

TIHANA MARTINJAK

ELENMARI PLETIKOS OLOF

## MASKIRANJE OKLUZIVA U HRVATSKOME JEZIKU

PRETHODNO PRIOPĆENJE

<https://doi.org/10.17234/9789533791821.19>

---

Ovaj se rad bavi percepcijom okluziva u neoptimalnim uvjetima. Kao teorijski okvir daje se pregled psihoakustičkih teorija koje mogu objasniti procese pri slušanju maskiranoga govora. U radu se istražuje rezistentnost okluziva hrvatskoga jezika na maskiranje šumom. Signali su bili bezznačenjske dvosložne riječi koje su se sastojale od bezvučnih i zvučnih okluziva u okruženju vokala /a/. Bezznačenjske riječi snimljene su u izgovoru muškoga i ženskoga govornika odrasle dobi. Maska je bila širokopojasni šum raspona do 16 kHz. Svaki signal maskiran je šumom intenziteta 62, 65, 68, 71, 74 i 77 dB. Rezultati su pokazali da su bilabijalni okluzivi najrezistentniji na maskiranje šumom, što se vjerojatno može pripisati njihovu kodiranju u zakrivljenosti tranzijenata. Velarni su okluzivi slabije rezistentni na maskiranje od bilabijalnih, a najmanje su rezistentni alveolarno-dentalni. Po kriteriju zvučnosti rezistentniji su na maskiranje bezvučni okluzivi u odnosu na zvučne.

**Ključne riječi:** hrvatski jezik, okluzivi, maskiranje, maskiranje šumom, širokopojasni šum

---

### UVOD

#### Percepcija zvuka u uvjetima buke i maskiranje

Kada se nađemo u bučnom okruženju percepcija govora otežana je jer su informacije koje želimo čuti ometene raznim zvukovima koji otežavaju ili

onemogućuju primanje govorne poruke. Percepcija govora, psihoakustika, audiologija, neurofonetika i druge znanstvene discipline istražuju kako ljudi percipiraju i interpretiraju auditivne informacije, uključujući načine na koje ljudski mozak obrađuje različite akustičke signale. Saznanja dobivena istraživanjima pomažu pri oblikovanju audiosustava, dizajnu zvuka, tonskom inženjeringu, dizajniranju slušnih pomagala te u drugim područjima govorne i glazbene produkcije.

Tema je ovog rada maskiranje okluziva šumom. *Maskiranje* se definira kao psihoakustički fenomen u situacijama u kojima „čujan zvuk prestaje biti čujan zbog prisutnosti drugog zvuka” (Stevens i Davis, 1960: 208). U hrvatskom jeziku naziv koji koristimo za primarni, odnosno maskirani zvuk jest *signal*, dok zvuk koji onemogućuje njegovu percepciju nazivamo *maskom* (Bakran, 1999). Veličina efekta maskiranja izražava se u decibelima, a ta nam mjera govori za koliko se podigao prag sluha (engl. *threshold value*) signalnog zvuka. Veličina efekta maskiranja računa se kao razlika između praga sluha za signal bez maske i povišenog praga sluha kada je signal maskiran (Stevens i Davis, 1960: 208). U kontekstu ovoga rada bitno je istaknuti kako se okluzivi smatraju glasovima koje je općenito teško maskirati zbog njihove nagle promjene energije, odnosno eksplozije, na granicama s ostalim glasovima (Bakran i sur., 2000: 43).

S obzirom na to da vrijeme uključivanja maske nije uvijek isto, razlikujemo tri vrste maskiranja: simultano maskiranje (engl. *simultaneous masking*), maskiranje unaprijed (engl. *forward masking*, *postmasking*, *post-stimulus masking*) i maskiranje unazad (engl. *backward masking*, *premasking*, *pre-stimulus masking*), (Zwicker i Fastl, 1999: 61–62). Bakran i sur. (2000: 43) ističu da je u kontekstu identifikacije zvučnih tokova bitan element asinkronija različitih vrsta zvukova koji upravo u prirodnim komunikacijskim situacijama vrlo rijetko počinju i završavaju istovremeno.

Kako bi se protumačili mehanizmi percepcije zvuka u otežanim uvjetima, važno je spomenuti nekoliko psihoakustičkih teorija i postavki. Jedna je od najpoznatijih teorija percepcije u otežanim uvjetima teorija ostataka (engl. *residue theory*) koja objašnjava kako se tzv. virtualna i rezidualna tonska visina rekonstruira na temelju „ostataka” u visokim frekvencijama iako je  $f_0$  uklonjen iz zvuka (de Cheveigné, 2005: 191). U kontekstu maskiranja važno je spomenuti i teoriju kritičnih pojaseva koju je postavio Fletcher tvrdeći da ljudski sluh obrađuje zvukove unutar određenih frekvencijskih pojaseva tzv. kritičnih pojaseva (engl. *critical band* ili *critical bandwidth*). Zvukovi unutar istog kritičnog pojasa vjerojatnije će se međusobno maskirati, dok zvukovi u različitim pojasevima imaju manji učinak maskiranja. Fletcher je, u eksperimentima gdje je signal bio čisti ton, a maska šum, pretpostavio da dio spektra

šuma koji uspješno maskira signal jest upravo onaj dio koji je frekvencijski blizak tonu koji se pokušava maskirati (Zwicker i Fastl, 1999: 149). Maska će stoga postići najučinkovitiji efekt onda kada njezin zvuk aktivira iste dijelove bazilarne membrane kao i zvuk signala (Truax, 1999). Važno je spomenuti i teoriju spektralne i temporalne rezolucije koja istražuje kako ljudski sluh razlikuje zvukove na temelju njihovih spektralnih (frekvencijskih) i temporalnih (vremenskih) obilježja, što je ključno za razumijevanje govora u bučnim uvjetima. Rezultati se mogu protumačiti u okviru teorije zvučnog okruženja (engl. *auditory scene analysis*) koju je razvio Bregman 1994., a u hrvatskome su je kontekstu objasnili Bakran i sur. (2000). Ova teorija objašnjava kako ljudski sluh razdvaja i organizira zvukove u „scenama” ili „zvučnim tokovima”, što omogućuje identificiranje i fokusiranje na određene zvukove u bučnom okruženju, a najpoznatiji efekt slušanja u bučnom okruženju naziva se *cocktail party* efekt.

U izvanlaboratorijskim uvjetima za slušanje iznimno su važni položaj izvora zvuka u prostoru u odnosu na recipijenta i mehanizmi lokalizacije zvuka kao što su efekt akustične sjene (engl. *acoustic shadow*), faktori interauralnog slušanja kao što su razlika u fazi i razlika u vremenu u kojem signal dolazi do jednog i do drugog uha (engl. *interaural phase/time difference*, ITD) te razlika u intenzitetu (engl. *interaural level difference*, ILD) koje uho percipira s obzirom na blizinu izvora. Budući da se u ovome radu eksperiment provodi s izvorom zvuka u slušalicama obostrano mehanizmi lokalizacije izvora zvuka nisu uzeti u obzir.

U povijesnom pregledu razvoja psihoakustike Yost (2015) iznosi rezultate od Hirsh (1948) i Licklidera (1948) koji su dokazali kako razlika između signalnog i maskirajućeg zvuka pogodno utječe na percepciju signalnog zvuka, odnosno da njegov prag čujnosti postaje niži. Ovaj pomak u razini praga čujnosti nazvali su diferencijalnim pragom maskiranja (engl. *masking level difference*), a njegovo je proučavanje i danas jedna od važnih tema istraživanja. Teorija detekcije signala (TDS), (Green i Swets, 1966, prema Tomić, 2019) pretpostavlja kako su najslabiji podražaji (slabo) kodirani, odnosno pretpostavlja gradacijski odnos snage podražaja i snage kojom je taj podražaj neuralno reprezentiran. Ova teorija objašnjava način na koji čovjek percipira signale te kako razlučuje bitne od nebitnih informacija. Teorija detekcije signala tumači i pragove osjeta te ljudsku sposobnost razlikovanja signala od buke u neidealnim uvjetima. Pretpostavlja se kako na percepciju zvuka ne utječe samo osjetljivost na podražaj, već i niz drugih faktora koji u ispitaniku uzrokuju neku vrstu pristranosti. U testiranju percepcije govornih signala važnu ulogu ima i multimodalna percepcija. Poznato je da na percepciju zvuka utječu i ostali modaliteti kojima slušač percipira podražaje sinkro-

no ili asinkrono sa zvučnim podražajem, a jedan od modaliteta koji izrazito utječe na percepciju zvuka je vid. Najpoznatija sprega vida i sluha opisana je 60-ih godina 20. stoljeća i prema autoru nazvana McGurkov efekt. Ovaj efekt dokazan je i na govornicima hrvatskoga jezika (Mildner i Dobrić, 2015).

Za uspješnu govornu komunikaciju neophodna je i razabirljivost, kojom se naziva dekodiranje glasnika. Na nju utječu vizualna povratna sprega i semantika riječi koje izgovaramo. U bučnom okruženju u komunikaciji uživo očitavanje mjesta artikulacije sa sugovornikovih usana može nam pomoći pri razumijevanju iskaza. Slabo razabirljiv govor lakše je razumjeti ako smo upoznati s kontekstom unutar kojeg se koriste određene riječi, odnosno ako poznajemo temu o kojoj se govori (Pierce i David, 1958: 195–197).

Čovjek se u svom okruženju ne susreće samo s govornim zvukom. Sve zvukove koje možemo percipirati neprestano izdajamo u akustičke elemente koje potom povezujemo u cjeline koje pripadaju istom izvoru. Čovječja se percepcija oslanja na dva međusobno komplementarna mehanizma, a to su razdvajanje i spajanje (Bakran i sur., 2000: 31). Primjerice, simultane komponente negovornog zvuka odvajamo na temelju faktora poput tonske visine, prostornog smještaja, spektralne regije i neovisnosti promjene. Kod određivanja kategorije glasnika naš auditivni mehanizam objedinjuje sve spektralne komponente koje pripadaju pojedinoj skupini. Spomenute vrste mehanizama možemo nazvati primitivnim mehanizmima procesiranja i mehanizmima zasnovanima na obrascima. Oba sudjeluju u analizi zvučnog okruženja. Prvi oblik temelji se na prepoznavanju svojstava zvuka kao podražaja, dok drugi ovisi o iskustvu slušanja i poznavanju zvukova u jezičnom kontekstu, npr. ako čujemo riječ *whiskey*, naši obrasci mogu pripremiti našu percepciju da bude spremna čuti npr. s ledom. Na temelju istih tih obrazaca moguća je i restauracija fonema u prisustvu šuma.

Budući da se u radu istražuje percepcija okluziva pri maskiranju šumom, navest ćemo neke osnovne značajke šuma i njegovu primjenu. Šum je zvuk aperiodičnog titranja u kojem nema osnovnog tona ni harmonika, a spektar je kontinuiran. Šumove razlikujemo po raspršenju njihove zvučne energije unutar spektra te ih dijelimo na širokopojasne (engl. *broad band noise*, *wide band noise*) i uskopojasne (engl. *narrow band noise*) (Truax, 1999). Najpoznatiji je bijeli šum, zvuk koji najčešće asocijiramo uz radijske ili televizijske smetnje u signalu. Bijeli šum vrsta je širokopojasnog šuma i obuhvaća široki raspon frekvencija, a njegov intenzitet jednak je na svim razinama od 20 Hz i 20 kHz (Zwicker i Fastl, 1999: 62). Filtriranjem bijeloga šuma na različitim frekvencijskim razinama nastaju druge vrste šumova, npr. ružičasti šum koji, u usporedbi s bijelim, ima utišane visoke frekvencije, a njegova spektralna gustoća

smanjuje se za 3 dB po oktavi (Connaghan, 2021). Upravo se ova dva šuma u istraživanjima najčešće koriste kao maska za buku (Ghasemi i sur., 2022). Šum se koristi i pri obradi zvuka u procesu koji se zove *audio dithering*, a označava namjerno dodavanje niske razine šuma u audiodatoteku kako bi se uklonile kvantizacijske distorzije (iZotope, 2023).

## Artikulacijska, akustička i perceptivna obilježja okluziva

Dosadašnja istraživanja okluziva daju uvid u različite aspekte njihove percepcije te njihovih akustičkih i artikulacijskih obilježja. Okluzivi su jedini konsonanti koji se pojavljuju u svim poznatim jezicima svijeta, a čak 98 % jezika ima konsonante /p/, /t/ i /k/ (Ladefoged i Maddieson, 1996, prema Horga i Liker, 2016: 252). Okluzivi u hrvatskom jeziku (nazivaju se i zatvornici, pregradnici, eksplozivi, engl. *plosives*, *stops*) javljaju se kao tipične realizacije fonema /p/, /t/, /k/, /b/, /d/, /g/, a grkljanski okluziv [ʔ] javlja se kao nefonemski glasnik između stanke i samoglasnika (Škarić, 1991: 354). Okluzive po načinu artikulacije od drugih glasnika razlikuje stupanj suženja govornog prolaza: njihova faza držanja sastoji se od potpune pregrade koja sprečava prolaz zračnoj struji kroz usta, a istovremeno mekim se nepcem ne dozvoljava prolaz zračnoj struji kroz nos (Horga i Liker, 2016: 252). Mjesto artikulacije kod okluziva određuje se upravo prema mjestu gdje dolazi do zatvaranja fonacijskog toka: /p/ i /b/ su dvousnjeni ili bilabijali, /t/ i /d/ su zubno-nadzubni ili dentalno-alveolarni glasnici, a /k/ i /g/ mekonepčani ili velarni (Bakran, 1996: 58). Okluzive također možemo razlikovati po zvučnosti koja se ostvaruje u prvom dijelu njihove realizacije, odnosno u okluziji: bezvučni hrvatski okluzivi su /p/, /t/ i /k/, dok su zvučni /b/, /d/, /g/. Kod zvučnih okluziva glasnice titraju i proizvode prigušen zvuk na najnižim frekvencijama (Bakran, 1996: 59, 70, 71). Važno obilježje okluziva je eksplozija do koje dolazi nakon prestanka okluzije. Eksplozija okluziva po svom je spektralnom sastavu šum, a spektar šuma i trajanje šuma ovise o artikulatorima uključeni-ma u stvaranje pregrade kod okluzije. Kod tromijih artikulatora šum eksplozije traje duže pa najkraću eksploziju stoga imaju bilabijalni, dok najdulju eksploziju imaju velarni okluzivi (Bakran, 1996: 59–60). Za vrijeme eksplozije traje uključivanje izgovornog pokreta za idući glasnik. Ovu značajku nazivamo vrijeme uključivanja glasa (VUG, engl. *voice onset time*, *VOT*) – to je vrijeme između početka otvaranja govornog prolaza i početka uključivanja napetosti grkljana (Škarić, 1991: 143). VUG kod hrvatskih zvučnih okluziva traje kraće (prosječno +12 ms) nego kod bezvučnih (prosječno +18 ms), a s obzirom na izgovorno mjesto, najkraći je kod bilabijala /p/ (prosječno oko 15 ms), duži kod alveolara /t/ (oko 19 ms), a najduži kod velarnog /k/ (oko 25 ms), (Škarić, 1991: 211). Frekvencije šuma eksplozije ovise o veličini slobodnog prostora ispred izgovornog mjesta: u bilabijala šum je pojačan na niskim frekvencija-

ma, u /t/ i /d/ šupljina je ispred pregrade malena te su zbog toga u njihovoj eksploziji pojačane visoke frekvencije između 3500 i 5500 Hz, a u velara je zbog većeg slobodnog prostora ispred mjesta artikulacije šum eksplozije pojačan na niskim frekvencijama (Škarić, 1991: 199).

Kod okluziva u intervokalskom okruženju, koje ćemo koristiti u ovom istraživanju, važnu akustičku informaciju nose tranzijenti na prijelazu glasova. Njihov nagib, tj. mjesto u tišini prema kojem zakreće drugi formant vokala nazivamo *lokusom*. Lokus okolnih vokala u bilabijala je zakrenut prema dolje, što upućuje na „vrlo niske virtualne rezonantne frekvencije” (Škarić, 1991: 199) koje su rezultat velike šupljine iza mjesta pregrade. Kod /t/ i /d/ lokus je usmjeren prema sredini spektra jer je slobodni prostor iza izgovornog mjesta manji, a kod velara /k/ i /g/ lokus kreće prema visokim frekvencijama, zbog najmanje šupljine iza pregrade. Pokazalo se kako tranzijent F2 nosi puno više informacija od tranzijenta F1. Varijabilni lokus u /k/, kao i raspršenje njegove zvučne energije u eksploziji, posljedica su najnestabilnijeg mjesta artikulacije (Bakran, 1996: 69, 70; Škarić, 1991: 199).

Različita istraživanja ističu važnost tranzijenata, frekvencija šuma eksplozije, zvučnost i mjesto artikulacije u percepciji okluziva. Miller i Nicely (1955) na temelju istraživanja percepcije engleskih okluziva zaključili su da širokopojasni šumovi bolje maskiraju više frekvencije glasova. Njihovi rezultati pokazuju kako je obilježje zvučnosti mnogo važnije za razabirljivost od mjesta artikulacije, a pokazuje i kako se /p/ češće zamjenjuje s /k/ nego s /t/ i obratno te da se zvučni /b/ rijetko zamjenjuje s /d/ ili /g/. Stevens i Blumstein (1978) istraživali su globalne akustičke parametre prema kojima identificiramo mjesto artikulacije kod zvučnih okluziva te pokazali kako akustička obilježja poput spektra eksplozije i lokusa u CV slogovima treba gledati kao jedno jedinstveno akustičko obilježje prema kojem identificiramo okluzive. Stalna obilježja koja su se mogla povezati s identifikacijom zvučnih okluziva odnosila su se na raspršenja spektra: difuzni padajući ili ravni spektar kod /b/, difuzni rastući spektar kod /d/ te kompaktno raspršenje kod /g/ koje se grupira oko srednjih frekvencija. U istraživanju neaspiriranih okluziva telugu jezika Datta i sur. (1980) s pomoću sustava za automatsko prepoznavanje govora otkrili su da se mjesto artikulacije prepoznaje na temelju prva dva formanta i potvrdili kako je vrijeme trajanja tranzicija između okluziva i vokala ključno za raspoznavanje mjesta artikulacije. Najbolje su prepoznati zvučni velarni i bilabijalni okluzivi, u prosjeku 90 % primjera. Stabilnost formanata navodi se kao razlog boljeg prepoznavanja bilabijala. Alveolarni i dentalni okluzivi bili su slabije prepoznati. Osim u slučaju mekonepčanih okluziva, sustav za automatsko prepoznavanje govora bolje je detektirao zvučne od bezzvučnih glasova.

Percepciju bezvučnih intervokalskih okluziva hrvatskoga jezika istražili su Bakran i sur. (1991/1992) te zaključili da se na temelju prvog vokala najbolje prepoznaje okluziv /p/, što su protumačili kao posljedicu zakrivljenosti tranzijenta. Iako je ustanovljeno kako početni vokal nije dovoljan za uspješnu identifikaciju okluziva, glas /p/ bio je prepoznat u čak 92,4 % slučajeva. Također je ustanovljeno kako uspješnost prepoznavanja glasa /k/ ovisi o njegovu vokalskom kontekstu, a najbolje je bio prepoznat u okruženju vokala /a/. Šum eksplozije kao zaseban akustički signal pokazao se kao dostatan za ispravnu identifikaciju okluziva u čak u 95 % slučajeva. Pritom se najbolje identificirao glas /k/, a najgore glas /t/. Bezvučni okluzivi najslabije su prepoznati u signalima koji su sadržavali samo drugi vokal, a i u ovom je slučaju glas /p/ prepoznat najbolje, u 80,4 % slučajeva. Ustanovljeno je kako vokalski kontekst najmanje djeluje na identifikaciju /p/, kako se /t/ loše identificira u kontekstu prednjih vokala te kako se /k/ najuspješnije prepoznaje prema tranzijentu /a/, a zatim /e/ i /i/.

Hant i sur. (1996) proveli su istraživanje u kojem su okluzive maskirali bijelim šumom s ciljem da pronađu model za predviđanje praga čujnosti za maskirane bezvučne okluzive. Parametri koje su uzeli u obzir bili su trajanje eksplozije, širina pojasa i centralna frekvencija. Analiza govornog materijala podudarala se s parametrima Blumsteina i Stevensa (1978) za zvučne okluzive, a to su: kompaktnost spektra kod velarnih, difuzno rastući spektar kod alveolarnih te difuzno padajući spektar kod bilabijalnih okluziva. Rezultati su pokazali kako je kratkotrajna eksplozija od svega 10 ms dovoljna za identifikaciju mjesta artikulacije bezvučnih okluziva, odnosno da ljudskom uhu nisu potrebne detaljne informacije o spektralnom obliku eksplozija okluziva. Hant i Alwan (2000) istraživali su percepcijsku konfuziju sintetičkih afrikata u buci i otkrili da postoji pad od 5 do 10 dB u pragovima SNR-a (engl. *signal-to-noise ratio*) između nemaskiranoga govora i govora u buci. Ovo sugerira da izvorni govornici engleskog jezika koriste visokofrekventne migove za razlikovanje okluziva u buci.

U tumačenju percepcije okluziva nužno je analizirati ne samo perceptivne i akustičke osobine, već i artikulaciju glasnika elektropalatografskim metoda-ma (Lee i sur., 2023). Zasad su nam poznati podatci o stabilnosti okluziva /t/ u izgovoru engleskih govornika kod kojih okluzija traje prosječno oko 70 ms (Liker i sur., 2007). Rezultati istraživanja hrvatskih dentoalveolara (Liker i Gibbon, 2015) pokazali su da /t/ i /d/ općenito dijele mjesto artikulacije u dentoalveolarnoj regiji, ali i da povezivanje kvantitativnih fizioloških podataka s određenim mjestima artikulacije treba raditi oprezno, uzimajući u obzir varijabilnost u individualnim produkcijama. Istraživanje okluzije kod okluziva i afrikata u engleskom, kantonskom i talijanskom (Liker, 2010) pokazuje da

se razlikuju u svim istraživanim parametrima u sva tri jezika, što upućuje na to da se o univerzalnih obilježjima okluziva treba zaključivati oprezno jer su mnoga njihova obilježja jezično specifična. Istraživanje bezvučnog okluziva /k/ (Liker i Gibbon, 2007) pokazuje da je značajna varijabilnost među govornicima u svim mjerama, posebno u preciznoj lokaciji mjesta na mekom nepcu i artikulacijskoj udaljenosti između alveolarnog i mekonepčanog dodira.

Istraživanja različitih jezika naglašavaju složenost i raznolikost akustičkih i artikulacijskih svojstava okluziva važnih za razumijevanje percepcije govora. Istraživanja različitih jezika pokazuju neka univerzalna akustička obilježja okluziva, ali i specifičnosti kao što su razlike u VUG-u, varijabilnosti ili postojanosti mjesta okluzije, trajanju okluzije i dr.

Ovaj će rad dati uvid u percepciju hrvatskih okluziva u uvjetima gdje se njihova percepcija maskira širokopojasnim šumom. U radu su se postavile sljedeće hipoteze:

Bilabijalni okluzivi bit će rezistentniji na maskiranje od alveolarnih i velarnih okluziva koji imaju nestabilnije mjesto artikulacije.

Zvučni okluzivi bit će rezistentniji na maskiranje od bezvučnih okluziva.

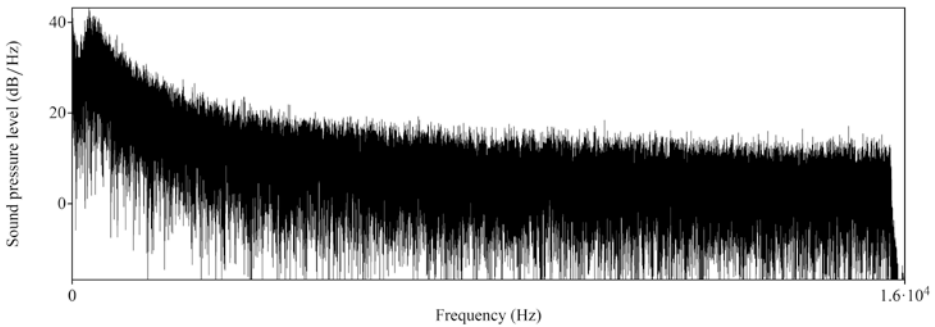
## METODOLOGIJA ISTRAŽIVANJA

### Govornici i stimulusi

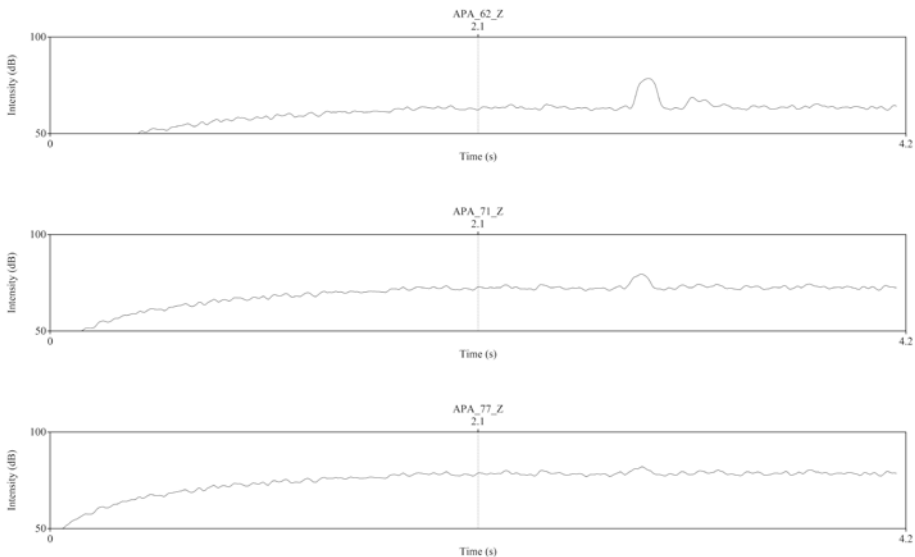
Mjerni instrument sastojao se od kreiranih zvučnih stimulusa te upitnika koji su ispitanici ispunjavali pri slušanju metodom papir/olovka. Zvučni stimulusi sastojali su se od signala i maske. Signali su bili bezznačenjske dvo-složne riječi (logatomi) koje su se sastojale od bezvučnih i zvučnih okluziva u okruženju vokala /a/, tj. riječi /apa/, /aba/, /ata/, /ada/, /aka/, /aga/. Logatomi su snimljeni u Studiju za akustička snimanja Odsjeka za fonetiku u svibnju 2023. godine, a izgovarali su ih jedan muški i jedan ženski govornik odrasle dobi, ugodnoga glasa i s dobrom dikcijom. Signal je potom obrađen u programu Praat (Boersma i Weenink, 2023). 12 signala (6 po govorniku) svedeno je na jednaki intenzitet od 74 dB. Maska je bila širokopojasni šum raspona do 16 kHz, koji je odabran zbog svoje spektralne gustoće, tj. jednakog intenziteta na svim svojim frekvencijama, što je uvjetovalo da svi dijelovi govornog zvuka budu jednako dobro maskirani. Svaki signal maskiran je šumom intenziteta 62, 65, 68, 71, 74 i 77 dB. Ove razine intenziteta šuma predstavljaju tipične razine buke u rasponu od blage do jače buke koja se pojavljuje u zatvorenim i otvorenim prostorima te utječe na percepciju govora. Razmak od 3 dB između intenziteta pojedinih maski odabran je jer prelazi minimalni

prag razlike percepcije intenziteta. Signali i maske spojeni su u jedinstvene stimulse, ukupno 72 stimulusa. Maska, čije je trajanje prije početka signala iznosilo 2,75 sekundi od početka trajanja postepeno se pojačavala, dok je u preostalom vremenu nakon završetka signala ostala jednakog intenziteta. Na slici 1 prikazana je spektralna ovojnica maske od 71 dB, a na slici 2 prikazane su krivulje intenziteta za stimulse ženskoga glasa s tri različita intenziteta maske.

Slika 1. Spektralna ovojnica maske intenziteta 71 dB



Slika 2. Krivulje intenziteta za /apa/ ženskoga glasa koje su maskirane šumom



od 62, 71 i 77 dB

## Ispitanici

U perceptivnom eksperimentu sudjelovalo je ukupno 24 ispitanika (19 žena i 5 muškaraca), prosječne dobi od 25 godina, u rasponu od 19 do 28 godina. Sudionici su bili studenti Sveučilišta u Zagrebu, većinom studenti Filozofskog fakulteta. Ispitanici su netom prije provođenja eksperimenta pisanim i usmenim putem upoznati s istraživanjem, a zatim su potpisali suglasnost za sudjelovanje. Svi ispitanici bili su osobe urednog sluha.

## Procedura

Eksperiment je proveden individualno u prostorima Odsjeka za fonetiku, a istraživanje s pojedinim ispitanikom trajalo je oko 20 minuta. Zvučni stimulusi reproducirani su s pomoću računala Lenovo Ideapad Gaming i slušalica Grado Labs SR60. Postavke na računalu preko kojeg su se prezentirali slušni stimulusi putem slušalica bile su istovjetne za sve sudionike eksperimenta te nikome nisu izazivale bol ili neugodu. Nacrt istraživanja odobrilo je 2023. Etičko povjerenstvo Odsjeka za fonetiku za provedbu istraživanja s ljudima.

## REZULTATI I RASPRAVA

U tablici 1 prikazani su objedinjeni rezultati za sve stimulse, 288 odgovora po okluzivu, neovisno o spolu govornika i o intenzitetu maske. Iz tablice je vidljivo da je bezvučni bilabijalni /p/ u prisustvu maske bio najbolje identificiran, a visoke rezultate postigao je i bezvučni velarni /k/. Najlošije je bio prepoznat zvučni dentalno-alveolarni /d/ s najnižih 160 točnih odgovora, a drugi nakon njega bio je zvučni velarni /g/ sa 187 točnih odgovora. Razlika između broja točnih i netočnih odgovora za pojedine grupe stimulusa, koji su kao postotci prikazani u tablici 2, s obzirom na mjesto artikulacije testirana je hi-kvadrat testom za nezavisne varijable. Razlika između broja točnih odgovora bilabijalnih (/p/ i /b/) i velarnih (/k/ i /g/) okluziva je statistički značajna,  $\chi^2(1, N = 1152) = 13,63$ ,  $p < ,001$ . Razlika između točnih odgovora bilabijalnih (/p/ i /b/) i dentalno-alveolarnih (/t/ i /d/) okluziva je statistički značajna,  $\chi^2(1, N = 1152) = 74,74$ ,  $p < ,001$ . Razlika između točnih odgovora dentalno-alveolarnih (/t/ i /d/) i velarnih (/k/ i /g/) okluziva također je statistički značajna,  $\chi^2(1, N = 1152) = 25,22$ ,  $p < ,001$ . Možemo zaključiti kako je prepoznavanje okluziva različito s obzirom na mjesto artikulacije.

Tablica 1. Prikaz broja točnih odgovora za pojedine okluzive

Odgovori ispitanika (N = 24)							
	/p/	/t/	/k/	/b/	/d/	/g/	drugo
Stimulusi							
/p/	<b>254</b>	2	8	13	0	0	11
/t/	22	<b>196</b>	22	3	6	8	31
/k/	10	11	<b>249</b>	1	1	4	12
/b/	40	3	3	<b>233</b>	1	2	6
/d/	2	10	1	29	<b>160</b>	78	8
/g/	6	1	13	11	25	<b>187</b>	45

**Napomena:** Ukupno je 12 različitih stimulusa po glasniku: jedna snimka muškog i jedna snimka ženskog izgovora riječi strukture VCV sa šest različitih razine maske. Pojedini stimulus slušalo je ukupno 24 ispitanika, što dovodi do 288 odgovora po glasniku.

U tablici 2 rezultati eksperimenta prikazani su u postotcima i uspoređeni po kategorijama mjesta artikulacije i zvučnosti.

Tablica 2. Prikaz postotka točnih odgovora s obzirom na zvučnost i mjesto artikulacije

	Zvučnost		Mjesto artikulacije		
	bezzvučni	zvučni	bilabijalni	dentalno- -alveolarni	velarni
Točni odgovori	81 %	67 %	85 %	62 %	76 %

Rezultati prikazani u tablici 1 i tablici 2 potvrđuju prvu hipotezu ovoga rada, odnosno da su bilabijalni okluzivi rezistentniji na maskiranje jer su značajno točnije percipirani od ostalih okluziva (velarni su prepoznati s točnošću od 76 %, a dentalno-alveolarni od 63 %). Objašnjenje ovakvih rezultata za bilabijale može se tražiti u obilježju koje ističu Bakran i sur. (1991/1992), a to je zakrivljenost tranzijenata na prijelazu između vokala i ovih okluziva. Rezultati se također mogu pripisati stabilnosti njihovih formanata, a potvrđuju i rezultate dobivene za indijski jezik telugu (Datta i sur., 1980). Iako se bilabijalni u odnosu na velarne okluzive kao skupina statistički značajno točnije prepoznaju, kada promotrimo pojedinačne rezultate (v. tablica 1) vidimo da je ova razlika velika kod zvučnih, ali ne i kod bezzvčnih okluziva, jer da je /p/ prepoznat 254 puta, a /k/ 249 puta (od 288 odgovora), dobro prepoznavanje glasnika /k/ može se povezati s vokalskim okruženjem koje se u prethodnim istraživanjima pokazalo kao optimalno za identifikaciju ovog okluziva (Bakran i sur., 1991/1992).

Što se tiče glasnika /k/, iako je njegovo mjesto artikulacije nestabilno, a stupanj koartikulacije visok, u ovom istraživanju /k/ je prepoznat u 86 % slučajeva. Ovakvi rezultati ukazuju na to da raspršenje zvučne energije u spektru eksplozije kod velarnog bezvučnog /k/ nije obilježje koje utječe na njegovo prepoznavanje u šumnom okruženju.

Tablica 2. prikazuje da su bezvučni okluzivi prepoznati s točnošću od 81 %, a zvučni s točnošću od 67 %. Hi-kvadrat test za nezavisne varijable između količine točnih odgovora zvučnih i bezvučnih okluziva pokazuje da je razlika statistički značajna,  $\chi^2(1, N = 1728) = 41,90, p < ,001$ . Ovi rezultati dokazuju da druga hipoteza ovoga istraživanja nije potvrđena, odnosno zvučni okluzivi nisu se pokazali rezistentnijima na maskiranje, štoviše značajno su manje točno prepoznati od bezvučnih. S obzirom na to da je ova hipoteza postavljena na temelju rezultata dobivenih u istraživanjima koja nisu provedena na hrvatskom jeziku (Miller i Nicely, 1955; Datta i sur., 1980), jedna od mogućih pretpostavki jest da ovi rezultati odražavaju specifičnost hrvatskih okluziva. No, odgovor koji se doima uvjerljivijim nalazi se u specifičnoj razlici izgovora glasnika /d/ dvaju govornika snimljenih za ovo istraživanje. Naime, kada se uspoređi izgovor istih okluziva kod muškog i ženskog govornika vidljivo je da je jedino kod /ada/ ženski okluziv bio značajno manje rezistentan od muškog, a i značajno manje rezistentan u odnosu na sve druge okluzive ove govornice (tablica 3).

Tablica 3. Prikaz postotka točnih odgovora za stimuluse muškog i ženskog govornika

	/p/	/t/	/k/	/b/	/d/	/g/
Ž	93 %	58 %	92 %	85 %	28 %	68 %
M	83 %	78 %	86 %	77 %	87 %	62 %

Činjenica da je glasnik /d/ u izgovoru ženske govornice bio najmanje rezistentan okluziv od svih glasnika u ovom istraživanju navodi na razmišljanje o odabiru reprezentativnog zvučnog signala za glasnik /d/. Također, neočekivan je rezultat da je glasnik /d/ muškoga govornika u prepoznavanju točniji od glasnika /b/ i /g/. Okluziv /d/ je u primjerima ženskoga glasa najčešće bio zamijenjen glasnikom /g/ (v. tablica 1), dok je kod izgovora /g/ čest odgovor bio /adga/ koji je u tablici 1 uvršten pod drugo. Ovakvi rezultati sugeriraju kako je potrebno prikupiti više uzoraka i muškoga i ženskoga glasa na kojima bi se sigurnije ustanovilo je li /d/ slabo rezistentan na maskiranje šumom ili je izgovor govornika u ovom eksperimentu na neki način bio atipičan. Razlog različite percepcije glasnika /d/ može se tumačiti i rezultatima istraživanja prepoznavanja filtriranih glasnika (Desnica-Žerjavić, 1990: 166) koja poka-

zuju da je glasnik /d/ jedan od glasova koji posjeduje „strukturni element diskontinuirane optimale”, što mu osigurava redundanciju u slušanju.

U tablici 4 prikazan je broj točnih odgovora po okluzivu s obzirom na intenzitet maske. Od sveukupno 48 primjera po glasniku, 24 muška i 24 ženska, u tablici je naveden broj točnih odgovora za svaki okluziv. Iz ovih je podataka ponovo vidljivo kako je /p/ najuspješnije identificirani okluziv, čak i u prisustvu najglasnije maske od 77 dB. Štoviše, okluziv /p/ više je puta uspješno identificiran u prisustvu najglasnije maske nego okluziv /d/ u prisustvu najtiše maske. Osim /p/ neočekivano dobre rezultate u prisustvu najglasnije maske ostvario je i glasnik /k/ koji je bio prepoznat 28 puta. Poznato je da intenzitet maske utječe na prepoznavanje signala, odnosno da glasna ili vrlo glasna maska uspješno prekriva signal (Bakran, 1999), zbog čega ove rezultate možemo protumačiti kao očekivane.

Tablica 4. Prikaz broja točnih odgovora za pojedine okluzive s obzirom na intenzitet maske

Intenzitet maske	Intenzitet maske					
	62 dB	65 dB	68 dB	71 dB	74 dB	77 dB
Stimulusi						
/p/	48	46	48	44	30	38
/t/	46	39	37	35	23	16
/k/	48	46	47	43	37	28
/b/	48	48	47	45	34	11
/d/	33	32	31	29	21	14
/g/	45	40	33	31	25	13

**Napomena:** Stimuluse čini ukupno 2 različita stimulusa (jedan u muškom i jedan u ženskom izgovoru) po jednom uvjetu intenziteta maske. Pojedini stimulus slušalo je ukupno 24 ispitanika, što dovodi do 48 odgovora po glasniku za jedan intenzitet maske.

Rezultati istraživanja ukazuju na to da su prema zvučnosti bezvučni okluzivi rezistentniji na maskiranje šumom od zvučnih okluziva, a prema kategoriji mjesta artikulacije bilabijali i velari rezistentniji su od dentoalveolara. Razina prepoznavanja svih glasnika potvrđuje očekivani rezultat da će što je maska glasnija, teže biti prepoznati signal (bezznačenjsku riječ). Za sigurnije tumačenje maskiranja dentoalveolara potrebno je provesti daljnja istraživanja na većem broju muških i ženskih govornika kako bi se ustanovilo je li riječ o specifičnosti hrvatskih okluziva ili se ovdje radilo o individualnim razlikama izgovora hrvatskih govornika čiji je govor korišten u kreiranju stimulusa.

U ovom istraživanju okluzivi nisu bili segmentirani, već su se njihova artikulacijska obilježja gledala globalno. Maskirani su u cijelom svom trajanju i nije se dodatno isticala njihova eksplozija, VUG ili lokus. Ovakav bi pristup stoga bio u skladu s opažanjima Blumsteina i Stevensa (1979) koji na kratke i brze promjene koje se odvijaju u eksploziji okluziva gledaju kao na jedinstveno akustičko obilježje. Ono što također izostaje u ovom eksperimentu jest proučavanje okluziva u različitom vokalskom okruženju. Kao signali korištene su samo bezznačenjske riječi strukture VCV. Može se pretpostaviti da se u stvarnim suboptimalnim komunikacijskim uvjetima, u kojima se okluzivi pojavljuju kao dio jezičnoga konteksta, ovi glasnici identificiraju bolje nego kad su u bezznačenjskom kontekstu. Za daljnja istraživanja predlaže se ispitivanje rezistentnosti hrvatskih okluziva u drugačijem vokalskom okruženju te na većem broju primjera i muških i ženskih glasova.

## ZAKLJUČAK

Okluzivi su glasnici koje čovjek može dobro prepoznati čak i u uvjetima kada komunikaciju ometa buka nalik na širokopojasni šum. Možemo zaključiti da je ovo istraživanje dokazalo da se u hrvatskome jeziku po kriteriju mjesta artikulacije dentalno-alveolarni okluzivi jače maskiraju u odnosu na bilabijalne i velarne, a velarni jače u odnosu na bilabijalne. Po kriteriju zvučnosti jače se maskiraju zvučni okluzivi u odnosu na bezvučne. Okluzivi hrvatskog jezika u ovom su istraživanju bili uspješno percipirani u primjerima u kojima je šum maske bio slabijeg intenziteta (od 61 do 71 dB), a znatno teže u uvjetima šuma maske jačeg intenziteta (od 74 do 77 dB). Okluziv koji se pokazao kao najrezistentniji bio je bezvučni bilabijalni /p/ koji je prepoznat u 88 % slučajeva. Budući da je njegov zvučni parnjak /b/ također dobro prepoznavan, s točnošću od 80 %, potvrđena je hipoteza kako će bilabijali biti uspješno percipirani zbog zakrivljenosti tranzijenata i stabilnosti formanata. Uz bilabijale dobro je prepoznat i zvučni velarni /k/, što možemo pripisati vokalskom okruženju, odnosno vokalu /a/ koji pogoduje prepoznavanju ovog okluziva. Spoznaje dobivene ovim istraživanjem mogu se primijeniti u različitim područjima, primjerice u razvoju audio tehnologije, terapiji govora te u dizajniranju sustava za okruženja s puno buke ili za osobe s oštećenjem sluha.

## LITERATURA

- Bakran, J. (1996). *Zvučna slika hrvatskoga govora*. Zagreb: Ibis grafika.
- Bakran, J. (1999). *Psihoakustika 2009*. URL: <https://fonet.ffzg.unizg.hr/djelatnici/bakran/psi/skripta2009.htm> (pristupljeno 7. rujna 2023.).

- Bakran, J., Erdeljac, V. i Lazić, N. (2000). Analiza zvučnog okruženja. *Govor*, 17 (1), 29–48.
- Bakran, J., Horga, D. i Stamenković, M. (1991/1992). Percepcija mjesta artikulacije bezvučnih okluziva. *Govor*, 8 (1–2), 31–47.
- Blumstein, S. E. i Stevens, K. N. (1979). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *The Journal of the Acoustical Society of America*, 67 (2), 648–662.
- Boersma, P. i Weenink, D. (2023). *Praat: doing phonetics by computer* [Computer program]. Version 6.4. URL: <http://www.praat.org/> (pristupljeno 1. studenoga 2023.).
- Connaghan, T. (2021). *The Difference Between Types Of Noise*. URL: <https://emastered.com/blog/different-types-of-noise> (pristupljeno 3. rujna 2023.).
- Datta, A., Ganguli, N. i Ray, S. (1980). Recognition of unaspirated plosives – A statistical approach. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28 (1), 85–91.
- De Cheveigne, A. (2005). Pitch perception models. U C. J. Plack, A. J. Oxenham, R. R. Fay i A. N. Popper (ur.) *Pitch: Neural coding and perception*, 169–233. New York: Springer New York.
- Desnica-Žerjavić, N. (1990). Neke slušne osobine glasova. *Govor*, 7 (2), 157–177.
- Ghasemi, S., Fasih-Ramandi, F., Monazzam-Esmaeelpour, M. R. i Ardakani, S. K. (2022). Different Colors of Noise and Their Application in Psychoacoustics: A Review Study. *Journal of Health and Safety at Work*, 12 (3), 459–482.
- Hant, J. J. i Alwan, A. (2000). Predicting the perceptual confusion of synthetic stop consonants in noise. *ICSLP*, 3, 941–944.
- Hant, J. J., Strobe, B. P. i Alwan, A. (1996). A Psychoacoustic model for the noise masking of voiceless plosive bursts. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, 1, 570–573. DOI: 10.1109/ICSLP.1996
- Horga, D. i Liker, M. (2016). *Artikulacijska fonetika: anatomija i fiziologija izgovora*. Zagreb: Ibis grafika.
- iZotope (2023). *What Is Dithering in Audio?* URL: <https://www.izotope.com/en/learn/what-is-dithering-in-audio.html> (pristupljeno 9. svibnja 2023.).
- Lee, A., Liker, M., Fujiwara, Y., Yamamoto, I., Takei, Y. i Gibbon, F. (2023). EPG research and therapy: further developments. *Clinical Linguistics & Phonetics*, 37 (8), 701–721, DOI: 10.1080/02699206.2022.2080588.

- Liker, M. (2010). Elektropalatografska analiza faza okluzije u /t/ i /tʃ/: višejezično istraživanje. U V. Mildner i M. Liker (ur.), *Proizvodnja i percepcija govora: profesoru Damiru Horgi povodom njegovog sedamdesetog rođendana*, 106–117. Zagreb: Odsjek za fonetiku Filozofskog fakulteta Sveučilišta u Zagrebu – Odjel za fonetiku Hrvatskoga filološkog društva – FF Press.
- Liker, M. i Gibbon, F. E. (2008). Tongue palate contact patterns of velar stops in normal adult English speakers. *Clinical Linguistics & Phonetics*, 22 (2), 137–148.
- Liker, M. i Gibbon, F. E. (2015). Place of articulation of anterior nasal versus oral stops in Croatian. *Journal of the International Phonetic Association*, 45 (1), 35–54.
- Liker, M., Gibbon, F. E., Wrench, A. i Horga, D. (2007). Articulatory characteristics of the occlusion phase of /tʃ/ compared to/t/in adult speech. *Advances in Speech Language Pathology*, 9 (1), 101–108.
- Mildner, V. i Dobrić, A. (2015). Reconsidering the McGurk Effect. U *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow, The Scottish Consortium for ICPhS 2015 (ur.). London – Delhi: the International Phonetic Association. URL: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0662.pdf> (pristupljeno 9. svibnja 2023.).
- Miller, G. A. i Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27 (2), 338–352.
- Pierce, J. R. i David, E. E. (1958). *Man's world of sound*. New York, Garden City: Doubleday Company.
- Stevens, K. N. i Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64 (5), 1358–1368.
- Stevens, S. S. i Davis, H. (1960). *Hearing: Its Psychology and Physiology*. New York: John Wiley & Sons, Inc.
- Škarić, I. (1991). Fonetika hrvatskoga književnog jezika. U S. Babić i sur., *Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika*, 61–377. Zagreb: Hrvatska akademija znanosti i umjetnosti – Nakladni zavod Globus.
- Tomić, I. (2019). *Uloga snage distraktora u vidnom radnom pamćenju*. Doktorski rad. Filozofski fakultet Sveučilišta u Zagrebu. Zagreb. URL: [http://darhiv.ffzg.unizg.hr/id/eprint/11491/1/Tomic\\_Ivan.pdf](http://darhiv.ffzg.unizg.hr/id/eprint/11491/1/Tomic_Ivan.pdf) (pristupljeno 7. rujna 2023.).

- Truax, B. (ur.). (1999). *Handbook for acoustic ecology (Second Edition)*. Cambridge Street Publishing. URL: <https://www.sfu.ca/sonic-studio-webdav/handbook> (pristupljeno 9. svibnja 2023.).
- Yost, W. A. (2015). Psychoacoustics: A brief historical overview. *Acoustics Today*, 11 (3), 46–53.
- Zwicker, E. i Fastl, H. (1999). *Psychoacoustics: facts and models*. Berlin – New York: Springer.

## ACOUSTIC MASKING OF STOP SOUNDS

### Abstract

This paper deals with the perception of stop sounds under suboptimal conditions. As a theoretical framework, an overview of psychoacoustic theories that can explain the processes of listening to masked speech is provided. The paper investigates the resistance of stop sounds in the Croatian language to masking by noise. The signals were meaningless two-syllable words consisting of voiceless and voiced stops surrounding the vowel /a/. The meaningless words were produced by adult male and female speakers. The mask was broadband noise with a range of up to 16 kHz. Each signal was masked by noise at 62, 65, 68, 71, 74 and 77 dB. The results showed that bilabial stops are the most resistant to masking by noise, which can probably be attributed to their encoding in the curvature of transients. Velar stops are less resistant to masking than bilabial sounds, and alveolar-dental sounds are the least resistant. According to the criterion of voicing, voiceless stops are more resistant to masking than voiced ones.

**Keywords:** Croatian language, stops, masking, noise masking, broadband noise