvih 100 milisekundi trajanja signala, a svaki sljedeći sadržao je 20 milisekundi više signala od prethodnoga. To znači da je drugi stimulus trajao 120 milisekundi od početka izričaja, treći 140 milisekundi itd.

## Postupak

Ispitanici, 22 studenta Filozofskog fakulteta, slušali su u normalnim okolnostima slušanja reprodukciju stimulusa redom - od vremenski najkraćeg do najduljeg (dakle, kao zadnji stimulus reproduciran je cijeli izričaj). Vremenski razmak između sukcesivnih stimulusa bio je 5 sekundi. Nakon svakog stimulusa ispitanici su morali zapisati riječ za koju pretpostavljaju da je izgovorena, bez obzira na količinu prezentiranog akustičkog materijala u stimulusu. Uz to, ispitanici su morali označiti stupanj sigurnosti svoje procjene zaokruživanjem, na pripremljenom obrascu, broja (postotka sigurnosti) u pet stupnjeva: 5, 25, 50, 75 i 100%.

## Rezultati i rasprava

Za svaki pojedini stimulus popisane su pretpostavljene riječi tako da je evidentiran broj njihova pojavljivanja u odgovorima ponderiran stupnjem sigurnosti procjene, što je konačno izraženo postotkom koji se smatra pokazateljem prepoznatosti riječi. Dakle, kad bi svi ispitanici sa 100% sigurnosti pretpostavili istu riječ, onda bi ona bila 100% prepoznata prema formuli:

$$100 \frac{\sum ss}{bmp} * maxss$$

gdje je  $\sum ss$  - zbroj stupnjeva sigurnosti za pojedinu riječ

$bmp$  - broj mogućih pojavljivanja (broj ispitanika)

$maxss$  - maksimalni stupanj sigurnosti (100%)

Konkretno to izgleda ovako: u odgovorima na prvi stimulus od moguća 22 ispitanika, troje je pretpostavilo da izgovoreni segment pripada riječi *pivo*.

Od toga, dvoje je svojoj procjeni pridalo 25 postotnu sigurnost, a jedan samo 5 postotnu. Uvrštavanjem u formulu:  $100 (25 + 25 + 5)/22 * 100$ , stupanj prepoznavanja riječi *pivo* iznosi 2,5%.

Posljedica je takva vrednovanja odgovora to da se jednak postotak prepoznatosti dobiva i na temelju dvaju odgovora s po 25% sigurnosti kao i na temelju jednoga odgovora s 50% sigurnosti.

Uzeti u obzir stupanj sigurnosti procjene ispitanika uobičajeno je pri vrednovanju odgovora u pokusima ovoga tipa (Marslen-Wilson, 1987; Grosjean, 1980; Walley et al., 1995).

U 28. stimulusu segment *viski* sadržali su svi predloženi kandidati. Međutim, čini se da se zbog toga što se u akustičkom signalu pojavila i dodatna informacija o fonemu *s*, ponovo došlo do nove raznolikosti predloženih riječi. Budući da su sva tri odgovora na 29. stimulus sadržala segment *s*, (pijem viski *s*, pijem viski *s* ledom, pijem viski *sad*), on se već u tom trenutku može smatrati potpuno prepoznatim.

A	B	C	D
1	100	13	2.3
2	160	11	0.0
3	180	10	0.0
4	200	11	0.0
5	220	5	23.8
6	240	3	53.4
7	260	2	63.6
8	280	2	72.7
9	300	1	86.4
10	320	1	93.2
11	340	1	93.2
12	360	1	93.2
13	380	1	92.1
14	400	3	100*

**Tablica 1.** Postotak prepoznavanja riječi /pijem/ i broj različitih riječi koje ispitanici predlažu kao "kandidate" u ovisnosti o količini akustičkog signala predstavljene rednim brojem stimulusa

A - redni broj stimulusa

B - vrijeme u milisekundama, proteklo od početka signala

C - broj različitih riječi predloženih u odgovorima

D - postotak prepoznatosti izgovorene riječi

A	B	C	D
14	400	3	0
15	420	5	1.1
16	440	5	5.7
17	460	5	6.8
18	480	2	61.4
19	500	1	72.7
20	520	1	77.3
21	540	1	79.5
22	560	1	84.1
23	580	1	88.6
24	600	1	92.0
25	620	1	93.2
26	640	1	97.7
27	660	1	96.6
28	700	4	100 *

**Tablica 2.** Postotak prepoznavanja riječi /viski/ i broj različitih riječi koje ispitanici predlažu kao "kandidate" u ovisnosti o količini akustičkog signala predstavljene rednim brojem stimulusa



A	B	C	D
30	760	4	0
31	800	3	0
32	820	3	0
33	840	5	0
34	860	7	0
35	880	6	0
36	900	6	0
37	920	4	0
38	940	4	0
39	960	4	0
40	980	5	0
41	1000	7	0
42	1020	8	10.0
43	1040	6	5.0
44	1060	5	5.7
45	1080	5	28.4
46	1100	5	30.7
47	1120	5	47.7
48	1140	2	67.0
49	1160	2	68.2
50	1180	1	82.9
51	1200	1	95.4
52	1220	1	92.0
53	1260	1	96.6
54	1320	1	98.9

*Tablica 3. Postotak prepoznavanja riječi /voćem/ i broj različitih riječi koje ispitanici predlažu kao "kandidate" u ovisnosti o količini akustičkog signala predstavljene rednim brojem stimulusa*

Opća primjedba na tumačenje svih tablica jest to da je na onim mjestima gdje stupanj prepoznavanja iznosi 100%, taj postotak dobiven računskim putem, tj. na temelju činjenice da konkretna riječ postoji u svim odgovorima (koji sadrže i pretpostavljenu novu riječ koja slijedi nakon analizirane) tako da samo u tim slučajevima postotak ne odražava procjenu sigurnosti ispitanika, koji se već koncentriraju na iduću riječ, što se vidi iz istodobnog ponovnog povećanja broja predloženih riječi u istom stimulusu (stimulusi broj 14 i 28).

Osim toga, dodatan je komentar nužan za ona mjesta u tablicama gdje se mogu opaziti vremenski skokovi veći od 20 milisekundi. Na tim mjestima akustički signal bio je u toj mjeri nepromijenjen da nije bilo novih informacija za prepoznavanje riječi.

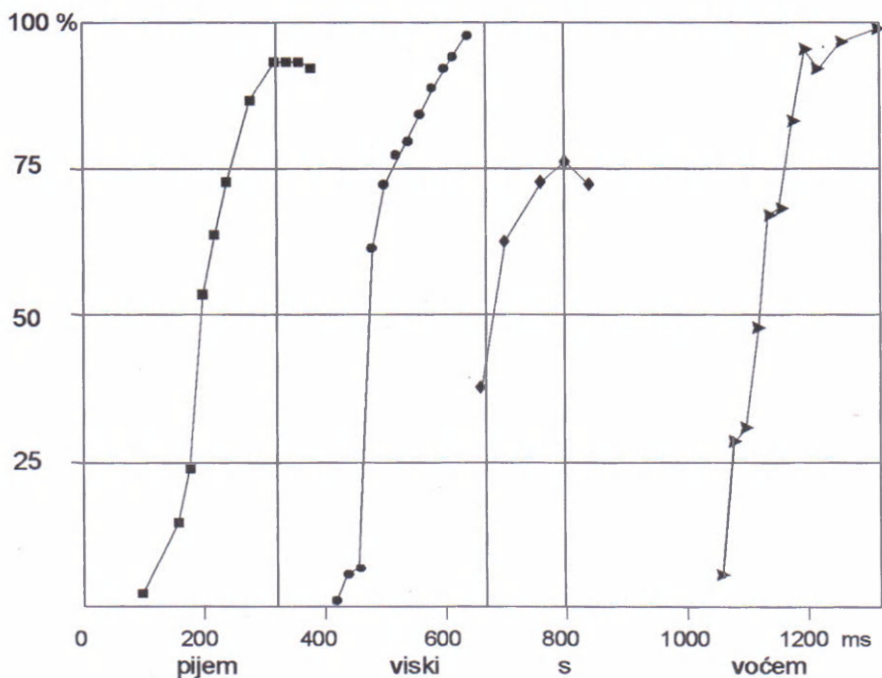
Prvi zaključak koji se na temelju ovih tablica može izvesti potvrđuje pretpostavku da se s porastom trajanja stimulusa u načelu smanjuje broj riječi predlaganih u odgovorima, dok se istodobno povećava sigurnost, odnosno postotak prepoznatosti jedne od izgovorenih riječi. Te tendencije nemaju uvijek

kontinuiran uzlazni, ili silazni tijek, no ta se odstupanja moraju pripisati razmjerno malom broju ispitanika.

Promatra li se stupac koji pokazuje broj različitih predloženih riječi u odgovorima, u prve dvije tablice uočava se potpuna pravilnost smanjivanja broja predloženih riječi. U trećoj tablici, koja prikazuje tijek prepoznavanja riječi *voćem*, to načelo vrijedi tek nakon trenutka u kojem se prvi put među odgovorima pojavi izgovorena riječ (od stimulusa 42).

Ovaj se pomak može objasniti razmjerno neočekivanim slijedom riječi u konkretnoj sintagmi *viski s voćem*, odnosno njihovom nepredvidljivom semantičkom povezanošću. Kasno prepoznavanje riječi *voćem* događa se unatoč uspješnoj identifikaciji fonema od kojih se riječ sastoji. To možemo zaključiti iz fonemskog sastava predlaganih riječi prije negoli se među odgovorima pojavila riječ *voćem*.

Dakle, uvidom u pretpostavljeni skup riječi za one stimulse u kojima su postojale akustičke informacije o slijedu fonema *viski s vo* -... može se zaključiti da su riječi *vodom*, *votkom*, *voćnim*, ... *svojevoljno*, *svojski* kao mogući asocijacijski povezani nastavci kombinacije ...*viski s* ... otežale izoliranje i prepoznavanje stvarno izgovorene riječi *voćem*.



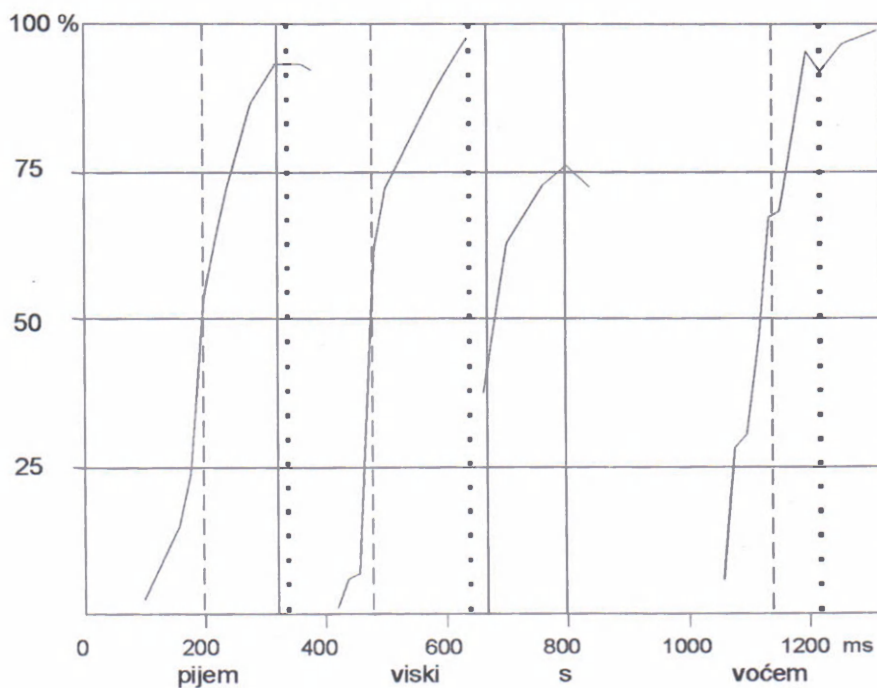
Slika 3. Grafički prikaz uspješnosti prepoznavanja /pijem viski s voćem /

U tablicama 1-3 postotak prepoznavanja izgovorenih riječi povezan je s rednim brojem stimulusa. Isti su podaci prikazani grafički tako da je na apscisi dijagrama označeno vrijeme, trajanje pojedinih stimulusa, a na ordinati postotak prepoznavanja pojedine riječi (Sl. 3).

Na apscisi dijagrama osim vremena u milisekundama (0 do 1200), označeno je (okomitom crtom) i trajanje pojedinih riječi određeno akustičkom analizom prema fonetskim kriterijima (spektrogram na slici 2).

Dobiveni rezultati mogu se interpretirati u skladu s već poznatim postavkama u vezi s procesiranjem prepoznavanja riječi - u prvom se redu može potvrditi zamisao o "kohortnoj" proceduri u prepoznavanju govorenih riječi (Marslen - Wilson et al. 1980).

U tablicama 1-3 broj različitih riječi predloženih u odgovorima opisuje, odnosno izravno odražava veličinu "kohorte" za svako pojedino trajanje akustičkog signala. Posve je očito, promatra li se zasebno svaka riječ, da se s povećanjem trajanja signala, i to onda kad dulje trajanje nosi nove akustičke informacije, smanjuje broj mogućih riječi - kandidata. (Produljivanje trajanja okluzije ili frikativnog šuma ne nosi dodatnu informaciju, što se vidi u tome da se za to vrijeme veličina skupa potencijalnih riječi ne mijenja.)



**Slika 4.** Točka izolacije (crtkana crta) i točka prepoznavanja (točkasta crta) za riječi /pijem/, /viski/ i /voćem/



U ovdje predloženom načinu opisivanja rezultata prepoznavanje riječi prati se tijekom cijelog trajanja (a ne u nekoliko odabranih točaka) pa se može uočiti karakteristična "S" krivulja za svaku riječ. Iste rezultate možemo prikazati prema Grosjeanovu prijedlogu (Grosjean, 1980) definiranjem dviju karakterističnih točaka u procesu prepoznavanja, određivanjem točke izolacije i točke prepoznavanja (Sl. 4).

Dakle, vidi se da se za riječ *pijem* točka izolacije (trenutak) nalazi 200 milisekundi od početka riječi, a točka prepoznavanja 140 milisekundi kasnije. Riječ *viski* "izolirana" je 160 milisekundi nakon početka riječi, a prepoznata je još 160 milisekundi kasnije. Međutim, riječ *voćem* izolira se razmjerno kasno, tek 340 milisekundi od početka riječi, ali se onda prepoznaje samo 80 milisekundi kasnije. Tako kasni trenutak izolacije riječi *voćem*, objašnjen je u komentaru tablice 3.

Grafički prikaz čitavog tijeka prepoznavanja pojedinih riječi korelira s prikazom s pomoću dviju karakterističnih točaka.

## Pokus 2.

U ovom pokusu namjerava se pokazati djelovanje konteksta koji neposredno slijedi traženu riječ na njezino prepoznavanje.

Zapravo, očekuje se da će rezultati pokusa ilustrirati proces određivanja granica riječi i posljedice koje ostavlja otežana delimitacija akustičkog signala na proces identificiranja i prepoznavanja riječi.

Poznati problem segmentiranja akustičkog signala u ovom je primjeru posebno istaknut naročitim odabirom niza riječi. S obzirom na to da u akustičkom signalu rijetko postoji nedvosmislen znak za granicu među riječima, odabrana je upravo takva rečenica u kojoj je segmentacija izrazito problematična. U rečenici "Vidim prstenast oluk" niz fonema od kojih se sastoji sintagma *prstenast oluk* može se segmentirati, u skladu sa zahtjevima (ili mogućnostima) korištene *gating* - metode, na jedanaest smislenih načina:

prst ...  
 prste ...  
 prsten ...  
 prstena ...  
 prste na ...  
 prsten na ...  
 prstenast ...  
 prstenasto ...  
 prste na sto ...  
 prste na stolu ...  
 prsten na stolu ...  
 prstenast oluk ...

U odgovorima ispitanika sve su se ove mogućnosti segmentiranja pojavile osim *prste na sto ...* (postoji mogućnost npr. - *isprepliće prste na sto načina*), što je ujedno najmanje vjerojatna kombinacija.

U slučaju kad veza između akustičkog signala (odnosno načina segmentiranja) i značenja nije dvosmislena, produljavanje akustičkog signala vodi do sve točnijeg prepoznavanja. To su pokazali rezultati prvoga pokusa (prepoznavanje riječi u rečenici *pijem viski s voćem*).

Višestruke mogućnosti određivanja granica među riječima manifestiraju se u prikazu rezultata kao postupno povećanje sigurnosti (koje nikada ne doseže maksimalan stupanj prepoznavanja) te obvezatno više ili manje postupno smanjenje stupnja sigurnosti jedne riječi i simultan rast sigurnosti za drugu riječ.

A	B	C	D
16	120	14	0.2
17	140	13	4.5
18	160	12	7.9
19	180	14	9.1
20	200	8	21.6
21	220	4	38.6
22	240	3	40.9
23	260	3	51.5
24	340	2	55.7
25	360	2	59.1
26	370	2	63.6
27	380	3	7.9
28	400	3	4.5
29	420	3	0.0*

**Tablica 4.** Tijek segmentiranja i stupanj sigurnosti odabira riječi /prst/ na temelju povećanja trajanja akustičkog signala

A - redni broj stimulusa

B - vrijeme eksponiranog signala od početka riječi *prstenast*

C - broj različitih predloženih riječi

D - stupanj sigurnosti odabira jedne riječi

Tablice u kojima su navedeni rezultati dobiveni u ovom pokusu jednim su svojim dijelom podudarne s tablicama 1, 2 i 3, tamo gdje prikazuju paralelno povećanje količine signala i smanjivanje broja riječi - kandidata za prepoznavanje. Razlika među njima jest u tome što se u tablicama 1, 2 i 3 opisuje sigurnost prepoznavanja stvarno izgovorene riječi, a u ovom se slučaju radi o stupnju sigurnosti odabira pretpostavljene riječi (u prvom primjeru, u tablici 4. riječ *prst*) kao jedne od mogućnosti segmentiranja tekućeg akustičkog signala.

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
18	160	12	0.2
19	180	14	0.2
20	200	8	1.1
21	220	4	5.7
22	240	3	14.7
23	260	3	13.6
24	340	2	15.9
25	360	2	19.3
26	370	2	17.0
27	380	3	47.7
28	400	3	45.4
29	420	3	64.7
30	460	2	73.8
31	480	3	80.7
32	500	2	94.3
33	520	2	95.4
34	540	7	12.5
35	560	5	0.0*

**Tablica 5.** Tijek segmentiranja i stupanj sigurnosti odabira riječi /prsten/ na temelju povećanja trajanja akustičkog signala

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
21	220	4	3.4
22	240	3	3.4
23	260	3	3.4
24	340	2	0.0
25	360	2	0.0
26	370	2	0.0
27	380	3	18.2
28	400	3	5.7
29	420	3	11.4
30	460	2	11.4
31	480	3	7.9
32	500	2	0.0*

**Tablica 6.** Tijek segmentiranja i stupanj sigurnosti odabira riječi /prste/ na temelju povećanja trajanja akustičkog signala



A	B	C	D
31	480	3	3.4
32	500	2	3.4
33	520	2	2.3
34	540	7	25.2
35	560	5	40.9
36	580	6	40.9
37	620	8	38.7
38	640	10	20.4
39	660	7	14.9
40	720	6	14.7
41	800	6	11.4
42	820	5	12.5
43	840	7	15.9
44	860	6	15.9
45	880	7	26.1
46	920	7	18.2
47	940	6	19.3
48	980	9	18.2
49	1000	8	20.5
50	1020	5	20.5
51	1040	4	23.8
52	1060	3	23.8
53	1100	3	27.3
54	1200	4	29.5
55	1220	5	22.7
56	1240	9	27.2
57	1260	8	25.0
58	1340	7	21.0

*Tablica 7. Tijek segmentiranja i stupanj sigurnosti odabira riječi /prsten na/ na temelju povećanja trajanja akustičkog signala*

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
34	540	7	2.3
35	560	5	4.5
36	580	6	6.8
37	620	8	12.5
38	640	10	7.0
39	660	7	14.7
40	720	6	14.5
41	800	6	15.7
42	820	5	17.0
43	840	7	21.6
44	860	6	25.0
45	880	7	16.0
46	920	7	27.2
47	940	6	29.4
48	980	9	30.6
49	1000	8	33.1
50	1020	5	36.4
51	1040	4	42.0
52	1060	3	52.3
53	1100	3	64.7
54	1200	4	55.7
55	1220	5	57.9
56	1240	9	46.6
57	1260	8	31.9
58	1340	7	22.7

**Tablica 8.** Tijek segmentiranja i stupanj sigurnosti odabira riječi /prste na/ na temelju povećanja trajanja akustičkog signala

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
34	540	7	6.8
35	560	5	15.9
36	580	6	13.6
37	620	8	11.3
38	640	10	20.4
39	660	7	36.3
40	720	6	43.1
41	800	6	43.1
42	820	5	37.5
43	840	7	31.8
44	860	6	21.8
45	880	8	17.1
46	920	7	18.2
47	940	6	13.7
48	980	9	14.8
49	1000	8	7.9
50	1020	5	9.1
51	1040	4	9.2
52	1060	3	6.8
53	1100	3	4.5
54	1200	4	7.9
55	1220	5	5.6
56	1240	9	7.0
57	1260	8	10.4
58	1340	7	14.7

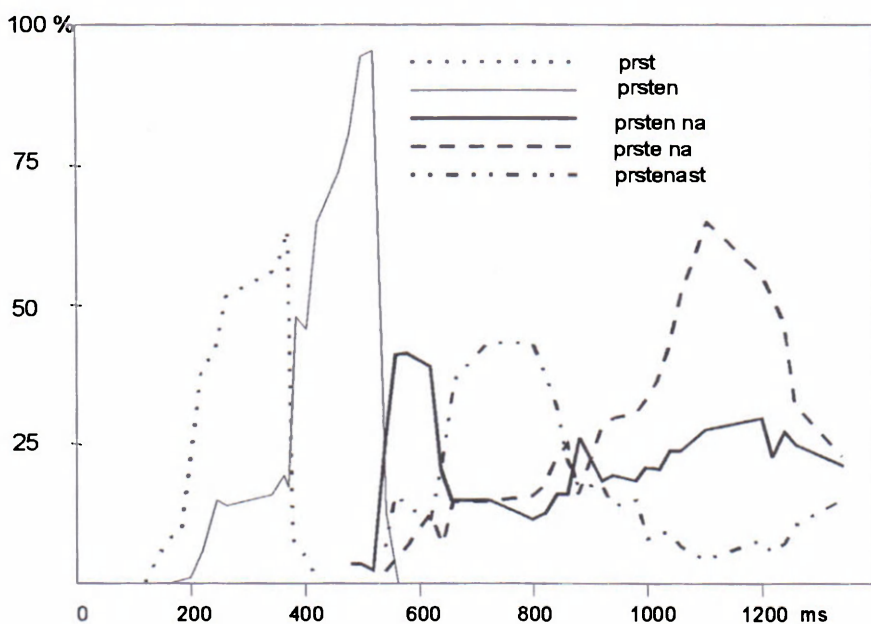
*Tablica 9. Tijek segmentiranja i stupanj sigurnosti odabira riječi /prstenast/ na temelju povećanja trajanja akustičkog signala*



Riječ *oluk* prvi se put pojavljuje tek u 54. stimulusu, sa 3.4% sigurnosti, a maksimalna dosegnuta sigurnost za tu riječ bila je samo 14.7%, i to u trenutku kad je emitiran cjelokupan akustički signal. Dakle, prema opisanim kriterijima za definiranje točke izolacije i točke prepoznavanja riječ *oluk* nije dosegla nijednu od ovih točaka.

Po ovom se primjeru vidi da je u slučaju kad akustička ili tzv. *bottom up* informacija nije jednoznačna nego ostavlja mogućnost za različite fonološke interpretacije, prepoznavanje riječi otežano (usporeno) ili se izgovorena riječ čak i ne može prepoznati.

Pretpostavlja se da su u takvim situacijama nužni i uključuju se u prepoznavanje podaci s viših jezičnih razina (rečenični kontekst, značenjske veze itd.).



**Slika 5.** Tijek segmentiranja i stupanj sigurnosti odabira riječi na temelju povećanja trajanja akustičkog signala

Zanimljiva je usporedba grafova na sl. 5. i sl. 3. koja je moguća ako se postotak sigurnosti prepoznavanja izgovorene riječi promatra paralelno sa stupnjem sigurnosti odabira pojedine pretpostavljene riječi.

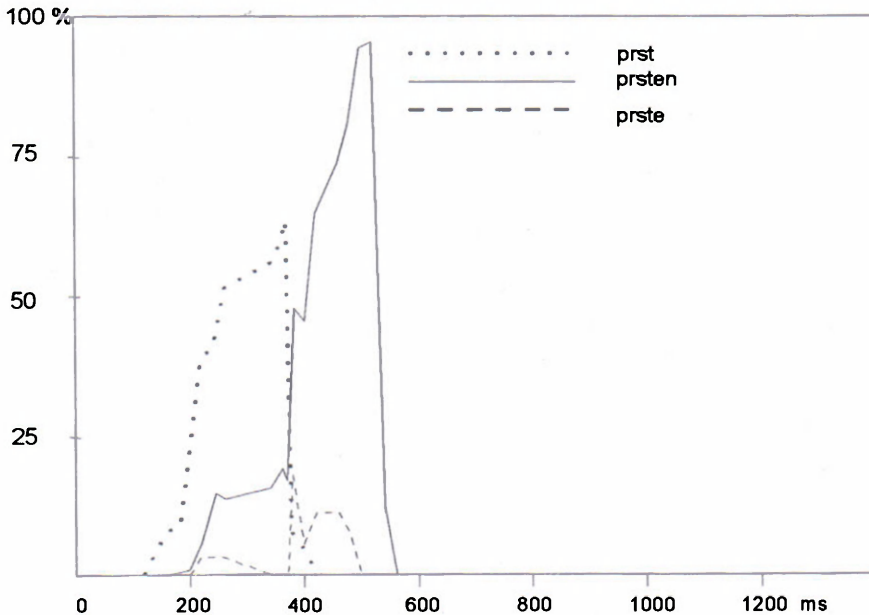
Najočitiji je podatak koji se može očitati sa sl. 5. taj da se tijekom prepoznavanja sintagme *prstenast oluk* ili glasovnog niza: *prstenastoluk*, ne ponavlja pretpostavljeni "S" oblik tijeka prepoznavanja.

Naime, u prvom je slučaju posljednja točka svake krivulje ujedno i maxi-

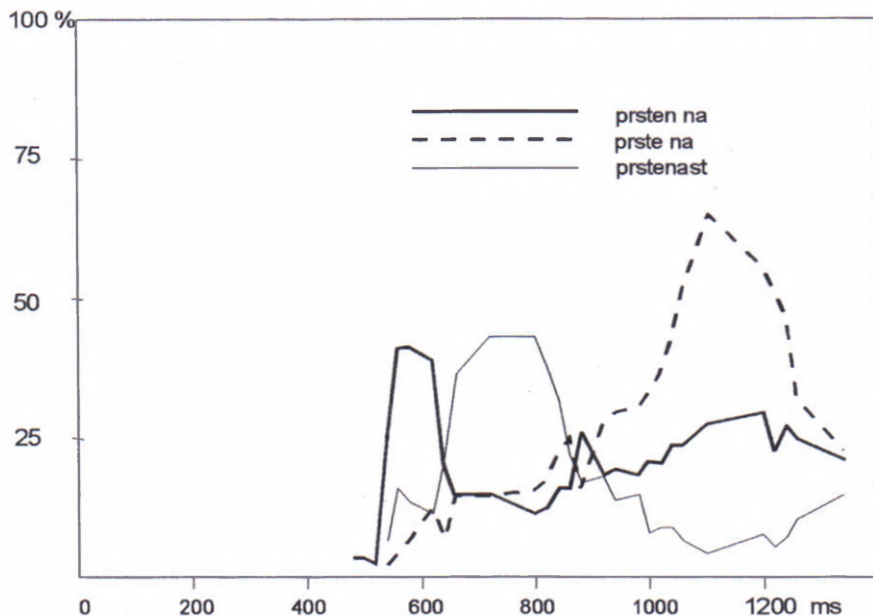
malan stupanj ispitaničove procjene sigurnosti, a na drugoj je slici vidljiv i više-manje postupan pad njihove sigurnosti u vlastitu procjenu identiteta odabrane riječi uz istodoban rast sigurnosti u korist nove procjene, tj. nove riječi.

Osim toga, krivulje sigurnosti za pojedine riječi prikazane na ovoj slici međusobno se u znatnoj mjeri isprepleću, što znači: dok dio ispitanika pretpostavlja da se radi o jednoj riječi, drugi dio ispitanika pretpostavlja da se radi o nekoj drugoj riječi ili riječima (prst - prsten; prste na - prsten na).

Vremenski tijek identificiranja mogućih riječi u ponuđenom glasovnom nizu može se interpretirati na dvije razine - jedna je intraleksička, a druga interleksička. Naime, sve dotle dok se na temelju fonetsko-akustičkog materijala može pretpostavljati da je tekući glasovni signal jedna riječ (temeljna ili neki od mogućih izvedenih oblika kao u slučajevima - *prst*, *prste*, *prsten*), stupanj sigurnosti odabira sl. 5. i stupanj sigurnosti prepoznavanja sl. 3. teku uglavnom podudarno. Naime, stupnjevi sigurnosti i u jednom i u drugom slučaju mogu biti vrlo visoki i ne pokazuju znatnije oscilacije u vremenu u tom smislu da s dodavanjem akustičkog signala tijekom vremena raste stupanj sigurnosti za pretpostavljenu riječ, a u trenutku kad pristigne informacija koja upućuje na drugu mogućnost, od prve se odustaje i slijedi isti postupak za sljedeće predviđanje.



Slika 6. Postotak sigurnosti odabira riječi /prst/, /prsten/, /prste/



Slika 7. Postotak sigurnosti segmentiranja na: /prsten na/, /prste na/, /prstenast/

Na interleksičkoj razini određuju se granice u govornom kontinuumu tako da rezultat bude ili nekoliko riječi koje sačinjavaju jednu sintagmu ili jedna cjelovita riječ koja nužno obuhvaća isti glasovni sastav (prste na, prsten na, prstenast). Veći broj mogućih kombinacija s istim akustičkim materijalom (a bez ograničavajućeg konteksta više razine) očituje se dvojako - kao nesigurnost u procjeni identiteta pojedinih pretpostavljenih riječi (što se vidi po tome da je maksimalna sigurnost procjene bila samo - 40,9% za *prsten na*, 64,7% za *prste na*, 43,1% za *prstenasti*), a oscilacije u tim procjenama u tijeku vremena očituju se u tome da se dominantan položaj jedne od varijanata u odnosu prema drugim dvjema može pratiti u vremenskom slijedu signala, ali na kraju, nakon zadnjeg odslušanog stimulusa, ostaju sve tri mogućnosti segmentacije u ravnopravnom odnosu sa 21,0%, 22,7% i 14,7% na skali procjene sigurnosti. To je, dakle, razlog zbog kojega izgovorena riječ nije prepoznata.

Analiza akustičkoga signala može pridonositi pojašnjavanju razloga nesigurnosti odluka ispitanika.

### Pokus 3.

Treći pokus jest varijanta drugog pokusa utoliko što je izveden na istom govornom materijalu, a razlikuje se od njega po različitoj uputi - zahtjevu koji je postavljen pred ispitanike.

Dok se u prethodnom pokusu inzistiralo na tome da ispitanici odgovaraju

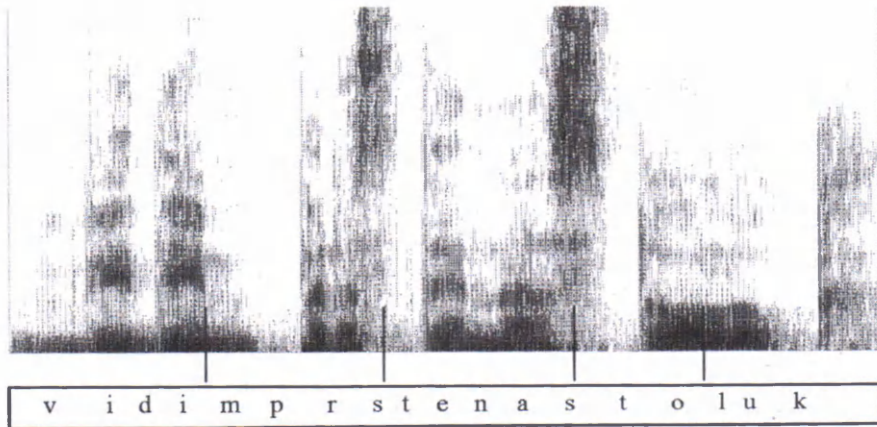
uvijek čitavom riječju bez obzira na nedovoljnu količinu emitiranog akustičkog signala, u ovom je pokusu zahtjev bio da kao odgovor upišu samo one foneme koje su u emitiranom akustičkom signalu identificirali. Dakle, u prethodnom pokusu ispitanici su mogli (morali) pogodati kojoj riječi pripada emitirani dio signala, a ovdje su smjeli navesti samo foneme za koje su sigurni da su ih čuli.

Očekuje se da će pokazatelji o tijeku identifikacije/prepoznavanja pojedinih fonema u fonetsko-akustičkom kontinuumu potvrditi temeljnu tezu da prepoznavanje ovisi o duljini (vremenskom tijeku) govornog signala.

Pretpostavlja se, osim toga, da će zapažanja o odnosu spektralnog oblika signala i procjene identiteta fonema omogućiti bolji uvid u procese prepoznavanja čitave riječi u kontinuiranom govornom iskazu.

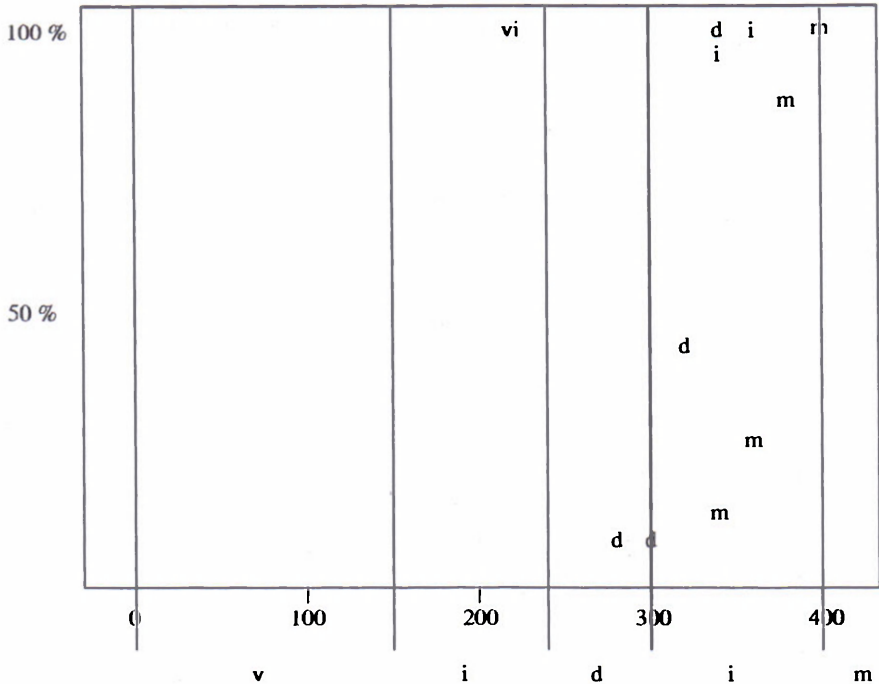
U ovom je pokusu sudjelovala druga skupina studenata ( $N = 23$ ) koja nije sudjelovala u prijašnjim *gating*-pokusima. Govorni niz i stimulusi koji su od njega pripremljeni jednaki su testiranom materijalu iz prethodnog, drugog pokusa: *Vidim prstenast oluk*.

Odgovori ispitanika bit će prikazani na slikama 9. do 13., a povezani su s odgovarajućim spektrogramima. Na taj se način može pratiti tijek identifikacije pojedinih fonema. Apscisa dijagrama na tim slikama razvučena je tako da je snimljena rečenica razlomljena na 5 segmenata. Granice pojedinih segmenata označene su na sl. 8.



**Slika 8.** /vidim prstenast oluk/ s granicama pojedinih segmenata prikazanih na slikama 9. do 13.





**Slika 9.** Postotak prepoznavanja pojedinih fonema povezan s vremenskim tijekom signala: segment /vidim/...

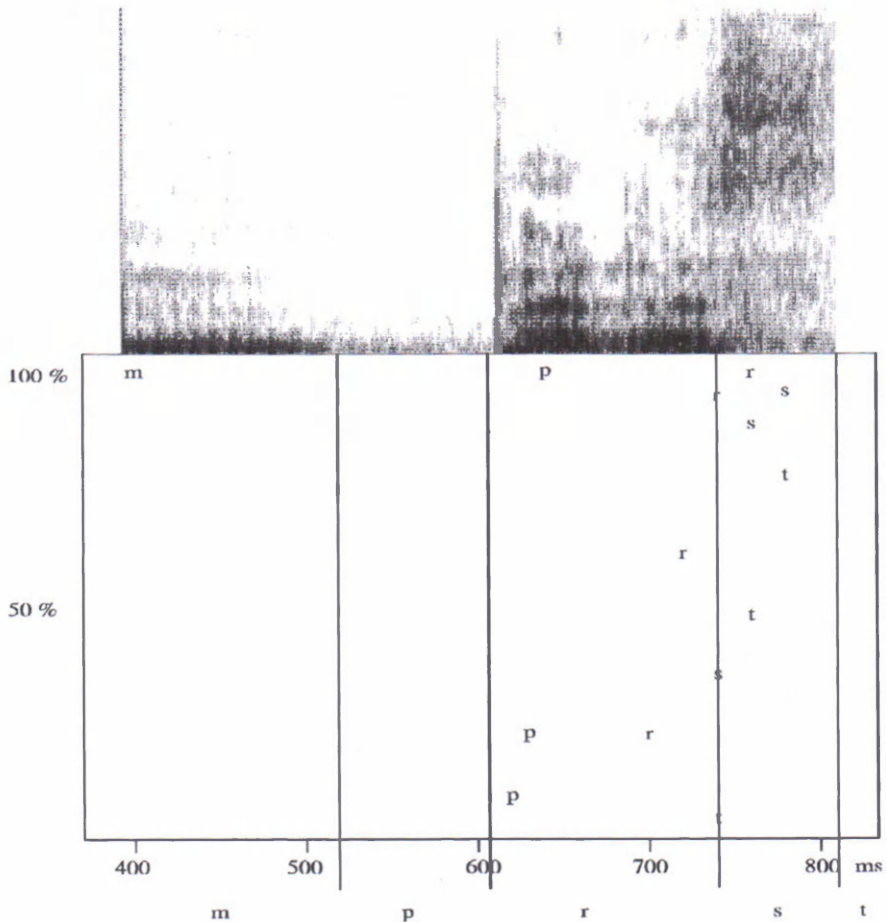
Na slici 9. vidi se da je prvi slog *vi* (ujedno i prvi stimulus) identificiran odmah nakon prezentiranja prvog zvučnog segmenta/odsječka. Za vrijeme trajanja okluzije samo je 8.7% ispitanika identificiralo *d*, a nakon eksplozije *d* kao odgovor na isti signal s razmjerno visokim postotkom identificiraju se *d* i *i*, a istodobno se u odgovorima pojavljuje i *m*.

Identifikacija *m* dosiže maksimum i prije negoli u akustičkom signalu postoje vidljivi znaci oralne okluzije. To znači da u onom dijelu akustičkog signala u kojem prepoznamo karakterističan oblik spektra *i* već postoji dovoljno informacija o nazalu koji slijedi.

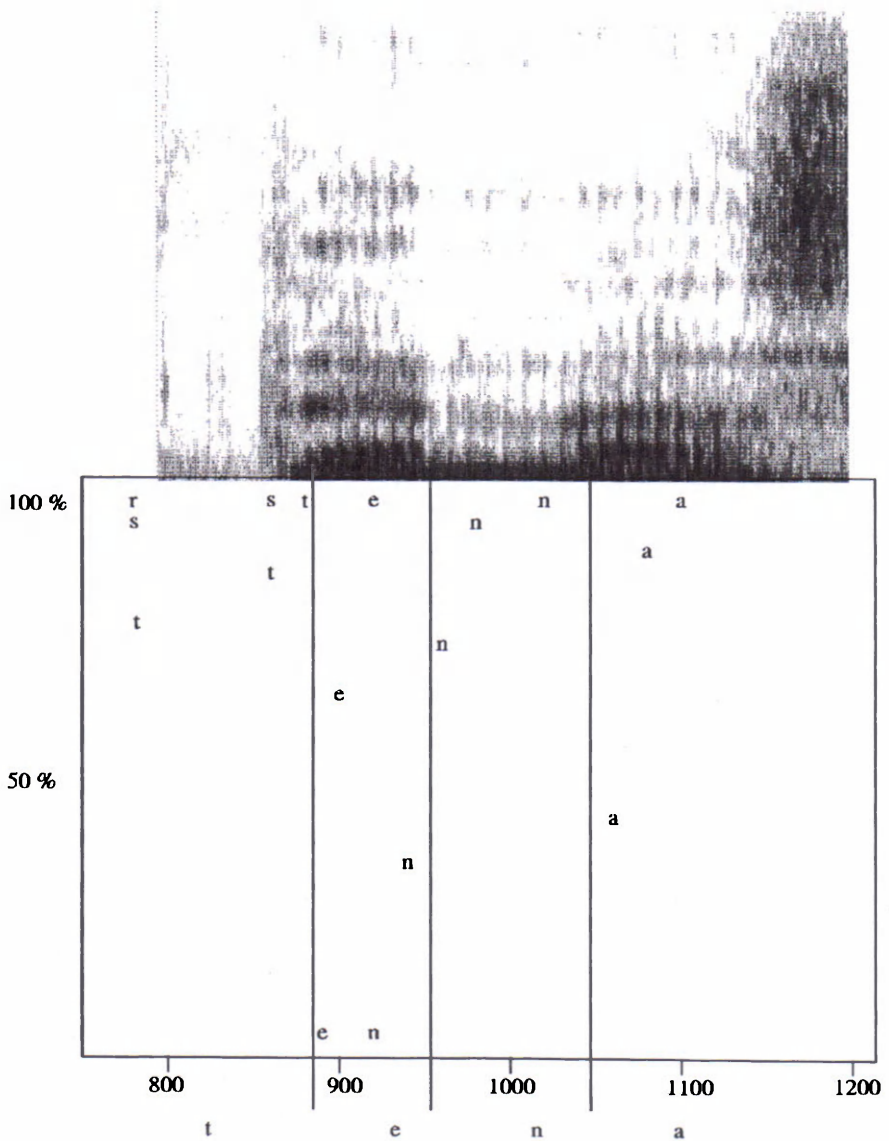
Suprotno od anticipacijske identifikacije *m* (na sl. 9), na sl. 10. može se uočiti da se *p* počinje prepoznavati tek nakon eksplozije, u vrijeme trajanja vokalskog dijela glasovnog segmenta *r*.

U odgovorima na stimulus koji je trajao 740 milisekundi *r* se identificira sa 95.6%, *s* ima postotak identifikacije 34.7%, a pojavljuje se čak i *t* sa 4.3%. Tako rano pojavljivanje *t* u odgovorima ispitanika može se protumačiti na dva načina. S jedne strane moguće je da je ono rezultat predviđanja cijele riječi na temelju proteklog niza fonema *prs...*, gdje *t* predstavlja vrlo vjerojatan nastavak riječi (iako bi, što se tiče fonotaktičkih mogućnosti moglo na istom

mjestu stajati i *-a, -e, -i, -k, -l, -n*, izgleda da najjače ograničenje u identifikaciji *t* predstavlja cijeli prethodni rečenični kontekst). S druge strane treba napomenuti da *gating*-posupak ima jedan inherentan nedostatak koji se sastoji u tome da se nagli prekid trajanja signala, njegov završetak, može se percipirati kao akustički znak za okluziju pa bi se u zaključivanju o okluzivima moralo voditi računa i o toj činjenici.

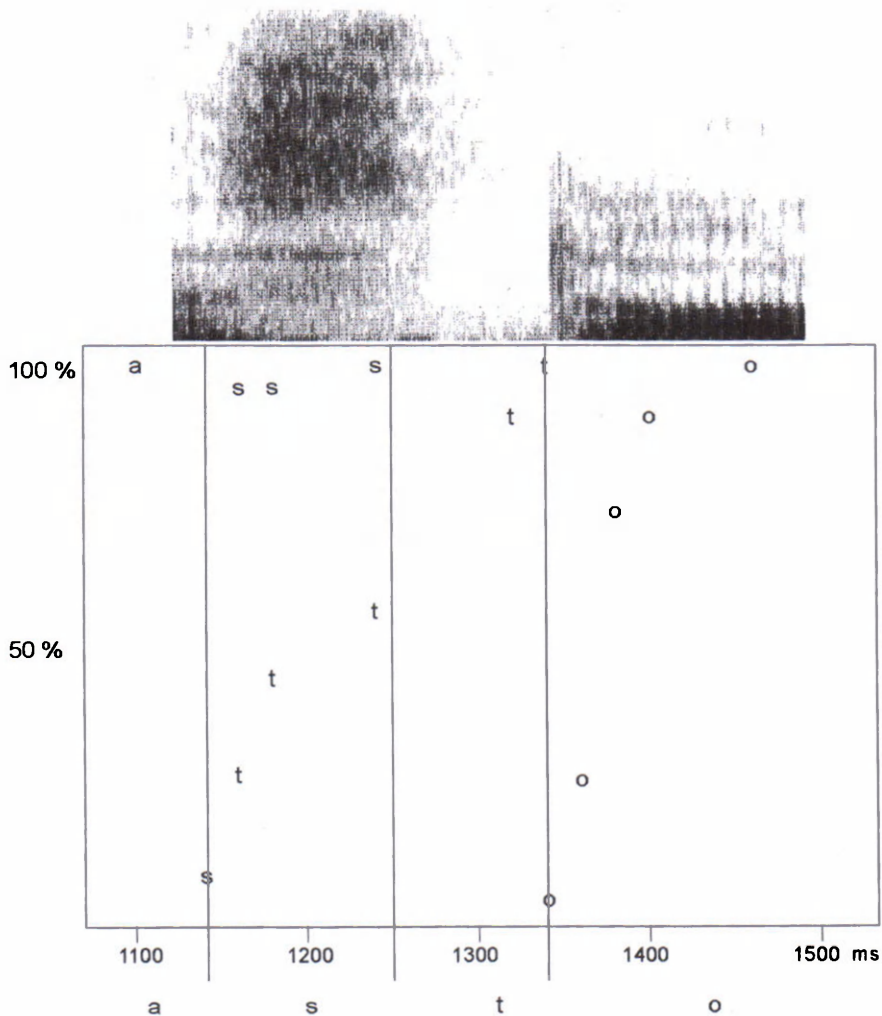


**Slika 10.** Postotak prepoznavanja pojedinih fonema povezan s vremenskim tijekom signala: segment /mprst/



**Slika 11.** Postotak prepoznavanja pojedinih fonema povezan s vremenskim tijekom signala: segment /tena/

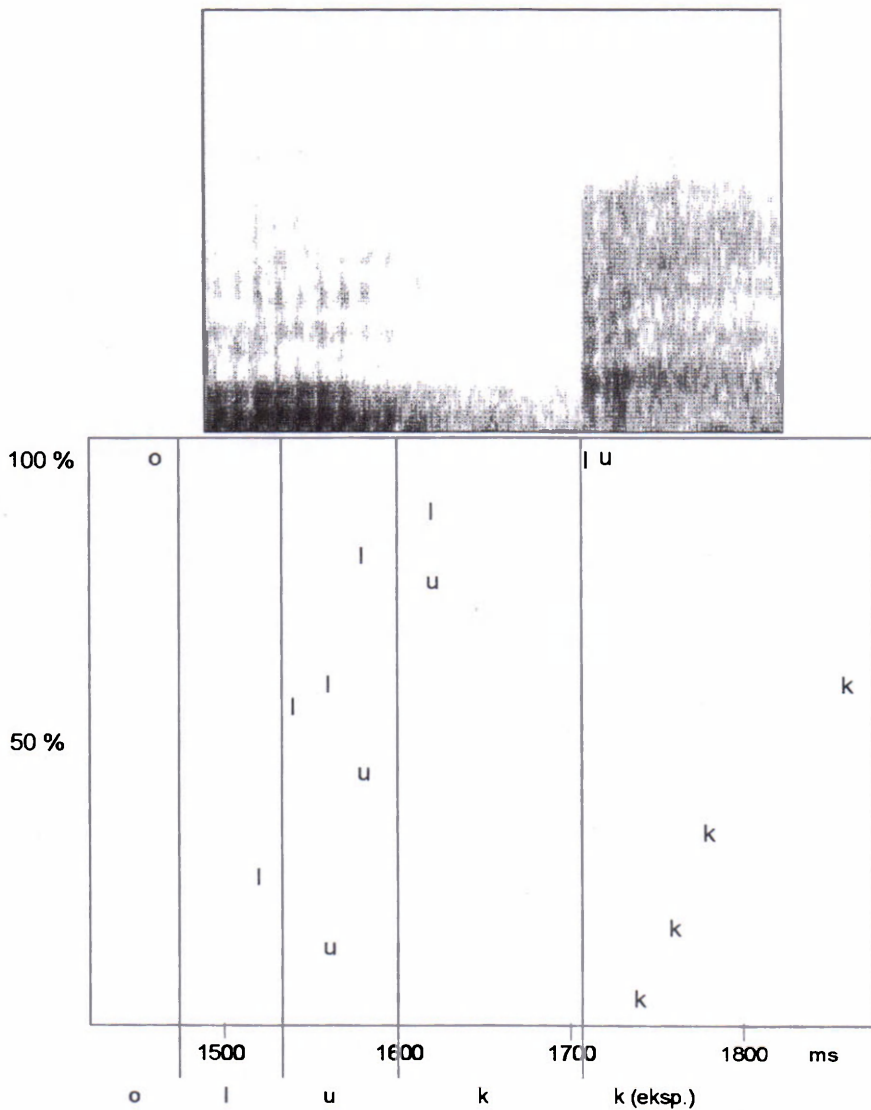
Identifikacija *n* s malim se postotkom (4.3%) pojavljuje kad je signal trajao 920 milisekundi, za vrijeme trajanja *e* tako da se opet može uočiti simultanost akustičkih informacija o vokalu koji traje i o nazalu koji slijedi. Maksimalna uspješnost identifikacije fonema *n* dosegnuta je tek pri kraju njegova fizičkog/nominalnog trajanja.



**Slika 12.** Postotak prepoznavanja pojedinih fonema povezan s vremenskim tijekom signala: segment /asto/



Na ovoj se slici može uočiti da se u ranoj fazi trajanja glasovnog ostvarenja fonema *s* pojavljuju i odgovori u kojima se identificira *t* koji vremenski slijedi neposredno iza *s*.



**Slika 13.** Postotak prepoznavanja pojedinih fonema povezan s vremenskim tijekom signala, segment /luk/

Ipak, *t* doseže maksimalnu identifikaciju tek u trenutku kad je u signalu jasno vidljiva (prisutna) eksplozija. A za trajanja eksplozije glasa *t* pojavljuje se u odgovorima potvrda za identifikaciju fonema *o* i to u karakterističnom prepoznatljivom "S" obliku koji zajedno s prikazom tijeka akustičkog signala pokazuje da je *o* dosegnuo maksimalan stupanj identificiranosti pri kraju trajanja vokalskog zvuka.

S malim pomakom na vremenskoj osi (i s malom razlikom u stupnju identifikacije) u odgovorima se pojavljuju *l* i *u*. Njihova identifikacija doseže maksimum tek u trenutku kad je prošlo čitavo trajanje okluzije, a u signalu je bila prisutna eksplozija *k*. Preuranjeno prepoznavanje *u* moglo bi biti da je posljedica jakog sintaktičko-morfološkog ograničenja zato što "vidim prste na ..." zahtijeva lokativ koji je u ovom slučaju obilježen upravo nastavkom *u*. Naime, u trenutku kad se definitivno prepozna fonem *l*, postaje jasno da u takvom obliku sintagma ne može biti završena, pa je najvjerojatnija pretpostavka da slijedi *u*.

Utjecaj informacija s viših jezičnih razina, tzv. *top-down* efekt, odnosno njegov izostanak, može se vidjeti po tome što je fonem *k*, unatoč prisutnosti cijelog akustičkog signala, dosegnuo samo 60 postotnu identifikaciju. S obzirom na to da ispitanici nisu znali da se radi o kraju izričaja, nije bilo nužno da *k* pridruže prethodnom dijelu signala s pretpostavkom da se radi o riječi *oluk* (koja je, osim toga, vrlo niskog frekvencijskog ranga), nego je pojava fonema *k* mogla značiti novu informaciju za neku drugu vrstu segmentacije.

U odgovorima na čitav govorni signal pojavili su se uz *k* još i fonemi *u* i *g*, a manji je dio ispitanika zamijetio samo šum eksplozije finalnog *k*, no taj zvuk nije interpretiran fonološki nego kao izvangovorni zvuk.

Na ovom primjeru pokazana je postupnost prepoznavanja fonema u ovisnosti o produljavanju akustičkog signala. Načelo prepoznavanja "slijeva nadesno" (koje je u slučaju prepoznavanja govorenih izvedbi, kao što je već rečeno, primjerenije nazvati načelom sukcesivnosti ili slijeda) može se reći da je vidljivo u dobivenim rezultatima, ali se isto tako jasno pokazalo da pojedini dijelovi akustičkog signala ne pripadaju samo jednom fonemu. Dakle, isti odsječci govornog kontinuuma ne predstavljaju nositelje informacija o samo jednom fonemu, nego se uočava simultana prisutnost informacija o dva ili čak tri sukcesivna fonema. Uz to, unatoč izričitom zahtjevu ispitanicima da odgovaraju samo fonemima koje čuju, a ne riječima, nesporno je da se neki fonemi identificiraju tek tijekom trajanja sljedećeg.

Na temelju podataka dobivenih u pokusu s prepoznavanjem čitave riječi u kontinuiranom govornom izričaju i na temelju podataka u pokusu s identifikiranjem pojedinačnih fonema u istim stimulusima moguće je usporediti dobivene rezultate. U obzir je uzeto objektivno trajanje pojedinoga govornog segmenta koji se smatra glasovnim ostvarenjem odgovarajućeg fonema određeno prema postojećim fonetsko-akustičkim kriterijima (Lehiste, 1970). Ti su podaci eksplicitno navedeni u stupcu B u tablici 10. Paralelno se promatralo u kojem je trenutku u trajanju glasovnog signala identificiran pojedini

fonem ako je zahtjev bio da se identificiraju samo fonemi (stupac D, tab. 10.), odnosno kako su identificirani pojedini fonemi unutar riječi koje je trebalo prepoznati (stupac E, u tab. 10.).

A	B	C	D	E
*		0		
v	155	155	220**	220
i	98	253	220**	220
d	52	305	340	320
i	98	403	340	340
m	111	514	380	380
p	98	612	640	640
r	116	728	740	720
s	87	815	760	740
t	70	885	860	740
e	69	954	920	900
n	87	1041	980	940
a	99	1140	1080	1060
s	110	1250	1160	1160
t	93	1343	1320	1180
o	122	1465	1400	1380
l	69	1534	1560	1500
u	66	1600	1700	1520
k	111	1711	***	***
****		1840		

**Tablica 10.** Usporedni prikaz trajanja glasovnih ostvarenja pojedinih fonema, duljine govornog stimulusa i vremena identifikacije fonema s obzirom na riječ i bez obzira na moguću riječ (ili u leksičkom i nelexičkom prepoznavanju), sve izraženo u milisekundama

A - fonem

B - trajanje glasovne realizacije pojedinog fonema u milisekundama

C - trajanje signala od početka do kraja glasovne realizacije pojedinog fonema

D - trenutak u trajanju stimulusa u kojem prepoznavanje prelazi 80% (za skupinu ispitanika koji su identificirali pojedine foneme)

E - trenutak u trajanju stimulusa u kojem prepoznavanje prelazi 80% (za skupinu ispitanika koji su prepoznavali cijele riječi)

\* - početak signala

\*\* - najkraći stimulus

\*\*\* - nikad nije identificiran preko 80% (samo 60,1%, odnosno 58,2%)

\*\*\*\* - trajanje najdužeg stimulusa (čitav izričaj)

Jednostavna usporedba dobivenih vremena identifikacije fonema pokazuje da se fonemi znatno brže identificiraju ako se smatraju sastavnim dijelovima riječi - od ukupno 18 fonema u samo 6 slučajeva vremena identifikacije bila su jednaka, dok je u svim ostalima iskazana prednost fonema smještenih u riječi ( $\chi^2$ -test pokazao je statistički značajnu razliku, na razini nesigurnosti manjoj od 0.01). Dakle, brže prepoznavanje fonema unutar riječi upućuje na zaključak da se radi o olakšavajućem leksičkom učinku na njihovo prepoznavanje, tj. da se leksičko i fonemsko procesiranje međusobno razlikuju. Ovakvi se podaci mogu interpretirati i u skladu s predviđanjima interaktivnoga modela prepoznavanja riječi koji uključuje povratnu vezu između riječi i fonema pretpostavljajući kraće vrijeme detekcije u slučaju da je fonem uklopljen u riječ nego onda kad se fonem nalazi unutar ne-riječi. Dakle, na temelju rezultata dobivenih u ovom pokusu potvrđeno je i na hrvatskom jezičnom materijalu postojanje povratne veze između leksičkog i fonemskog procesiranja.

## ZAKLJUČAK

Proučavanje fonetsko-fonološke komponente u procesu prepoznavanja riječi svojevrsna je osnovica za opisivanje sustava prepoznavanja i razumijevanja jezika, ali se pokazuje nužnim i uključivanje informacija s viših jezičnih razina (*top-down*), budući da, kako je pokazao drugi pokus, sama akustička informacija, pa i kad je potpuna, nije uvijek dovoljna za prepoznavanje riječi.

Rezultati provedenih pokusa protumačeni su u skladu s "kohortnom" teorijom prepoznavanja riječi W. Marslen-Wilsona, jer je potvrđeno da se s porastom količine akustičkog signala smanjuje broj predloženih riječi-kandidata za prepoznavanje, dok istodobno raste ukupan stupanj prepoznatosti za jednu riječ u "kohorti".

Dokazano je da problemi s delimitacijom zvučnog signala (s određivanjem granica među riječima) utječu na prepoznavanje riječi u povezanom govoru - prepoznavanje je znatno usporeno ili je potpuno onemogućeno.

U trećem je pokusu potvrđena hipoteza o uzajamnoj upravo proporcionalnoj povezanosti duljine akustičkog signala i uspješnosti prepoznavanja na razini fonema praćenjem ovisnosti identifikacije fonema o vremenskoj strukturi govornog signala. Pokazalo se da i u identificiranju fonema postoje naznake o sukcesivnosti, ali da on na ovoj razini nema takvu važnost kao na razini riječi (zato što je jasno pokazano da u pojedinim dijelovima zvučnoga signala istodobno postoje informacije o nekoliko fonema).

Usporedbom rezultata 2. i 3. pokusa potvrđen je leksički efekt u identifikaciji fonema tj. postojanje utjecaja riječi kao cjeline na otkrivanje pojedinih fonema u govornom iskazu.

Taj podatak potvrđuje tvrdnju da postoji razlika u leksičkom i fonemskom procesiranju, a kako je leksički utjecaj na prepoznavanje fonema olakšavajući, dobiveni rezultati prilog su postavkama o interaktivnom odnosu u procesiranju podataka na fonološko-fonetskoj i leksičkoj razini.



## LITERATURA

- Bakran, J.** (1996). *Zvučna slika hrvatskoga govora*, Zagreb, Ibis.
- Becker, C.A., Killion T.H.** (1977). Interaction of visual and cognitive effects in word recognition, *Journal of Experimental Psychology: Human Perception and Performance*, 3, 389-401.
- Bond, Z.S., Garnes S.** (1980). Misperception of fluent speech, u: R.A. Cole (ed) *Perception and production of fluent speech*, Hillsdale, N.J. Erlbaum.
- Cole, R.A., Jakimik J.** (1979). A Model of Speech Perception, u: R. Cole (ed.), *Perception and production of fluent speech*, Hillsdale, N.J. Erlbaum, 133-163.
- Elman, J.L., McClelland J.L.** (1984). Speech as a cognitive process: The interactive activation model, u: N. Lass (ed) *Speech and Language*, Vol. 10, N.Y. Academic Press.
- Forster, K.I.** (1976). Accessing the mental lexicon, u: Wales, R.J., E. Walker (eds), *New approaches to language mechanisms*, Amsterdam, North-Holland.
- Forster, K. I.** (1979). Levels of processing and the structure of the language processor, u: W.E. Cooper i E.C.T. Walker (eds) *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, N.J.: Erlbaum
- Garnham, A.** (1985). *Psycholinguistics: Central topics*, Methuen, London, N.Y.
- Grosjean, F.** (1980). Spoken word recognition processes and the gating paradigm, *Perception & Psychophysics*, 28, 4, 267-283.
- Grosjean, F.** (1985). The recognition of words after their acoustic offset: evidence and implications, *Perception & Psychophysics*, 38, 4, 299-301.
- Marslen-Wilson, W.D., Welsh, A.** (1978). Processing Interactions and Lexical Access during Word Recognition in Continuous Speech, *Cognitive Psychology*, 10, 29-63.
- Marslen-Wilson, W.D., Tyler, L.K.** (1980). The temporal structure of spoken language understanding, *Cognition*, 8, 1-71.
- Marslen-Wilson, W.D.** (1987). Functional parallelism in spoken word-recognition, u: Frauenfelder & Tyler (eds) *Spoken Word Recognition*, Elsevier Science Publishers B.V., Amsterdam, 71-102.
- McClelland, J.L., Rumelhart D.E.** (1981). An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings. *Psychological Review*, Vol.88, No. 5, 375-407.
- Miller, G.A., Bruner J., Postman L.** (1954). Familiarity of letter sequences and tachistoscopic identification, *Journal of Genetic Psychology*, 50, 129-139.
- Morton, J.** (1969). Interaction of information in word recognition, *Psychological Review*, 76, 165-178.
- Morton, J.** (1970). Word recognition, u: J. Morton i J.D. Marshall (eds), *Psycholinguistics 2: Structure u processes*. Cambridge, MA: MIT Press

- 
- Pisoni, D.B., Luce P.A.** (1987). Acoustic-phonetic representation in word recognition, u: Frauenfelder, U.H., L.K. Tyler (ed), *Spoken Word Recognition*, A Bradford Book, The MIT Press, Cambridge, Mass. London, England 21-52.
- Taft, M.** (1991). *Reading and the mental lexicon*, Lawrence Erlbaum Associates Ltd. Hove, London, Hillsdale.
- Vellutino, F. R.** (1982). Theoretical Issues in the Study of Word Recognition: The Unit of Perception Controversy Reexamined, u: Rosenberg, S. (ed), *Handbook of Applied Psycholinguistics. Major Thrusts of Research and Theory*, LEA, Hillsdale, N.J.
-

**Vlasta Erdeljac**  
Faculty of Philosophy, Zagreb

**WORD RECOGNITION IN CONTINUOUS SPEECH  
(ON EXAMPLE OF CROATIAN LANGUAGE)**

*SUMMARY*

*Experiments with Croatian language show that the phonetic-phonologic component represents the basis for the entire system of language understanding, but that this component operates together with higher level information.*

*Key words: word recognition, language understanding, Croatian language*

---