Otpornost hrvatskih glasnika na maskiranje bukom

Porupski, Ivan

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:982984

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-03-24**

Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences

zir.nsk.hr

DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

Sveučilište u Zagrebu
Filozofski fakultet
Odsjek za fonetiku

Ivan Porupski

OTPORNOST HRVATSKIH GLASNIKA NA MASKIRANJE BUKOM

Diplomski rad

Zagreb, rujan 2024.

Sveučilište u Zagrebu
Filozofski fakultet
Odsjek za fonetiku

Ivan Porupski

OTPORNOST HRVATSKIH GLASNIKA NA MASKIRANJE BUKOM

Diplomski rad

Mentor: Dr. sc. Elenmari Pletikos Olof, izv. prof.

Zagreb, rujan 2024.

PODACI O AUTORU

Ime i prezime: Ivan Porupski

Naziv oba studija: Znanstveno usmjerenje fonetike, računalna lingvistika

PODACI O RADU

Naslov rada na hrvatskome jeziku:

Otpornost hrvatskih glasnika na maskiranje bukom

Naslov rada na engleskome jeziku:

Resistance of Croatian phonemes to noise masking

Datum predaje rada: 5. rujna 2024.

IZJAVA O AUTORSTVU DIPLOMSKOGA RADA

Ovim potvrđujem da sam osobno napisao diplomski rad pod naslovom

Otpornost hrvatskih glasnika na maskiranje bukom

i da sam njegov autor.

Svi dijelovi rada, podaci ili ideje koje su u radu citirane ili se temelje na drugim izvorima (mrežni izvori, udžbenici, knjige, znanstveni, stručni članci i sl.) u radu su jasno označeni kao takvi te su navedeni u popisu literature.

Ime i prezime studenta:
Ivan Porupski

Zagreb, rujan 2024.

Zahvala

Htio bih se zahvaliti svojoj obitelji, pogotovo majci i sestri, svim prijateljima te nastavnom i nenastavnom osoblju koji su mi pružali konstantnu podršku kroz moj studij.

Sadržaj

0. Uvod 3

1. Teorija 3

1.1. Percepcija govora u buci 3

1.2. Grafički prikaz zvuka 5

1.3. Auditivno maskiranje 8

1.4. „Glimpsing“ teorija 9

1.5. Hrvatski konsonanti 10

2. Metodologija 11

2.1. Dizajn govornog materijala 11

2.2. Generiranje rečenica 13

2.3. Snimanje govornika 14

2.4. Maska šuma oblika govora (SSN) 14

2.5. Izrada i provođenje upitnika 17

2.6. Statistička analiza rezultata upitnika 18

2.7. Akustička analiza govornog materijala 18

3. Rezultati i rasprava 19

3.1. Akustička analiza *glimpsova* u govornom materijalu 19

3.2. Matrice zabune 20

4. Zaključak 26

5. Literatura 27

Sažetak 30

Abstract 31

Prilog 32

Prilog A – Referentna čestota fonema 33

Prilog B – Govorni materijal 36

Prilog C – SoSci Upitnik 38

Prilog D – MFCC prikazi 41

/apa/, /aba/ 41

/ata/, /ada/ 42

/aka/, /aga/ 43

/asa/, /aša/ 44

/aza/, /aža/ 45

/afa/, /aha/ 46

/ača/, /ađa/ 47

/aca/, /aja/ 48

/ama/, /ana/ 49

/ara/, /ala/ 50

/ava/ 51

0. Uvod

Razumijevanje otpornosti glasnika na maskiranje bukom ključno je za napredak našeg znanja o percepciji govora. Maskiranje glasnika nastaje kada pozadinska buka ometa sposobnost razlikovanja pojedinih glasnika, što dovodi do značajnih izazova u komunikaciji. U suvremenim kontekstima, ljudi se često suočavaju s okolinama gdje vanjski zvukovi maskiraju ključne komponente govora, otežavajući verbalnu interakciju. Ovaj rad istražuje otpornost hrvatskih glasnika na energetsko maskiranje bukom, s ciljem otkrivanja obrazaca koji bi mogli informirati razvoj učinkovitijih komunikacijskih pomagala i tehnologija. Istraživanjem složenosti otpornosti glasnika nastojimo poboljšati razumijevanje auditivnog procesiranja i poboljšati razumljivost govora u bučnim okruženjima. Takvi su uvidi ključni za dizajniranje boljih slušnih pomagala, poboljšanje sustava prepoznavanja govora i stvaranje učinkovitijih komunikacijskih strategija u raznim okruženjima.

1. Teorija

Prema bibliometrijskoj analizi koju su proveli Chen i Chang (2022), percepcija govora (engl. *speech perception*) predstavlja jedno od ključnih područja istraživanja unutar govornih znanosti. Autori predviđaju daljnji razvoj u istraživanju percepcije govora s posebnim naglaskom na dva smjera: (1) istraživanje percepcije govora kroz cijeli životni vijek, uključujući dojenčad, starije populacije i kliničke skupine poput djece s gubitkom sluha ili razvojnom disleksijom, te (2) primjenu neurokognitivnih tehnika za razumijevanje regija mozga uključenih u percepciju govora i vremenski tijek moždanih aktivnosti tijekom obrade govora (engl. *speech processing*).

1.1. Percepcija govora u buci

Percepcija govora u buci složen je problem. Kako bismo mogli probati riješiti taj problem, potrebno ga je prvo rastaviti na njegove elemente. Percepcija govora podrazumijeva govornika koji govori i slušatelja koji sluša te su željni uspostavit komunikaciju. „Da bi se komunikacija uspostavila, moraju se usklađeno poredati dijelovi komunikacijskoga lanca“, nadodaje Škarić (2007: 17). Komunikacijski lanac teorijski je okvir prijenosa obavijesti, od govornika do slušatelja. Unutar lanca, dok je obavijest u komunikacijskom kanalu, moguć je

unos buke. „Entropičnost koju dobiva odredište (slušatelj), a koja ne potječe od izvora, zove se *buka* ili *šum*“ (Škarić, 2007: 19). Drugim riječima, buka je svaki zvuk koji slušatelj čuje, a govornik ne stvara. Valja naglasiti kako Škarić (2007: 33) zasebno definira *šum* kao doživljaj aperiodičnosti titraja. Jelaković (1978: 16) definira šum kao „nepravilno titranje u kojem nema ni stalnih frekvencija ni stalnih amplituda“. U kontekstu ovoga rada, šum je aperiodično titranje, dok je buka entropičnost koja ne potječe od govornika.

Uho je slušni čovjekov organ koji ne perceptira sve frekvencije podjednako. Krivulja koja opisuje taj odnos zove se *prag sluha* (Jelaković, 1978: 31-41). Zbog toga, svaka buka ne utječe na percepciju podjednako.

Prisutstvo buke u komunikacijskom kanalu česta je pojava. Među najpoznatijim primjerima je i takozvan „*cocktail party*“ problem, odnosno zadatak čuti zvuk od interesa, često govorni signal, u složenom slušnom okruženju. Škarić (2007: 21) definira govorni signal, odnosno govor kao „čovječja zvučna optimalna komunikacija oblikovana ritmotvornim jedinicama slogova, riječi i rečenica.“

Škarić (2007: 18) navodi kako se u komunikaciji informacijski elementi izvora moraju označiti, pa se tako fonemi označuju glasnicima ili slovima. *Fonem* je najmanji odsječak jezične poruke koji jednu riječ može razlikovati od druge riječi (ili ne riječi), dok u govoru foneme signaliziraju glasnici. *Glasnik* je „najkraći cjelovit hotimičan izgovorni pokret“. Raznolikost glasnika u govoru na bilo kojem jeziku ne iskorištava sve raspoložive osobine govora. U hrvatskom jeziku ima znatno veći broj različitih glasnika nego što ima različitih fonema (Škarić, 2007: 58-59). Zadnju definiciju koju još treba razlikovati jest fonemski glasnici. *Fonemski glasnici* su glasnici koji najpotpunije govorno ostvaraju fonemske crte. Fonemske crte su apstrakt kojima se opisuje predodžba fonema, a odnose se na ponovljive govorne osobine glasnika, u neutralnom izgovornom položaju (npr. zvučnost, frikcija na različitim frekvencijama, itd.) (Škarić, 2007: 73).

U kontekstu ovoga rada, spomenuti pojmovi koristit će se zamjenjivo. U najužem smislu, ovaj rad ispituje percepciju fonemskih glasnika u bučnom okruženju.

1.2. Grafički prikaz zvuka

Grafički prikaz zvuka igra ključnu ulogu u vizualizaciji i analizi akustičkih svojstava govornog signala. Spektrogram (engl. *spectrogram*), kao jedan alat za analizu, omogućava detaljan prikaz frekvencijskih komponenti zvuka kroz vrijeme. Korištenjem spektrograma možemo preciznije ustanoviti obrasce i otklone u govornom signalu. Svaki periodičan proces može se rastaviti na niz jednostavnih sinusoidnih titranja uz pomoć brze Fourierove transformacije (engl. *fast Fourier transform*, dalje u radu FFT) (Jelaković, 1978: 14). FFT je metoda digitalne obrade signala za izračunavanje Fourierove transformacije, odnosno spektar signala (Johnson, 2012: 60). Zvučni spektar prikazuje frekvencije i amplitude određenih komponenata zvuka. Prema tome, Jelaković (1978: 15-16) dijeli zvukove na tri vrste: *čisti ton* (jednostavno sinusoidno titranje), *mužički* ili *harmonički ton* (sastoje se od osnovnog tona i broja harmoničkih tonova, nadvalova), te *šum* (nepravilno titranje u kojem nema ni stalnih frekvencija ni stalnih amplituda).

Ljudski govor ima širok raspon mogućih zvučnih spektra jer se sastoji od različitih frekvencija koje se mijenjaju tijekom vremena (Bakran, 1996: 17). Primjenom FFT na kratke segmente (npr. 20 ms) govornog signala mogu se identificirati dominantne frekvencije i njihove amplitude (Johnson, 2012: 60). Drugim riječima, FFT pomaže u razumijevanju kako se zvučne energije raspoređuju po različitim frekvencijama unutar govora.

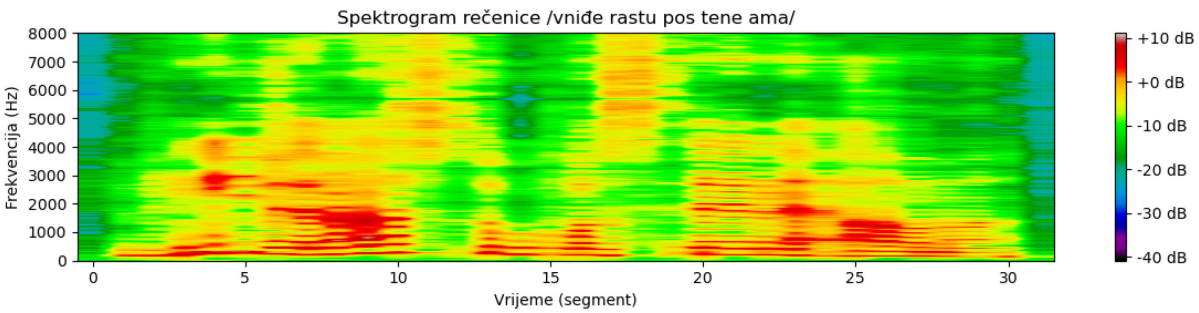
Spektrogram ili *spektrograf* je grafički prikaz frekvencijskog sadržaja govornog signala u vremenu (Slika 1.). Takav prikaz omogućava vizualno identificiranje različitih fonetskih obrazaca ili karakteristika kao što su formanti, harmonici, tranzijenti (Bakran, 1996). Spektrogram nam prikazuje linearnu ovisnost frekvencija, koja nije reprezentativna spomenutom pragu sluha.

Davis i Mermelstein (1980) u svojem radu predlažu kako se kompaktna reprezentacija zvuka prilagođena ljudskoj percepciji mogla postići uz pomoć m*el-frekvencijskih cepstralnih koeficijenata* (engl. *mel-frequency cepstral coefficients*, dalje u radu MFCC). Cepstrum koeficijenti su rezultat diskretne kosinusne transformacije (engl. *discrete cosine transformation*, DCT) realnog logaritma spektra kratkotrajne energije izraženog na mel skali (engl. *mel scale*) (Davis i Mermelstein, 1980; Zheng i sur., 2001; Jurafsky i Martin, 2009: 326).

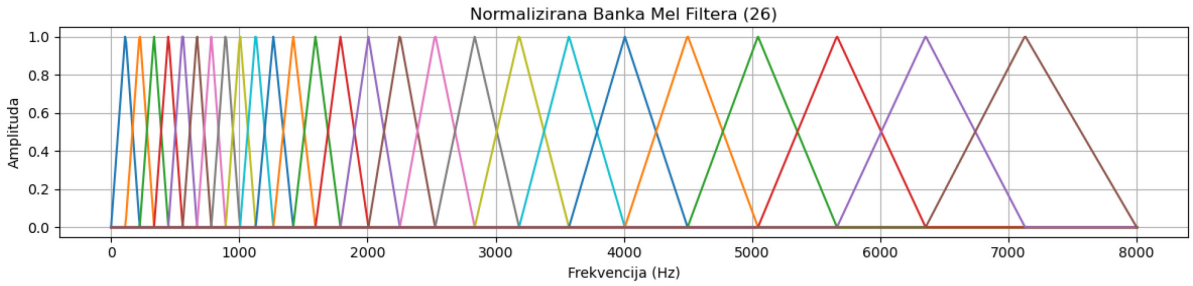
Drugim riječima, primjenom diskretnih filtera (Slika 2.) na zvuk te DCT transformacijom dobivamo MFCC prikaz[1] (Slika 3.), odnosno kompaktniji prikaz ljudske percepcije govora u odnosu na spektrogram (Slika 1.).

Davis i Mermelstein (1980) nadalje navode kako bi se mel skala trebala shvatiti kao skala koja ima linearni razmak između frekvencija ispod 1000 Hz, ali logaritamski razmak frekvencija iznad 1000 Hz. Primjenom mel skale smanjuje se potencijalna nesigurnost pri višim frekvencijama gdje se očekuje najviše varijacije u produkciji govora među govornicima, uz minimalni efekt na percepciju slušatelja (Johnson, 2012: 164).

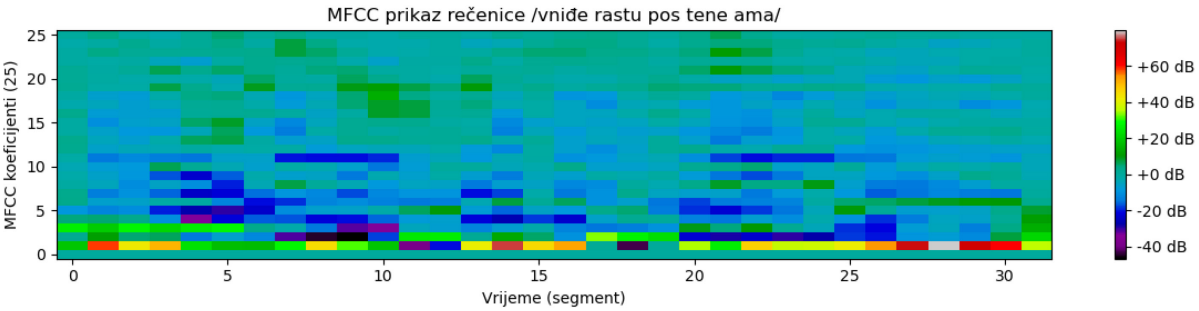
[1] Primjer vodiča za izradu MFCC: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/v (Pristupljeno 17. srpnja 2024.)



Slika 1. Grafički prikaz rečenice /vniđe rastu pos tene ama/ (Prilog B, 18.) spektrogramom.



Slika 2. Prikaz filtarske banke, odnosno 26 filtera koji demonstriraju linearne karakteristike ispod 1000 Hz i logaritamske iznad 1000 Hz.



Slika 3. MFCC prikaz rečenice /vniđe rastu pos tene ama/[2].

[2] Zheng i sur. (2001) savjetuju uvijek uključiti nulti koeficijent u izračun, no ovdje je izostavljen zbog boljeg konstrasta.

1.3. Auditivno maskiranje

Slušatelji obično pokušavaju pratiti govorni signal jedne osobe usred kakofonije drugih zvukova (npr. žamor, smijanje, udaranje čaša itd.). U takvim situacijama govorni signal od interesa često je blizu praga razumljivosti. Ako se intenzitet govornog signala imalo smanji relativno buci okoline, često postaje nerazabirljiv (McDermott, 2009).

Auditivno maskiranje odnosi se na pojavu u kojoj buka prekriva ili maskira zvuk govora (Jelaković, 1978: 61). U kontekstu percepcije govora u „*cocktail party*“ okruženju, može se klasificirati u nekoliko oblika: *energetsko maskiranje* i *modulacijsko maskiranje*. Sva maskiranja unose buku koja prekriva govorni signal u spektro-temporalnom smislu. *Informacijsko maskiranje* karakterizira unos zbrke izvan periferne razine (Wang & Xu, 2021). Drugim riječima, energetsko maskiranje se „nadglašava“ sa željenim signalom, dok informacijsko otežava „praćenje“ željenog signala na semantičkoj razini.

Postoje različite vrste maski koje mogu maskirati govorni signal. Maska može biti govor, šum (npr. bijeli ili ružičasti šum), čisti ton, harmoničan ton i sl. Iako su sve maske u nekom obliku energetske maske, glavna razlika između maske govora i drugih maski je u tome što maska govora može biti i informacijska maska. Generalno gledano, maske govora mogu se kategorizirati prema broju govornika. *Maska jednog govornika* (engl. *Single-Speaker Mask*) najjednostavniji je slučaj govornog maskiranje gdje se signal maskira bukom govora jednog govornika. Takva maska je informacijska maska, ukoliko se traži semantičko razumijevanje govornog signala. Ako se maskira bukom dva ili više sugovornika, takva maska naziva se „*bukom brbljanja*“ (engl. *babble-speech*).

Buka brbljanja, ovisno o broju sugovornika, može informacijski maskirati. Dva istovremena sugovornika najjača su informacijska maska buke brbljanja. Ako imamo „beskonačno“ sugovornika, buka brbljanja prelazi u takozvanu *buka oblika govora* (engl. *speech-shaped noise*, SSN[3]) (Cooke, 2006: 1565) ili *žamor*[4], koja može biti isključivo energetska maska.

[3] U nekim radovima je moguće spominjanje (engl.) *speech-induced emission noise*, SIEN.
[4] Definicija prema HJP: Žamor je „šum mnogih glasova, razgovor koji se slijeva u jedan šum različite jačine.“ (https://hjp.znanje.hr/index.php?show=search_by_id&id=f15iXhd0,) (Pristupljeno 17. kolovoza 2024).

Kada govorimo o maskiranju, odnosno spajanju dva ili više zvuka u jedan, možemo govoriti o omjeru signala i buke, odnosno *SNR* (engl. *signal-to-noise ratio*) izražen u dB (Johnson, 2012: 115), a decibel (dB) je logaritamski odnos dviju zvučnih snaga (Jelaković, 1978: 13), odnosno odnos dva zvuka.

U principu, prisutnost buke djeluje nepovoljno na precepciju govora. Unatoč tome, neka istraživanja pokazala su kako prisutnost optimalne buke malog intenziteta (engl. *low-level noise*) zapravo može povećati razabirljivost govora slušateljima s urednim sluhom uz pomoć fenomena stohastičke rezonancije (engl. *stochastic resonance*). Ako unutar nekog sustava postoji prag koji se mora nadmašiti prije nego što sustav okine (npr. slušni živci), unosom buke u govorni signal povećava se statistička vjerojatnost da će prag biti nadmašen te govorni signal uspješno percipiran (Kishon-Rabin i sur., 2008).

1.4. „Glimpsing“ teorija

No, što ako je maska buke većeg intenziteta? Cooke (2006) postavlja pitanje, koriste li slušači (engl.) „*glimpses*“[5] – spektro-temporalne dijelove zvuka u kojima je govorni signal pod najmanjim utjecajem maske – kako bi razumijeli govor u buci? Drugim riječima, *glimpsovi* su svi dijelovi govora koji energetski nadmašuju masku. Kako bi se razabirljivost govora mogla kvantificirati, Cooke (2006: 1562) jasno definira *glimpsove* kao spektro-temporalni dijelovi govornog signala koji su barem 3 dB intenzivniji od maske (SNR ≥ 3 dB).

Cookeov rad ispituje teoriju *glimpsova* koristeći vlastiti sustav za automatsko prepoznavanje govora (engl. *automatic speech recognition*, ASR), ali i ljudske slušatelje u procjeni razabirljivosti 16 engleskih konsonanata (/b, p, d, t, g, k, m, n, l, r, f, v, s, z, š, č/) iz korpusa Shannona i sur. (1999).

Lu i Cooke (2009) proširuju teorijski okvir, ali ovaj put ispitivajući percepciju govora preko Lombardijevog govora. Valja razlikovati „*cocktail party*“ efekt od Lombardijevog efekta, pri čemu je Lombardijev efekt promjena postave govornog aparata kao odgovor na bučnu okolinu i jedan od najpoznatijih utjecaja buke na proizvodnju govora, dok se „*cocktail*

[5] *Glimpses* nemaju izravni prijevod na hrvatski jezik. Potencijalni prijedlozi bi mogli biti *migovi* ili *blicovi*, pri čemu Škarić (2007: 27) već prevodi (engl.) *cues* kao migove. Međutim, dok *glimpses* tehnički spadaju pod *cues* iz razloga što jesu oblik (spektralne) redundantnosti, u ovom radu koristi će se kroatizirana inačica *glimpsovi*.

party“ efekt odnosi na mogućnost usmjeravanja pažnje na željenog govornika unutar bučnog okruženja (Li i Lutman, 2006, Korvel i sur., 2020). U bučnom okruženju, govorni aparat se prilagođava, povećava se intenzitet govora te se energija spektra pomiče prema višim frekvencijama (engl. *spectral tilt*). To podržava Cookeovu (2006) teoriju jer se buka, odnosno maska češće javlja s niskim frekvencijskim sastavnicama nego s višim, pa zbog toga dolazi do više izraženih *glimpsova*.

Li i Lutman (2006) dotiču se *glimpsova*, tvrdeći kako se princip temelji na oskudnosti (engl. *sparsness*) i zalihnosti (engl. *redundancy*) govora. Zalihnost je komunikacijski višak koji služi kao osiguranje da se neka obavijest ne izgubi, a u običnom razgovoru ima 70-80% zalihnosti (Škarić, 2007: 18). Jedna od pretpostavki je kako je govorni signal sve oskudniji informacijama, segmenti čistog govora su izraženiji, dozvoljavajući više *glimpsova*. Nadalje, postoje fiziološki dokazi koji sugeriraju da je oskudnost u zvuku važan princip za uspješno neuronsko kodiranje okoline. Jedan primjer smanjivanje zalihnosti (ili dekorelacije) je već spomenuti MFCC prikaz.

1.5. Hrvatski konsonanti

Budući da će se u radu istraživati hrvatski konsonanti, slijedi kratak pregled njihovih karakteristika i klasifikacije u skupine. Po uzoru na Cookea (2006), ovaj rad će procijeniti razabirljivost, odnosno otpornost 21 hrvatskog konsonanta (/p, t, k, b, d, g, s, š, z, ž, h, f, c, č, đ, m, n, j, r, l, v/) u tri bučna okruženja, koristeći energetsko maskiranje SNN-om.

Odabrana 21 hrvatska konsonanata mogu se ugrubo podijeliti u 4 skupine: okluzivi (/p, t, k, b, d, g/), frikativi (/s, š, z, ž, h, f/), afrikate (/c, č, đ/) i ostali (/m, n, j, r, l, v/).

Okluzive (ili zatvornike) karakterizira faza potpune pregrade i povećavanje intraoralnog tlaka (okluzije) te faze naglog otpuštanja zračne struje (eksplozije) (Horga i Liker, 2016: 252). *Frikativi* (ili tjesnačnici) su konsonanti koji se proizvode jakim suženjem govornog prolaza, pri čemu se stvara zvučna turbulencija, odnosno šum (frikcija) (Horga i Liker, 2016: 252). *Afrikate* (ili zatvoreni tjesnačnici) nastaju stvaranjem potpune pregrade (okluzije) popračeno otpuštanjem prednjeg dijela te stvaranjem pregrade u stražnjem dijelu (frikcija) (Horga i Liker, 2016: 253). Ove tri kategorije konsonanta nazivaju se i pravim konsonantima.

Osim njih, preostalu kategoriju dijelimo na: nazale (/m, n/), aproksimante (/j, v/), lateral (/l/) i vibrant (/r/). *Nazali* su konsonanti s pregradom zračnoj struji u ustima, ali sa slobodnim prolazom kroz nosnu šupljinu (Horga i Liker, 2016: 257). *Aproksimanti* (ili približnici) su konsonanti koji, nalik frikativima, rade pregradu zračnoj struji, ali nedovoljno uski prolaz da bi došlo do stvaranja šuma, pri čemu prevladava harmoničan ton (Horga i Liker, 2016: 253). *Laterali* (ili bočni) su konsonanti koji imaju kontakt između pasivnog artikulatora (nepce, zubi, itd.) i središnjeg dijela jezika, pri čemu zračna struja prolazi bočnim prolazom. *Vibrant* (ili treptajnik) /r/ je glasnik kod kojega aktivni artikulator (jezik) vibrira i udara u pasivni artikulator zbog aerodinamičkih uvjeta stvorenih zračnom strujom (Horga i Liker, 2016: 259).

Vizualni prikaz (spektrograf) svih hrvatskih konsonanata dostupan je u knjizi Zvučna slika hrvatskog govora (Bakran, 1996: 58-171). Otpornost pojedine skupine glasnika na buku ovisi o karakteristikama skupine, odnosno karakteristikama pojedinog glasnika ili riječi (Kalikow i sur., 1977: 1337).

2. Metodologija

2.1. Dizajn govornog materijala

Ovaj rad koristiti će metodologiju prema Kalikow i sur. (1977) uz nekoliko izmjena. Jasno, ispitivanje percepcije govora u buci (engl. *speech perception in noise*, SPIN) obradili su već mnogi autori na različite načine pri čemu svaki pristup ima svoje prednosti i mane.

Rad Kalikow i sur. (1977) naglašava sam proces dizajniranja testa koji ispituje percepciju govora u buci te diskutiraju o svakom pojedinom parametru koji utječu na percepciju govora u buci, kao što su fonetski i prozodijski faktori, utjecaj konteksta, poznavanje riječi, utjecaj buke i faktori vezani uz slušatelja. Ovaj rad će na dosljedan način proći sve faktore, izuzev utjecaja konteksta i poznavanje riječi.

A) Pod fonetske i prozodijske faktore autori navode fonetsku raspodjelu (balans) govornog materijala i prozodijska obilježja većih jedinica unutar rečenice. Drugim riječima, čestotna raspodjela fonema u govornom materijalu treba odražavati čestotnu raspodjelu fonema u jeziku. Referentna fonetska raspodjela (vidi Prilog A, Slika A1.) prikazuje čestotnu

raspodjelu glasnika u tekstnom korpusu ParlaSpeech-HR[6] v1.0[7] (Ljubešić i sur., 2022). Raspodjela je dobivena opojavničenjem korpusa na 30 hrvatskih glasnika, a opis raspodjele je rađen na uzor Tambovtseva i Martindalea (2007), iako nemaju krivulju za raspodjelu hrvatskih glasnika.

Usporedbe radi, priložena je Škarićeva (1991: 352) fonetska raspodjela (vidi Prilog A, Slika A2.). Prozodijska obilježja (kao što su ritam, naglasak, intonacija i dr.) važna su za razumijevanje govora nekog jezika (Škarić, 2007), pa je s toga važno održati prozodiju prema standardnom hrvatskom jeziku.

B), C) Utjecaj konteksta i poznavanje riječi su zanemareni u ovome radu iz razloga što se rad ne bavi informacijskim, već isključivo energetskim maskiranjem. Dapače, lingvistički prazan materijal svejedno je potreban kako bi se izbjegla bilo kakva mogućnost predviđanja zadnje riječi zbog samog lingvističkog konteksta materijala.

D) Utjecaj buke odnosi se na slušne situacije današnjice. Vrsta i način buke kako utječu na već složen akustički prostor oko nas. Mnogi autori (Kalikow i sur, 1977; Cooke, 2006; Kishon-Rabin i sur., 2008) odabrali su buku brbljanja, dok su drugi (Cooke, 2006; Lu i Cooke, 2009) odabrali SSN kao svoju masku. Lu i Cooke (2009: 1256) koristili su prosjek (LTAS) cijelog govornog korpusa (8 različitih govornika) za izradu SSN maske. U ovome radu odlučila se napraviti SSN maska za svakog govornika zasebno prema prosjeku (LTAS-u) ukupnog govornog materijala pojedinog govornika. Ovim pristupom postiže se kontrolirano maskiranje, relativno jednako za sve glasnike, odnosno govornike.

E) Faktori vezani uz slušatelja – osim akustičkih, lingvističkih i situacijskih utjecaja na percepciju govora, postoji niz faktora koje ovise o slušatelju, a mogu negativno utjecati na razabirljivost. Kao na primjer oštećenje sluha, utjecaj starosti na sluh i veličina vokabulara (Kalikow i sur, 1977: 1338). Vodit će se računa o oštećenju sluha i starosti, dok vokabular može biti zanemaren jer se ne ispituje informacijsko (lingvističko) maskiranje. Ukoliko se test

[6] https://www.clarin.si/repository/xmlui/handle/11356/1494.
[7] U radu se analiza v1.0 verzija tekstnog korpusa, unatoč tome što su Ljubešić i sur. (2024) objavili novi ParlaSpeech-HR v2.0 korpus, koji se razlikuje se po većem opsegu (ParlaMint 4.0 umjesto ParlaMint 2.1).

administrira *online*, valja uzeti u obzir i uređaj kojim će slušatelj reproducirati zvuk (preko mobitela, preko zvučnika, preko slušalica, itd.).

Na kraju, Kalikow i sur. (1977: 1338) spominju uočene probleme s postojećim testovima. Generalno, testovi percepcije govora ne opisuju adekvatno situacije i uvjete svakodnevne komunikacije. Ističu kako je pogotovo teško dizajnirati rečenice kao govorni materijal koji ima kontroliranu predvidljivost riječi, fonetski balans i sve druge atribute potrebne za jedno istraživanje. Poželjno je da se test može lako administirati, te da je pouzdan i valjan.

2.2. Generiranje rečenica

Sukladno prethodnom poglavlju, idealni govorni materijal je onaj koji ima prozodijska obilježja (standardnog) hrvatskog jezika, lingvistički (semantički) je prazan te fonetski balansiran. Lingvistički prazan govorni materijal sa željenim prozodijskim obilježima može se dobiti uz pomoć tekstnog dijela govornog korpusa ParlaSpeechHR-v1.0 (Ljubešić i sur., 2022), opsega 14,2 milijuna riječi, 31,6 milijuna slogova ili 70,8 milijuna fonema.

Opojavničenje (engl. *tokenization*) je zadatak razdvajanja niza znakova u prepoznatljive jezične jedinice koje čine dio jezičnih podataka (Bird i sur., 2009). Opojavničenje tekstnog korpusa izvršeno je vlastitim programom „SyllableTokenizer_HR“[8] u Jupyter Lab interaktivnoj okolini. Program je napisan u Pythonu i temelji se na NLTK[9] opojavničivaču (engl. *tokenizer*). Koristila se prilagođena hijerarhija sonornosti specifično za hrvatski jezik prema Škariću (2007: 88). Korpus se opojavničio prvo na 173 225 jedinstvenih riječi, a zatim te riječi na 9 207 jedinstvena sloga.

Temeljem statističke analize broja slogova po riječi, broja fonema po slogu i čestotne distribucije pojedinih jedinstvenih slogova (vidi Prilog A, Slike A3. i A4.), program „SyllableTokenizer_HR“ nasumičnim odabirom slaže rečenice zadanih parametara: broj rečenica, broj riječi po rečenici, maksimalan broj slogova po riječi; maksimalna duljina sloga (u slovima), odabir jedinstvenih slogova prema čestoti, broj iteracija (vidi Prilog B).

[8] https://github.com/porupski/diplomski (jupyter notebooks/NLTK_SyllableTokenizer_HR_custom_v6.0.ipynb)
[9] https://www.nltk.org/index.html (Pristupljeno 15. srpnja 2024.)

Novonastale riječi sastoje se od novih nizova slogova koji nalikuju hrvatskim riječima, ali pri čemu nemaju gotovo nikakvo semantičko značenje. Program „SyllableTokenizer_HR“ zatim analizira frekvencije slova, odnosno fonema, novog materijala te radi poredbu na prethodno utvrđenu referentnu distribuciju fonema (vidi Prilog A, Slika A1.). Program koristi Pearsonov koeficijent korelacije (R^2) kako bi utvrdio koliko svaka iteracija podudara željenoj distribuciji. Program zatim analizira odabran broj iteracija te odabire govorni materijal s najvišom korelacijom, odnosno R^2.

Odabrana su 21 hrvatska glasnika, isključujući hrvatske vokale i konsonante /ć/, /dž/, /lj/ i /nj/. Govorni materijal (vidi Prilog B, R^2 = 0,8522) sastoji se od četiri seta po 21 jedinstvenih rečenicu, pri čemu je na kraju svake traženi stimulus oblika VCV (engl. *vowel – consonant – vowel*). Svaki glasnik, odnosno konsonant nalazi se isključivo u okolini hrvatskog vokala /a/ (npr. /apa/, /aba/, itd.).

2.3. Snimanje govornika

Govorni materijal čitalo je ukupno 4 govornika, dva muška i dva ženska govornika. Svaki govornik čitao je po jedan set (21 rečenicu). Govorni materijal bio je snimljen u Studiju za akustična snimanja Odsjeka za fonetiku, u prostoriji sa zvučnom izolacijom namjenjenom specifično za snimanje govora. Snimio se mikrofonom AKG C414B-ULS i zvučnom karticom Fireface UFX te pohranjen u .wav formatu (48 kHz).

Dugotrajni prosječni spektar (engl. *Long-term average spectum*, dalje u radu LTAS) svakog pojedinog govornika prikazan je narančastom crtom u Slici 4. Svakoj rečenici normalizirala se glasnoća (Normalize Loudness…) u programu Audacity (3.3.2.) na -23 LUFS (engl. *Loudness units full-scale*) kako bi se uklonila bilo kakva moguća varijacija u glasnoći prilikom snimanja te spremljena je kao .wav audio datoteka.

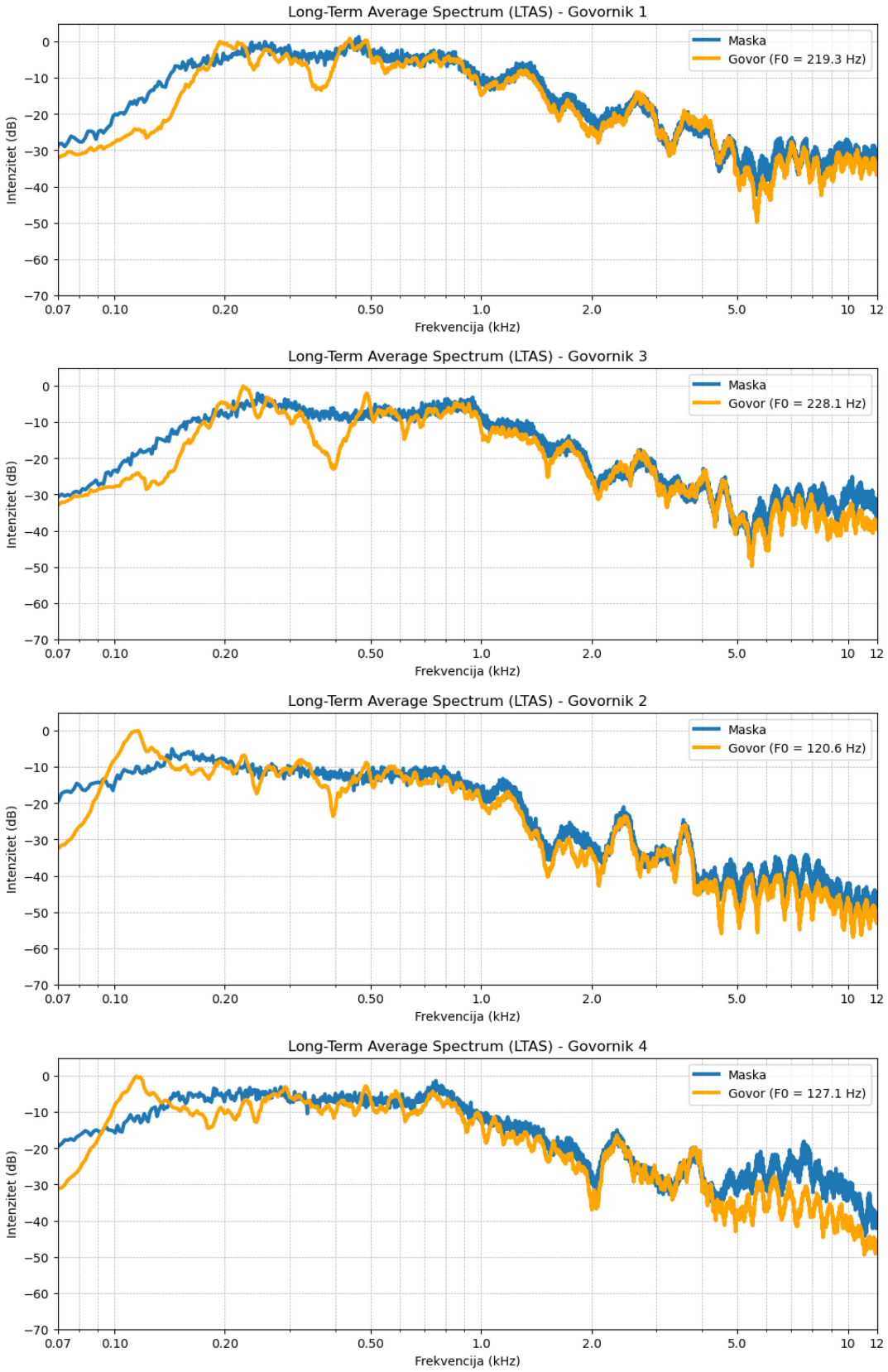
2.4. Maska šuma oblika govora (SSN)

Maska se napravila u programu Praat (Boersma i Weenink, 2024) uz pomoć već postojeće Praat scripte (Quene i Van Delft, 2010) za generiranje šuma prema obliku LTAS-u

govora (SSN). Svakom govorniku napravila se vlastita maska temeljenom njihovim snimljenim materijala (plava crta, Slika 4.), kako bi se izbjeglo uprosječivanje koje bi moglo više maskirati pojedine govornike. Dobivena maska i govorni materijal spojili su se koristeći vlastiti Praat script[10] na način da je maska počela 500 ms prije početka govornog materijal i završila 500 ms nakon. Početak i kraj imaju postepeno poglašavanje prvih, odnosno stišavanje zadnjih 100 ms kako bi se izbjegli klikovi prilikom reprodukcije snimki. Tim načinom, maska i govorni materijal bili su spojeni u tri različita SNR omjera (-2 dB, 0 dB i 2 dB). SNR omjeri postigli su se unutar Praata (Modify → Scale intensity…) tako što se govornom signalu namjestio intenzitet na 68 dB SPL (engl. *sound-pressure level*; prema uzoru na Lu i Cooke, 2009), a intenzitet pripadajuće maske na 66, 68 i 70 dB SPL prije spajanja.

Sveukupno je dobivene su tri rotacije različitog omjera signala i buke po 84 rečenice. Spremljene su kao .mp3 audio datoteke zbog prostornog ograničenja servera upitnika.

[10] https://github.com/porupski/diplomski/ (praat scripts/praat_script_Prepare_for_GLIMPSING_v3.praat)



Slika 4. LTAS sva četiri govornika (1 i 3 ženski, 2 i 4 muški govornici) u narančastom te njihova pripadajuća SSN maska u plavom.

2.5. Izrada i provođenje upitnika

Provođenje istraživanja provelo se preko online platforme SoSci Survey. Istraživanje dobilo je odobrenje Etičkog odbora održanoj 3. lipnja 2024. godine na Odsjeku za fonetiku. Upitnik se sastojao od informacija o istraživanju i suglasnosti o sudjelovanju, demografskim pitanjima (sluh, dob, materinski jezik i dr.), ugođavanje glasnoće, 3 primjera zadatka za vježbu te govornog materijala. Govorni materijal činile su 84 rečenice, odnosno 21 rečenica po četiri govornika. Niti jedan govornik nije imao istu rečenicu, ali svi su imali iste stimuluse – hrvatske konsonante u /aCa/ obliku na kraju svake rečenice. Govorni materijal maskirao se u tri različita SNR omjera, odnosno rotacije. Prva rotacija sastojala se od 84 najbučnijih snimki (SNR = -2 dB), druga rotacija od 84 manje bučnih snimki (SNR = 0 dB) te treća rotacija od 84 najmanje bučnih snimki (SNR = +2 dB). Svaki studionik dobio je nasumičan redoslijed snimki za svaku pojedinu rotaciju.

Upitnik je bio gamificiran (engl. *gamified*, pretvoren u igru) prema savjetima Harms i sur (2015) te Oliveira i Paula (2020) na način da se primjenila fiktivna narativna[11] kako bi se potaknulo sudionike na što kvalitetnije ispunjavanje upitnika (vidi Prilog C). Između svake rotacije pojavila bi se nova slika s opisom tvrdeći „da pomaže u razabriljivosti sljedećih snimki“, imajući na umu da je svaka rotacija bila sve više razgovjetnija po dizajnu.

[11] Narativa započinje s prikazom dvije AI (engl. *artificial inteligence,* umjetna inteligencija) generirane slike. Obje prikazuje nekog čovjeka („arheologa“) koji drži neobično velike glinene ploče, ne previše različite od gramofonskih. Opis ispod slika navodi kako su „pronađeni drevni audiozapisi u glini nedaleko od Krapine. Grupa znanstvenika uspjela je digitalizirati zvuk. Nažalost, zbog materijala i starosti, šum je neizbježan element snimke. U svrhe dokumentacije ovog drevnog jezika, obratite pažnju na zadnju riječ u snimci te odaberite glasnik kojega ste čuli.“ Nakon završene prve trećine (SNR = -2 dB), nova stranica upitnika otkriva sliku papirnatog svitka u *pixel art* stilu (ilustrirala Nela Gluhak) ispod kojeg piše „Drevna gramatika“ i njezinim atributom „+1 razumijevanje“. Upitnik objašnjava kako su „u međuvremenu, našli drevni zapis gramatike ovog neobičnog jezika i vjeruju da će Vam pomoći u boljem razumijevanju nadolazećih snimki!“ Ovo je tehnički faktualno istinito jer nadolazeće rotacije imaju povoljniji (veći) SNR omjer. Nakon druge rotacije (SNR = 0 dB), nova stranica upitnika otkriva novu sliku – ovaj put sliku „Drevnih krpica sa zeljem“ i atributom „+1 razumijevanje“. Upitnik objašnjava: „U međuvremenu, naši znanstvenici otkrili su novu kulturološku poveznicu s ovim drevnim jezikom i naprednim računalnim modelima došli do zaključka da krpice sa zeljem povećavaju razabirljivost ovog jezika! Tko bi rekao!“ Najuzbudljiviji dio događa se nakon završene zadnje trećine, kada se prikaže slika ogromnog plavog dijamanta i naziva „Izgubljeni dijamant Kralja Tomislava“, atribut „Neprocjenjivo blago“. Autor upitnika zahvaljuje se sudionicima i objašnjava situaciju: „Hvala Vam puno što ste došli do kraja upitnika! Htio sam Vam dati neprocjenjivi dijamant Kralja Tomislava, ali bojim se da nemate više koristit od toga, pa ću Vam dati nešto što mislim da će Vam biti od veće važnosti...“ Prelaskom na zadnju stranicu upitnika, prikazana je zadnja slika u kojoj piše „1h eksperimenta“, naziva „Najveće blago fonetičara“ i atributa „+1 eksperimentalni sat“. Upitnik zatim automatski upisuje eksperimentalni sad za svakog sudionika (studenta fonetike) koji je došao do kraja. Slike upitnika mogu se naći u Prilogu C.

Upitnik nije dozvoljavao unos odabira prije nego što li je snimka završila s obzirom na to da se traženi stimulus nalazio na kraju rečenice, a nakon odabira automatski prešao na slijedeće pitanje kako bi se smanjio zamor sudionika.

Upitnik je bio dostupan sudionicima kroz vremenski period od 3 tjedana.

2.6. Statistička analiza rezultata upitnika

Sakupljeni su odgovori sveukupno 110 sudionika, prosječne dobi od 22,6 godina. Rezultati su bili obrađeni u Jupyter Lab interaktivnoj okolini, vlastitim programom „SoSci_Data“ [12] napisan u Pythonu. U prošlom poglavlju spomenuli su se faktori vezani uz slušatelja koji mogu utjecati na pouzdanost testa. Rezultati sudionika koji su označili da su atipičnog sluha bili su odbačeni. Nadalje, primjenila su se tri eliminacijska parametra – točnost odgovora, varijabilnost odgovora i z-vrijednost varijabilnosti odgovora – kako bi se uklonili odgovori koji previše odstupaju. Arbitrarno se odabrao prag točnosti od 30 %, odnosno unos pojedinog sudionika koji je točan manje od 30 % je izbačeni. Arbitrarno se odabrao prag varijabilnosti odgovora od 40 %. Varijabilnost odgovora uklanja potencijalne unose koji su bili nasumično uneseni. Z-vrijednost varijabilnosti odgovora iznosi ± 1,96 i odvaja sve rezultate koji su izvan 95 % intervala sigurnosti. Nakon eliminacije, preostalo je 103 validnih rezultata.

2.7. Akustička analiza govornog materijala

Akustična analiza govornog materijala nudi uvid u stimulus koji je prezentiran sudionicima. Imajući na umu kako se govorni stimulus sastoji od govornog signala i maske, oduzimanjem maske SSN buke od govornog stimulusa preostat će samo dijelovi govornog signala koji nisu bili maskirani, i držeći se Cookea (2006), svaki dio preostalog govornog signala koji je barem 3 dB intenzivniji, definira se kao *glimps*. Za svaki stimulus odredili su se *glimpsevi* u Jupyter Lab interaktivnoj okolini, vlastitim programom „Fon_Glimpses“ [13] napisan u Pythonu.

[12] https://github.com/porupski/diplomski (jupyter notebooks/dipl_fon-SOSCI_data_v4.2.ipynb)
[13] https://github.com/porupski/diplomski (jupyter notebooks/dipl_fon-GLIMPSES_v6.3.ipynb)

3. Rezultati i rasprava

3.1. Akustička analiza *glimpsova* u govornom materijalu

Dobiveni *glimpsovi* prikazani su MFCC-om (vidi Prilog D) za svaki od 21 glasnika, u svakom od 3 omjera SNR-a. Tipične akustičke karakterisike uveliko su već poznate (Bakran, 1996: 58-171), te nude polaznu ili referentnu točku, prema kojoj se radi poredba.

Dok vizualizirani MFCC prikazi nemaju veliku primjenu sami po sebi, ipak nude uvid gdje očekivati pojavu *glimpseva* za 21 hrvatska glasnika, s obzirom na razinu buke. Nadalje, osim mjesta pojave, može se gledati i relativna otpornost svakog glasnika. Količina *glimpsova* nam direktno ukazuje na redundantnost pojedinog glasnika. Tako je primjerice vidljivo kako oblik, mjesto i količina *glimpsova* u stimulusu /ača/ je relativno konstantna kroz sva tri SNR omjera, dok s druge strane, stimulus /aža/ je značajno oskudniji u bučnijoj (SNR = -2) nego li u manje bučnoj (SNR = +2) okolini. Logično, što je povoljniji SNR uvjet, to je više redundantnosti, odnosno *glimpsova* u nekom stimulusu.

Pregledom MFCC prikaza kroz SNR uvjete unutar istog stimulusa, može se uočiti mjesto pojave redundantnosti. Stimulus /asa/, odnosno konsonant /s/ ima pojavu (frikciju) svojstvenih *glimpsova* između otprilike 4 i 20 kHz u prikazu. Promjenom razine buke, vidljivo je kako konsonant /s/ zadržava svoj oblik te se popunjava od donjeg dijela oblika (4 kHz) prema gore. Razlog toga je vjerojatno velika redundantnost konsonanta /s/ u sva tri okruženja, odnosno nedovoljno jako maskiranje. S druge strane, stimulus /aša/, odnosno konsonant /š/ ima uočljivo zadebljanje na otprilike 4,7 kHz u prikazu. Kako se buka smanjuje, tako je moguće uočiti jače popunjavanje ispod, nego iznad zadebljanja. Takvo opažanje odgovara spektralnom obliku maske, koja jače maskira niže frekvencije nego više frekvencije (Slika 4.). Kod nekih stimulusa moguće je uočiti i tranzijente kao razlikovno obilježje. Tako npr. između stimulusa /ata/ i /aka/, vidi se razlika između vodoravnih formanata u vokalu /a/ prije okluzije /k/, u kontrastu vokala /a/ prije okluzije /t/ gdje drugi formant ima nagib. Neki stimulusi djeluju gotovo identični kada su prikazani kao MFCC, kao npr. /ama/ i /ana/, no vrijedi naglasiti da to što čovjek može uočiti u MFCC prikazu, ne znači da stroj isto može, ali i obratno.

3.2. Matrice zabune

Pošto se svaki rezultat sastoji se od 3 rotacije (SNR -2, 0 i +2), a svaka rotacija od 84 stimulusa, pri čemu se svaki stimulus snimio četiri puta (govornika), to znači da je jedan stimulus prezentiran četiri puta u svakoj rotaciji. Odgovor sudionika za svaki stimulus uprosječio se za sva četiri govornika, ali unutar pojedine rotacije. Rezultat je prikazan u obliku tablice, odnosno matrice zabune za svaku rotaciju (Tablica 1.; Tablica 2.; Tablica 3.). Tablice zabune pokazuju ako je slušatelj zamijenio glasnik u stimulusu drugim glasnikom, i ako je, s kojim glasnikom. Rezultati manji od 2,5 % (< 11) izostavljeni su iz Tablica 1.-3. zbog preglednosti, a zabune veće od 10% (> 41) prikazane su u Tablici 4.

Zbog eliminacijskog koraka opisanog u poglavlju 2.6., odnosno zbog osigurane razine točnosti odgovora, dijagonala vidljiva u tablicama zabune je očekivana. Drugim riječima, ako bi svi sudionici točno odgovorili na svaki stimulus, u tablicama zabune bi se vidjela isključivo dijagonala. Izvan dijagonale, vidljiva je pak raspodjela netočnih odgovora, odnosno zabuna. Gledajući tablice vertikalno, može se uočiti kako su sudionici često mijenjali stimulus s okluzivima /p, t, k, b, d, g/ i ostalim /m, n, j, r, l, v/ glasnicima neovisno o stimulusu. Naprotiv toga, frikativi /s, š, ž/ čak i /f/ te afrikate /c, č, đ/ imaju relativno malo zabuna, sugerirajući na veću otpornost.

Najveću otpornost (više od 50% točnih odgovora, > 206) u najbučnijem okruženju (SNR = -2) imali su frikativi /s, š, z/, afrikate /c, č/, aproksimant /j/ i vibrant /r/. Otpornost frikativa i afrikata može se pripisati karakterističnim frikcijama pri višim frekvencijama. Aproksimant /j/ najvjerojatnije stvara otpornost svojim temporalnim karakteristikama, odnosno jedinstvenim načinom promjene spektra kroz vrijeme, kao i vibrant /r/ svojim jedinstvenim udarcima. Iznenađujuće, okluziv /b/ skoro prelazi tu granicu, s 205 točnih odgovora, dok su ostali zvučni okluzivi /d, g/ imali poprilično nisku točnost. Razliku u otpornosti zvučnih okluziva teško je pripisati zvučnosti, iz razloga što i u boljim SNR uvjetima, zvučni okluzivi /d, g/ i dalje imaju relativno nisku točnost (< 90). S druge strane, teško je tu razliku pripisati bilabijalnosti glasnika /b/ jer u istom okruženju (SNR = -2) bezvučni bilabijalni okluziv /p/ (142) pak pokazuje manju otpornost u odnosu na bezvučni velarni okluziv /k/ (182). Moguće je da ovo odstupanje potiče iz nedovoljno strogo kontroliranog snimanja govornog materijala.

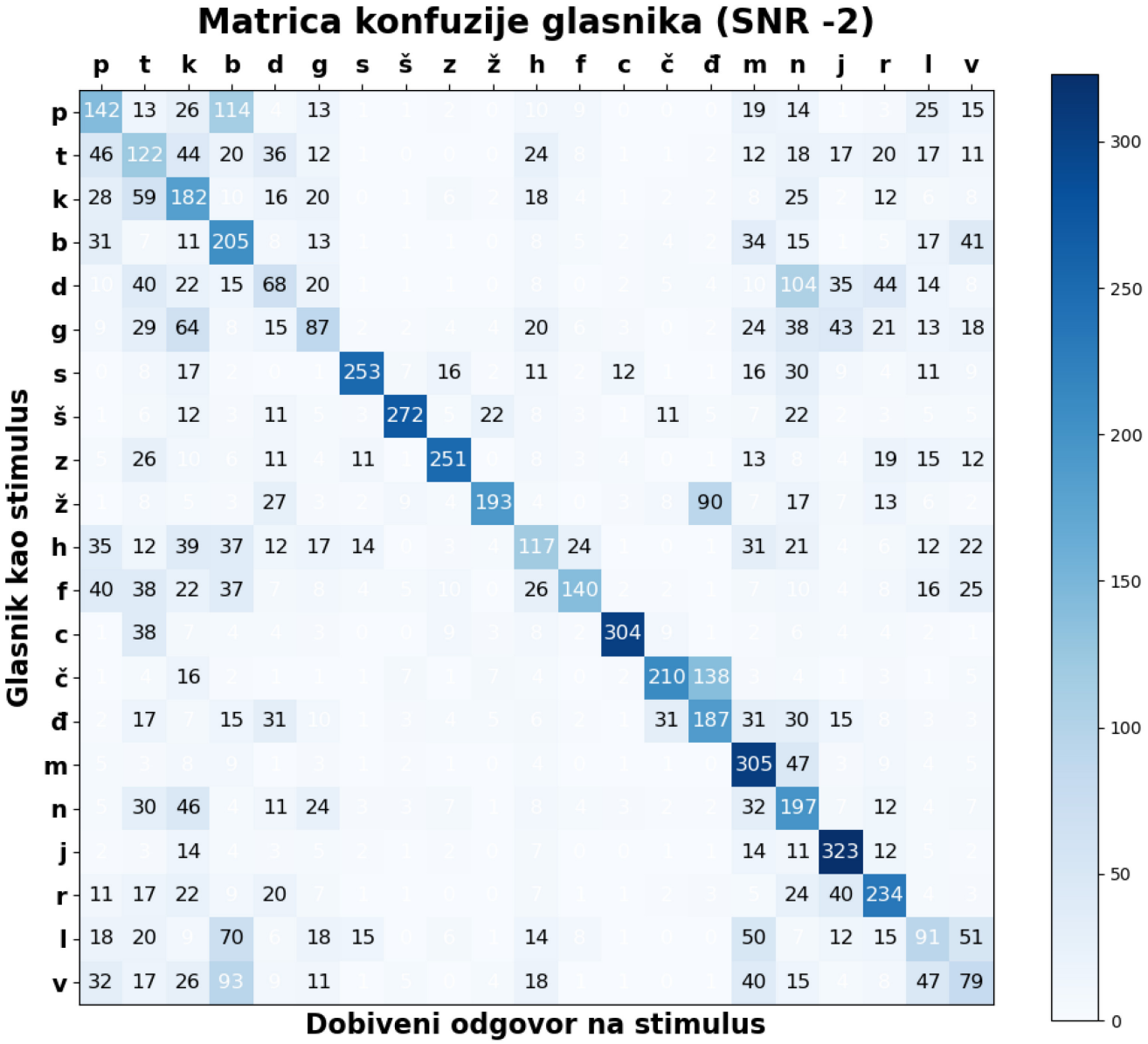
Iako je afrikata /č/ među najotpornijim glasnicima, otprilike se 40 % vremena afrikata /č/ bunila s afrikatom /đ/. Štoviše, smanjenje buke (SNR = +2) nije smanjilo zabunu, već ju je povećalo sa 138 zabuna na 166. Ovu anomaliju bi se moglo pripisati velikoj sličnosti afrikata /č/ i /đ/, pri čemu su sudionici znali da je stimulus jedna od tih afrikata, ali ne i točno koja. Osim toga, također je moguće da ova razlika proizlazi iz nedovoljno pažljivog snimanja govornog materijala.

Još jedna anomalija vidljiva je proučavanjem stimulusa zvučnog okluziva /d/. Glasnik se u velikoj mjeri mijenja s /n/, do te mjere da u najbučnijoj okolini (SNR = -2), više je sudionika stimulus /d/ identificiralo kao /n/ (104) nego kao /d/ (68). U sljedećoj okolini (SNR = 0), sudionici su već otprilike podjednako bili točni (123) i bunili /d/ s /n/ (119), dok u zadnjoj okolini (SNR = +2) je /d/ bio poprilično jasan, te ga se u manjoj mjeri bunilo s /n/. Bakran (1996: 123) spominje sličnosti između tranzijenta vokala koji okružuju nazale i koji okružuju okluzive, navodeći sličnosti unutar para /b/ – /m/, no specifično upozorava na uspoređivanje /d/ i /n/ jer se ne artikuliraju na istome mjestu. Prema tome, rezultati sugeriraju da neovisno o tome što /d/ ima mjesto artikulacije više dentalno, a /n/ alveolarno, u bučnoj okolini (SNR ≤ -2) moguća je značajna zabuna /d/ s glasnikom /n/.

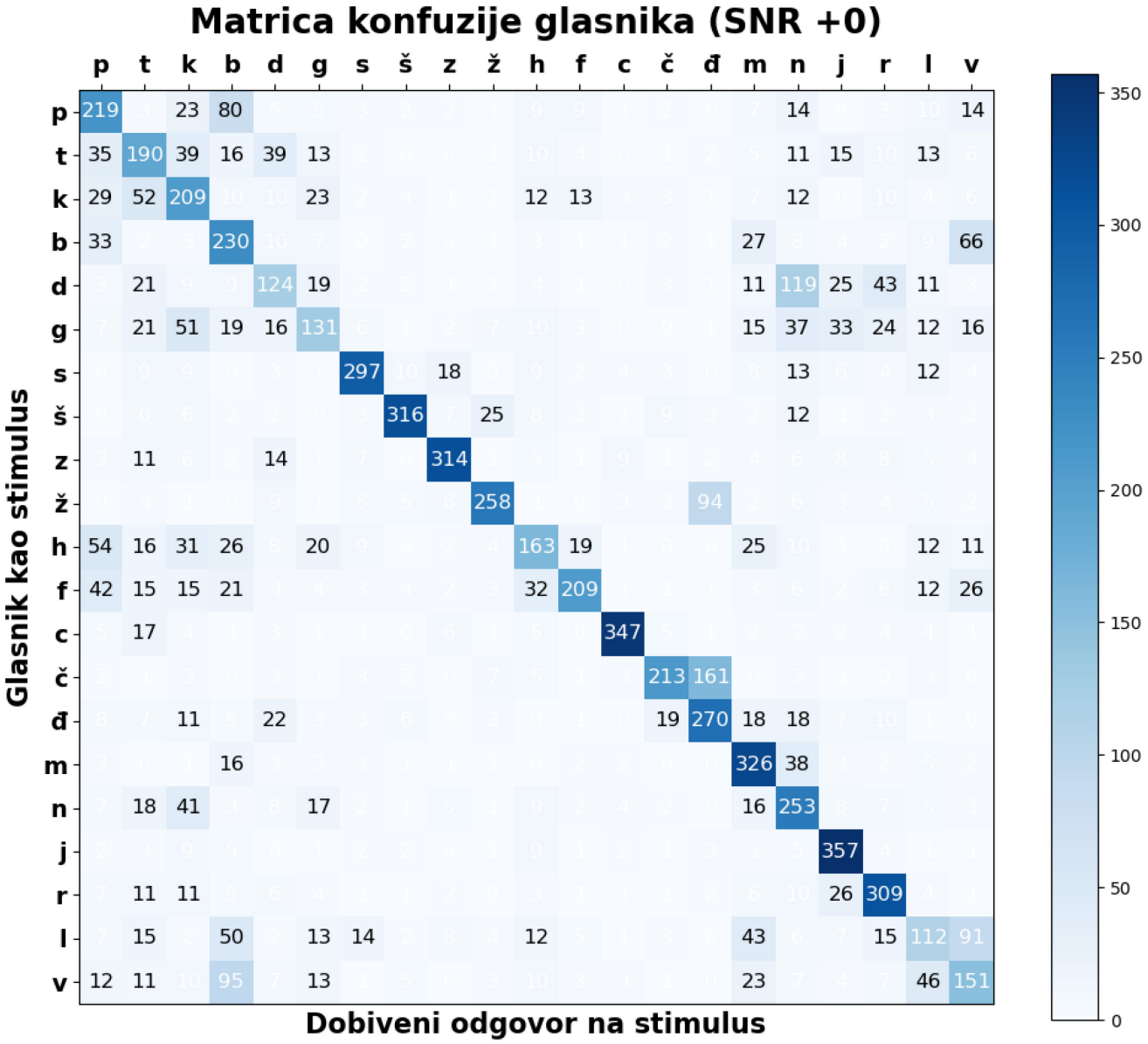
Aproksimant /v/ u najbučnijoj okolini (SNR = -2) najčešće je zamijenjen zvučnim okluzivom /b/ (93) – čak više nego točnih odgovora (79) – dok je u manjoj mjeri zamijenjen lateralom /l/ i nazalom /m/. Ovaj trend opada sa smanjenjem buke, pri čemu zamjena s okluzivom /b/ i dalje ostaje značajna (otprilike 40 % pri SNR = +2), što je u skladu s literaturom (Bakran 1996: 165). Razlog ove pojave je sličnost artikulacije aproksimanta /v/ s okluzijom okluziva /b/.

Dobiveni rezultati (Tablica 4.) pokazuju sličnosti s Cookeovim (2006: 1570) rezultatima. U oba slučaja su se dogodile zabune za parove: /b → v/, /d → n/, /k → t/, /l → m/, /v → b/, i u manjoj mjeri /f →p/. Postoje odstupanja kao što je zamjena /d → g/ u engleskom, dok u hrvatskom nema te, ali ima zamjene /d → r/ koje nema u engleskom. Isto tako, u hrvatskom je /l/ imao zabune s /b/ i /v/, dok je u engleskom /k/ imao zabune s /p/ i /f/.

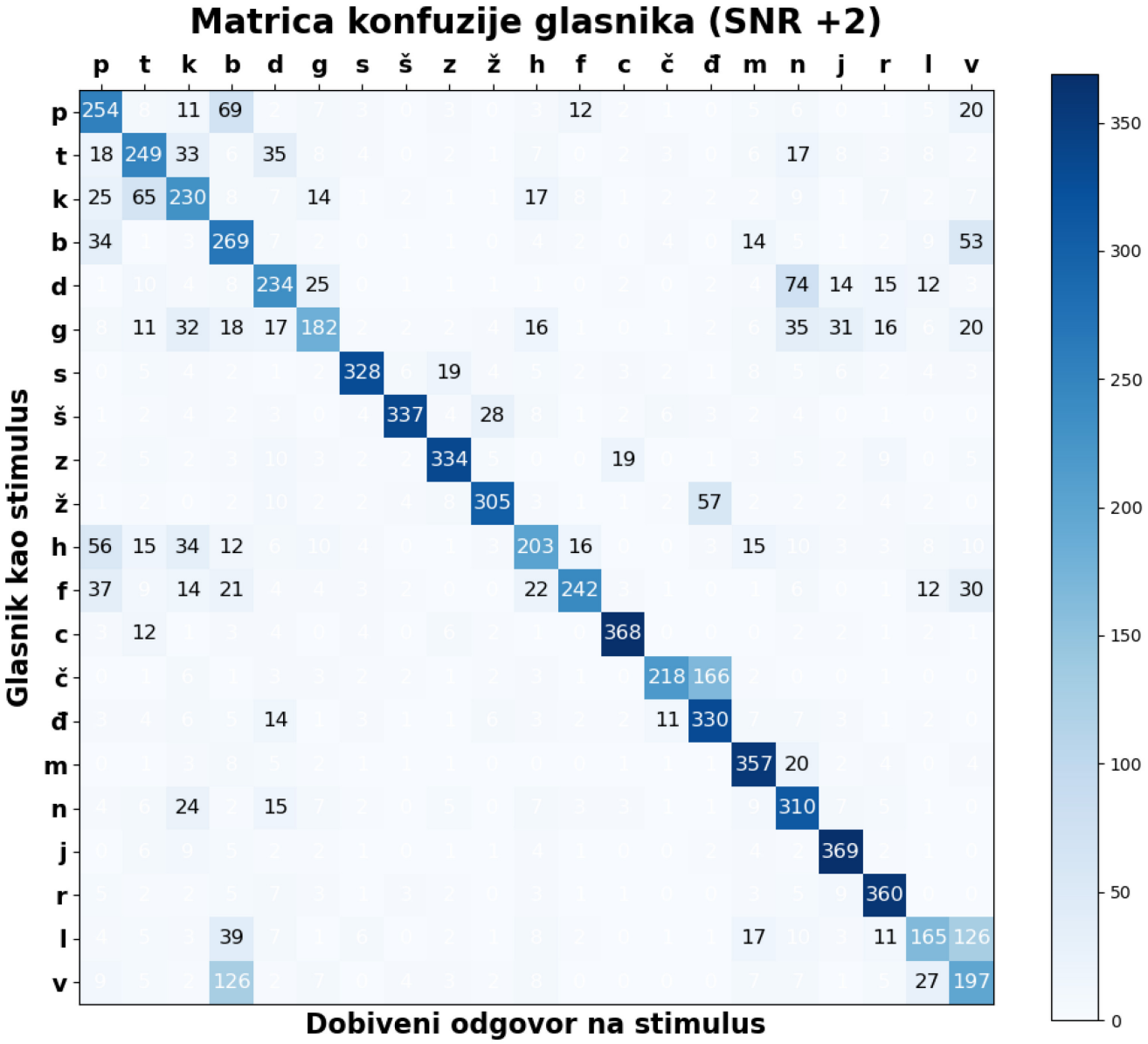
Tablica 1. Matrica zabune glasnika za SNR = -2 dB



Tablica 2. Matrica zabune glasnika za SNR = +0 dB



Tablica 3. Matrica zabune glasnika za SNR = +2 dB



Tablica 4. Prikaz zabune stimulusa u istraživanju veće od 10% (> 41) prema SNR uvjetima.

Stimulus	Zabuna (SNR = -2)	Zabuna (SNR = +0)	Zabuna (SNR = +2)
p	b	b	b
t	p, k	-	-
k	t	t	t
b	-	v	v
d	**n**, r	n, r	n
g	k, j	k	-
ž	đ	đ	đ
h	-	p	p
f	-	p	-
č	đ	đ	đ
m	n	-	-
n	k	-	-
l	b, m, v	b, m, v	v
v	**b**, l	b, l	b

4. Zaključak

Rezultati istraživanja pokazali su da su na tragu Cookea (2006). *Glimpsovi* svakako nose dovoljno informacija u hrvatskom jeziku za točnu identifikaciju, iako to nije slučaj za svaki glasnik. Glasnici se bolje prepoznaju u manje bučnom (pogodniji SNR uvjet) nego li u više bučnom okruženju. Glasnici koji imaju karakteristike među višim frekvencijam (kao npr. /s, š, z, c/) pokazali su se relativno otpornima na maskiranje bukom. Glasnici /m, j, r/ su se također pokazali relativno otpornima. Najmanje otporni pokazali su se glasnici /d, g, h, l, v/. Glasnici /ž, č/ često su se zamijenili s glasnikom /đ/, dok se glasnik /đ/ nije često mijenjao njima. Dobiveni rezultati (Tablica 4.) pokazuju sličnosti s Cookeovim (2006: 1570) rezultatima. Dok su hrvatski glasnici relativno otporni na maskiranje bukom, ovaj rad ističe iznimke koje bi mogle stvarati problema u svakodnevnoj, bučnoj komunikaciji.

5. Literatura

1. Bakran, J. (1996). Zvučna slika hrvatskoga govora. Ibis grafika.

2. Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: Analyzing text with the natural language toolkit. O'Reilly Media.

3. Boersma, P., & Weenink, D. (2024). Praat: Doing phonetics by computer (Version 6.4.12) [Computer software]. http://www.praat.org/ (Pristupljeno 2. svibnja 2024.)

4. Chen, J., & Chang, H. O. (2022). Sketching the Landscape of Speech Perception Research (2000–2020): A Bibliometric Study. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.822241

5. Cooke, M. (2006). A glimpsing model of speech perception in noise. The Journal of the Acoustical Society of America, 119(3), 1562-1573.

6. Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366. doi:10.1109/tassp.1980.1163420

7. Harms, J., Biegler, S., Wimmer, C., Kappel, K., & Grechenig, T. (2015). Gamification of online surveys: Design process, case study, and evaluation. INSO Research Group, Vienna University of Technology, Vienna, Austria.

8. Horga, D., & Liker, M. (2016). Artikulacijska fonetika - Anatomija i fiziologija izgovora. Ibis grafika.

9. Jelaković, T. (1978). Zvuk, sluh, arhitektonska akustika. Školska knjiga.

10. Johnson, K. (2012). Acoustic and auditory phonetics (3rd ed). Wiley-Blackwell. ISBN 9781444343076.

11. Jurafsky, D., & Martin, J. H. (2009). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (2nd ed.). Pearson.

12. Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. The Journal of the Acoustical Society of America, 61(5), 1337-1351. https://doi.org/10.1121/1.381436

13. Kishon-Rabin, L., Gam, S., Shiff, T., Rembrand, R., & Ari-Even Roth, D. (2008). Speech perception enhanced by noise in listeners with normal hearing. Journal of the Association for Research in Otolaryngology, 19. https://doi.org/10.1515/JBCPP.2008.19.3-4.237

14. Korvel, G., Kąkol, K., Kurasova, O., & Kostek, B. (2020). Evaluation of Lombard speech models in the context of speech in noise enhancement. IEEE Access, 8, 155156-155170. https://doi.org/10.1109/ACCESS.2020.3015421

15. Li, G., & Lutman, M. E. (2006). Sparseness and speech perception in noise. Proceedings of Interspeech. https://doi.org/10.21437/Interspeech.2006-562

16. Li, T., & Fu, Q.-J. (2010). Effects of spectral shifting on speech perception in noise. Hearing Research, 270(1-2), 81-88. https://doi.org/10.1016/j.heares.2010.09.012

17. Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. Speech Communication, 51(12). https://doi.org/10.1016/J.SPECOM.2009.07.002

18. Ljubešić, N., Koržinek, D., Rupnik, P., & Jazbec, I. (2022). ParlaSpeech-HR - a Freely Available ASR Dataset for Croatian Bootstrapped from the ParlaMint Corpus. PARLACLARIN. http://hdl.handle.net/11356/1494

19. Ljubešić, N., Koržinek, D., & Rupnik, P. (2024). Parliamentary spoken corpus of Croatian ParlaSpeech-HR 2.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042. http://hdl.handle.net/11356/1914

20. McDermott, J. H. (2009). The cocktail party problem. Current Biology, 19(22), R1024–1027. https://doi.org/10.1016/j.cub.2009.09.005

21. Oliveira, K. W. R., & Paula, M. M. V. (2020). Gamification of online surveys: A systematic mapping. IEEE Transactions on Games. https://doi.org/10.1109/TG.2020.3004366

22. Quené, H., & van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. Speech Communication, 52(11-12), 911-918. https://doi.org/10.1016/j.specom.2010.08.004

23. Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., & Wang, X. (1999). Consonant recordings for speech testing. Journal of the Acoustical Society of America, 106, L71–L74.

24. Škarić, I. (1991). Fonetika hrvatskoga književnog jezika. U S. Babić, D. Brozović, M. Moguš, S. Pavešić, & S. Težak (Ur.), Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika: Nacrti za gramatiku (str. 61-454). Hrvatska akademija znanosti i umjetnosti, Globus, Nakladni zavod.

25. Škarić, I. (2007). "Fonetika hrvatskoga književnoga jezika." U Glasovi i oblici hrvatskoga književnoga jezika, str. 17. Zagreb: Nakladni zavod Globus.

26. Tambovtsev, Y., & Martindale, C. (2007). Phoneme frequencies follow a Yule distribution. SKASE Journal of Theoretical Linguistics, 4(2), 1-11. http://www.skase.sk/Volumes/JTL09/ (Pristupljeno 19.12.2023.)

27. Wang, X., & Xu, L. (2021). Speech perception in noise: Masking and unmasking. Journal of the Acoustical Society of America, 16(2). https://doi.org/10.1016/J.JOTO.2020.12.001

28. Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. Journal of Computer Science and Technology, 16(6), 582–589. doi:10.1007/bf02943243

Otpornost hrvatskih glasnika na maskiranje bukom

Sažetak

Rad se bavi ispitivanjem otpornosti 21 hrvatskih glasnika na različite razine buke, primjenjujući teoriju *glimpseva* percepcije govora u bučnim okruženjima. Za analizu je korišten govorni materijal koji je prilagođen hrvatskom jeziku, dok je kao maska korištena buka govora (speech shaped noise, SNN) specifična za te govornike. Istraživanje je provedeno putem online upitnika. *Glimpsevi* u govornom materijalu analizirani su korištenjem MFCC prikaza, a rezultati upitnika prikazani su kroz matricu zabune. Glasnici se bolje prepoznaju u manje bučnom nego li u više bučnom okruženju, a glasnici koji imaju karakteristike među višim frekvencijama pokazali su se relativno otpornima na maskiranje bukom.

Ključne riječi: otpornost glasnika, teorija *glimpseva*, percepcija govora, buka govora, hrvatski jezik

Resistance of Croatian phonemes to noise masking

Abstract

This paper investigates the resistance of 21 Croatian phonemes to various levels of noise, applying the "glimpsing" theory of speech perception in noisy environments. The analysis utilized speech material adapted to the Croatian language, with the masking noise being speech-shaped noise (SNN) specific to the speakers. The study was conducted using an online questionnaire. Glimpses in the speech material were analyzed using MFCC representations, while the results from the questionnaire were presented through a confusion matrix. Phonemes are better recognized in less noisy than in more noisy environments, and phonemes with characteristics in higher frequencies proved to be relatively more resistant to noise masking.

Keywords: phoneme resistance, glimpsing theory, speech perception, speech-shaped noise, Croatian language

Prilog

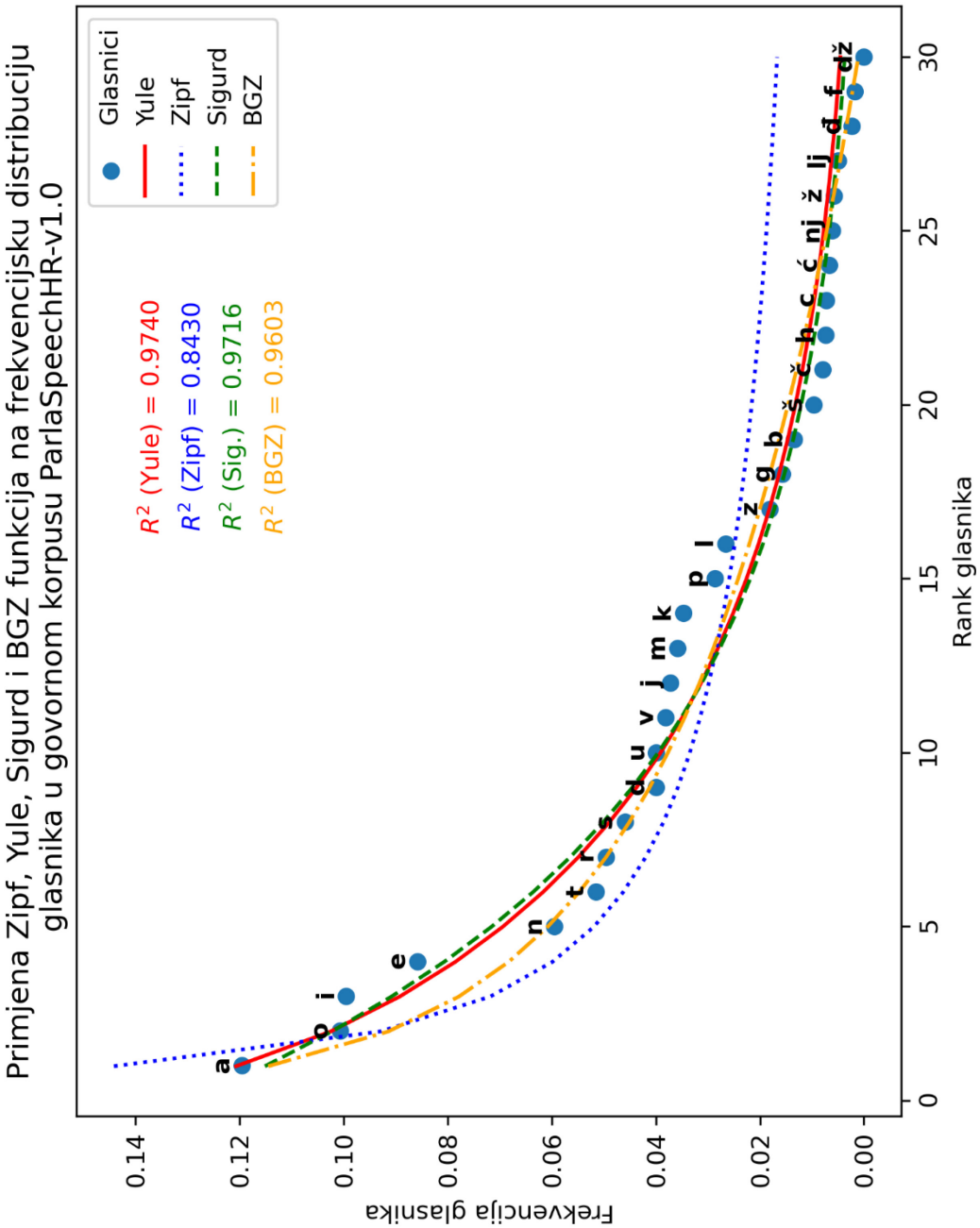
Prilog A – fonetska frekvencija u ParlaSpeechHR tekstnom korpusu

Prilog B – stvoreni govorni materijal, rečenice

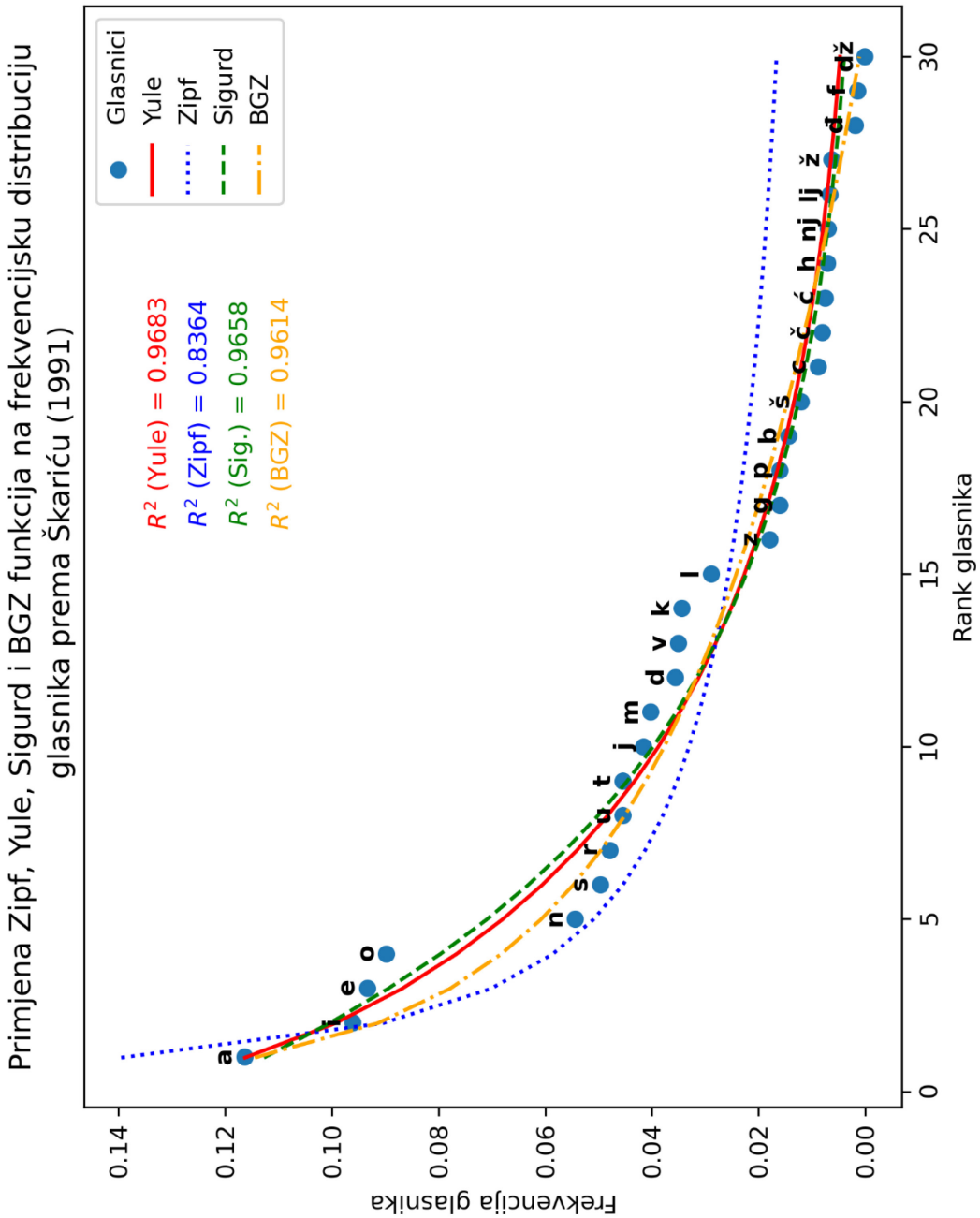
Prilog C – slike SoSci upitnika

Prilog D – MFCC prikaz *glimpsova* za svaki od 21 odabrani hrvatski konsonant

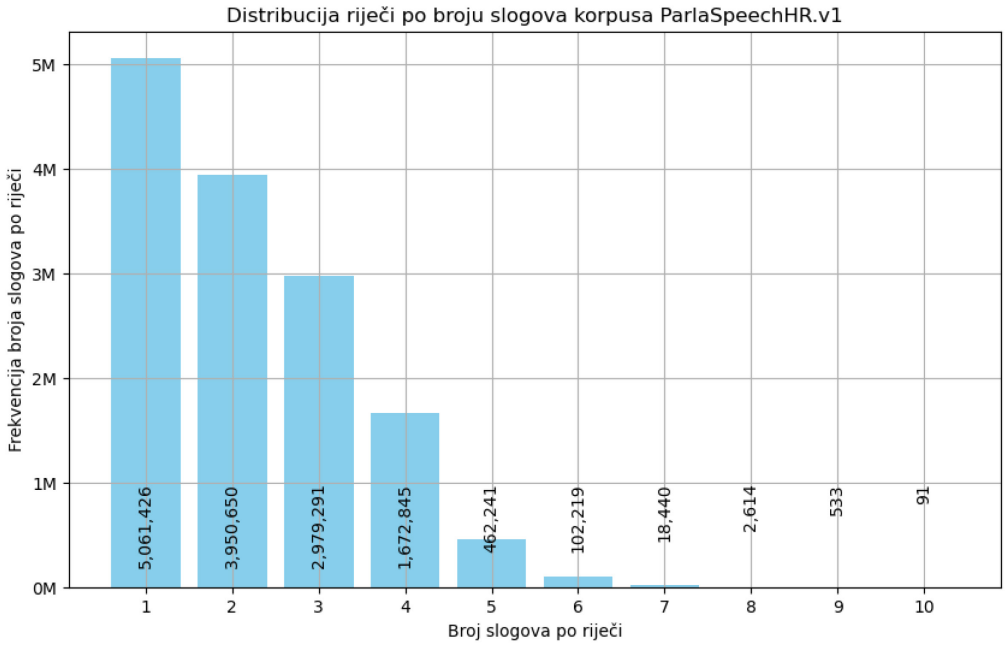
Prilog A – Referentna čestota fonema



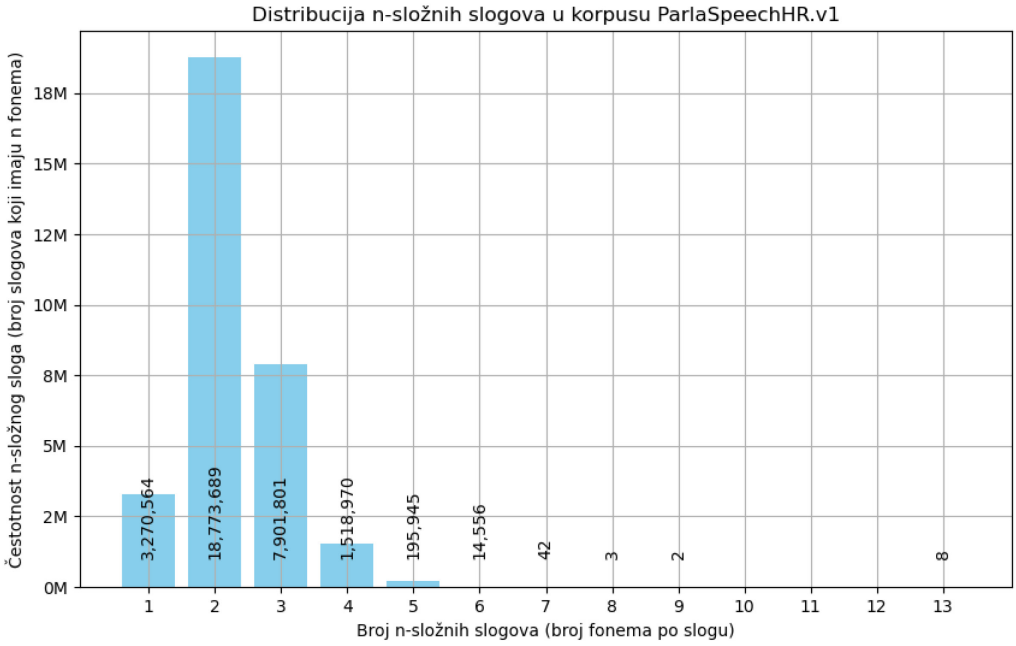
Slika A1. Prikaz frekvencije fonema u tekstnom korpusu ParlaSpeechHR-v1.0.



Slika A2. Prikaz frekvencije fonema prema Škariću (1991: 352).



Slika A3. Raspodjela riječi s obzirom na to koliko slogova imaju, iz ParlaSpeechHR-v1.0 tekstnog korpusa. Riječi s 6 ili više slogova su najvjerojatnije loše transkripcije unutar tekstnog korpusa.



Slika A4. Raspodjela slogova s obzirom na to koliko fonema imaju, iz ParlaSpeechHR-v1.0 tekstnog korpusa. Slogovi s 6 ili više fonema su najvjerojatnije loše transkripcije unutar tekstnog korpusa.

Prilog B – Govorni materijal

1. ji pri diti kom apa
2. seti ku ni di ata
3. nas vašbe be ona aka
4. eji ko kovi jumi aba
5. daje nia invač ža ada
6. alju rod jaje tui aga
7. brojor bi li modrš asa
8. oza bislo ti koćen aša
9. navom vatsnu bol suko aza
10. ci todni bivi ćeđe aža
11. uu sente isje jeza afa
12. innog ti ido tosus aha
13. čnenje nosna oli toza aca
14. svoj mepo gotro vistal ača
15. toja jui oa ćešto ađa
16. čišto tu na loku aja
17. ku ni uiš cika ana
18. vniđe rastu pos tene ama
19. kostčin tio dog jek ara
20. kono pris skloka jes ala
21. vičla na priko kune ava
22. amžje sevo pode kimko apa
23. neo vi kasli nake ata
24. modi une viko niko aka
25. nau na uveć nati aba
26. ov pr i ada
27. prijoj uba su pra aga
28. dnota getsja na vnibra asa
29. zatri seze bor deve aša
30. vom čajpla no piza aza
31. kor jigle to tovan aža
32. virpra e vreči udo afa
33. to vjeno kotram cidi aha
34. kvimve htioci rža poho aca
35. tone tavše vra podo ača
36. ko pota vnomiz dje ađa
37. pra prati zma doko aja
38. no ko vi ja ana
39. nisbli te nisto uje ama
40. ve dije vakno vanpo ara
41. i suidr tedo pojen ala
42. bu nog pa oddi ava

43. za snišni ma ona apa
44. okal babra ravins dnak ata
45. ste e avlas da aka
46. na bi vjetje va aba
47. srps njevni ji ida ada
48. korje berđe pretu ti aga
49. nani o sladru viće asa
50. šendi i is li aša
51. za djelo ža ra aza
52. đusno itit ida go aža
53. klubkva koj klevor slgov afa
54. svimto muvre sao velo aha
55. zbog gor vace unu aca
56. đoo prakka o je ača
57. pona pakri tizo neje ađa
58. duste tibu vi na aja
59. sveva će propos zah ana
60. svimko pi karsi pu ama
61. ci vats tajte por ara
62. zlogci tur melu tako ala
63. ćnime isi u žeti ava
64. ce ase tevnim zor apa
65. dao zaže stanje vablo ata
66. dra por doti jepro aka
67. rime tnemi niod neva aba
68. dre ključe mae kaho ada
69. bavan na je juzna aga
70. jeu nini če no asa
71. pra vnovu jina manzo aša
72. tigos tnišin de lo aza
73. e mibi nudi ce aža
74. glasiz pr vlas kažiš afa
75. u produr njuvla žnigo aha
76. davo kata dao jeka aca
77. ako jem nao če ača
78. uno bzirne fa jiko ađa
79. slode vreži su de aja
80. sta ko voru prasvat ana
81. zašvi surjih ču prosje ama
82. to zi či juji ara
83. vre žuje sute toje ala
84. da jevins dadan zašti ava

Informacije o generiraju:
max_syllables(2),
n_sentences(84),
n_words_per_sentence(4)
max_SPW_length (5),
max_PPS_length (4),
max_UQS_length (7000)

Korelacija s fonetskim balansom:
R^2 = 0,8522

Prilog C – SoSci Upitnik

Pronađeni su drevni audiozapisi u glini nedaleko od Krapine. Grupa znanstvenika uspjela je digitalizirati zvuk.

Nažalost, zbog materijala i starosti, šum je neizbježan element snimke.

U svrhe dokumentacije ovog drevnog jezika, molim Vas **obratite pažnju na zadnju riječ u snimci te odaberite glasnik kojega ste čuli**.

Snimke možete preslušavati neograničen broj puta.

Ako nakon drugog ili trećeg slušanja i dalje niste sigurni, odaberite što mislite da je najbliže onomu što ste čuli.

0:00 / 0:02

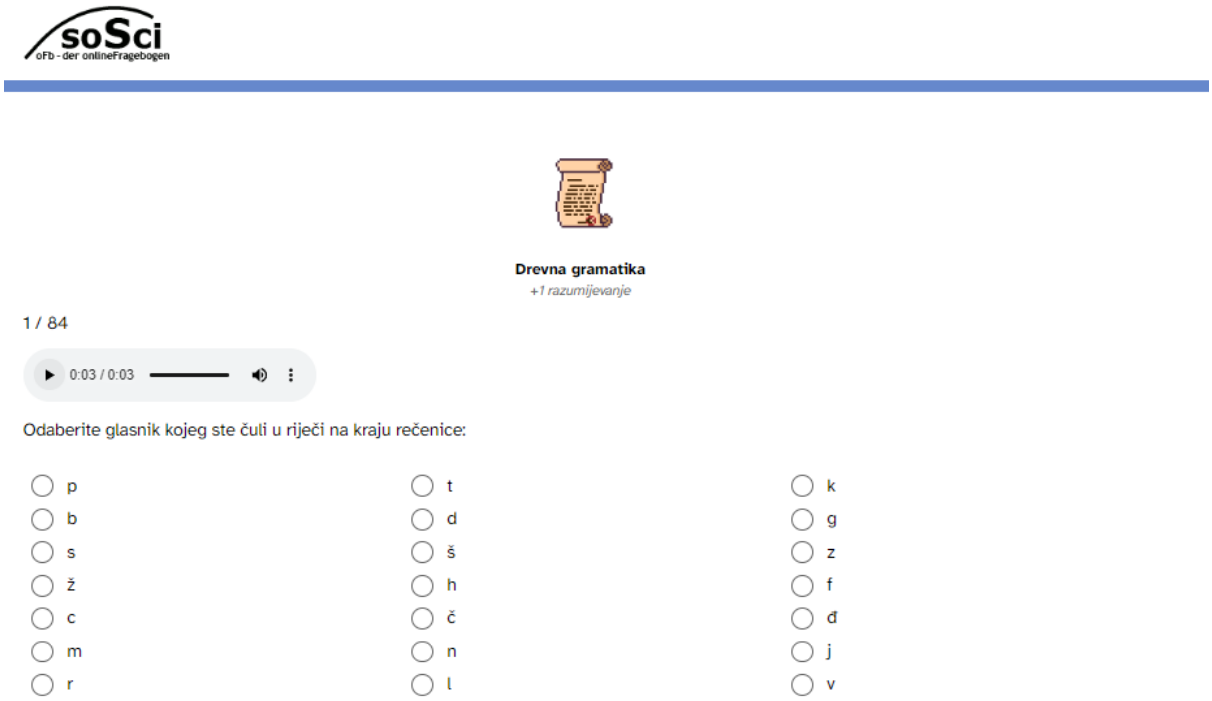
/ji pri diti kom a_a/

Odaberite glasnik kojeg ste čuli u riječi na kraju rečenice:

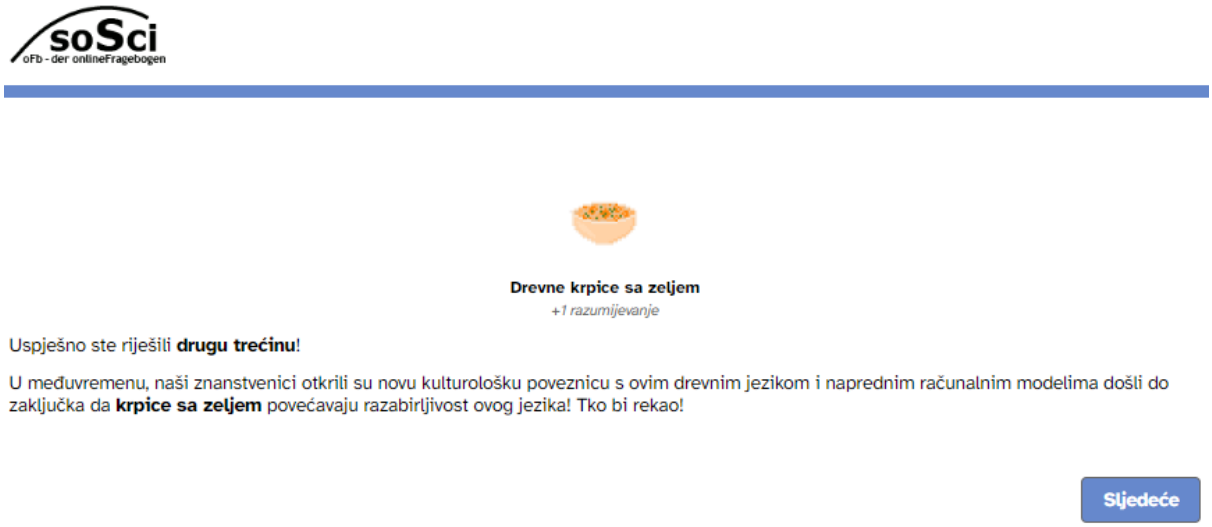
p	t	k
b	d	g
s	š	z
ž	h	f
c	č	đ
m	n	j
r	l	v

Sljedeće

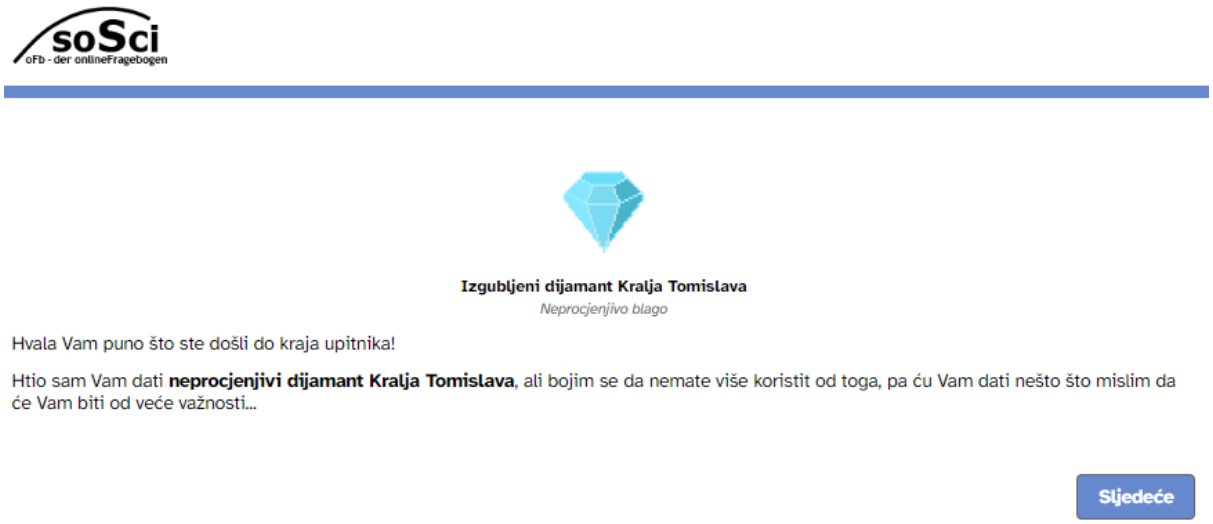
Slika C1. Prikaz fiktivne narative i prvog zadatka za vježbu



Slika C2. Prikaz prvog od 84 stimulusa u drugoj rotaciji (nakon dobivene „Drevne gramatike“).



Slika C3. Prikaz druge trećine upitnika, gdje se sudionika nagrađuje „poboljšanim razumijevanjem“ u obliku Drevnih krpica sa zeljem.

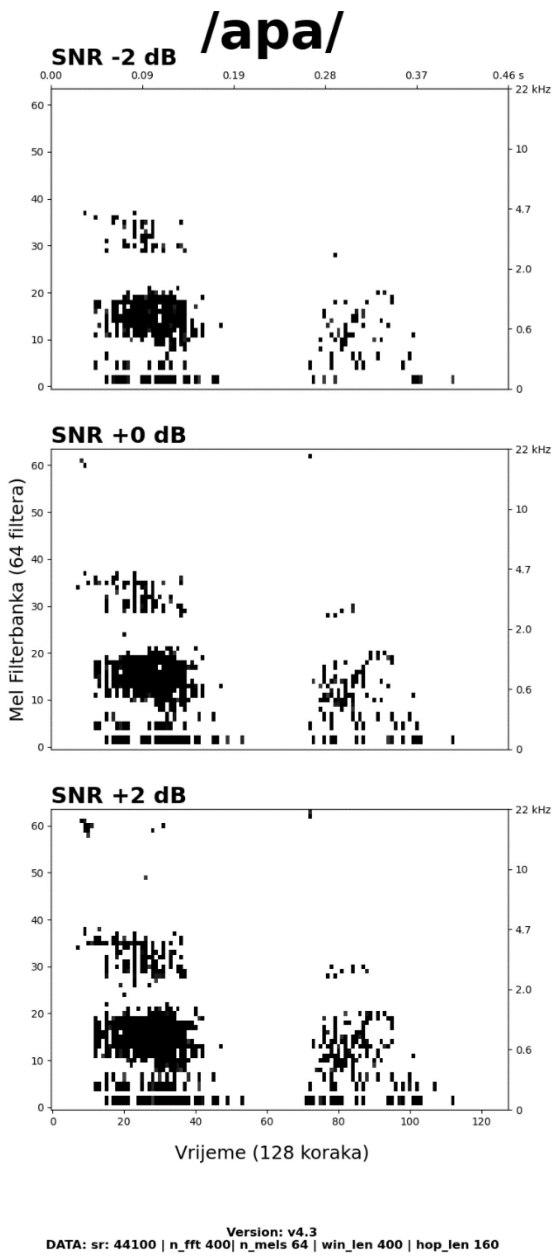


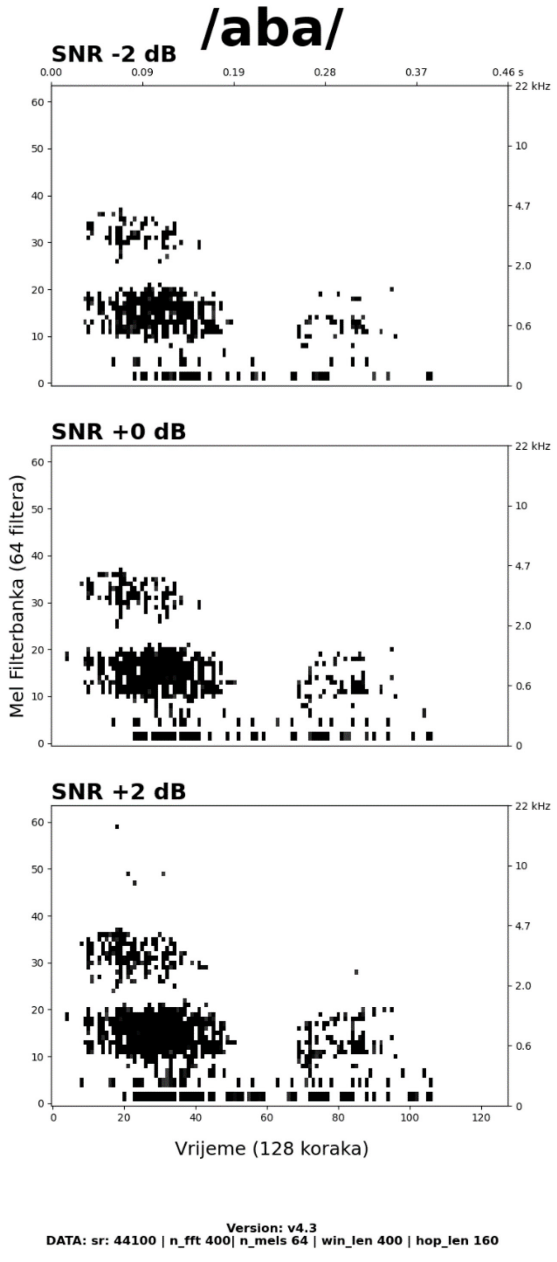
Slika C4. Predzadnja stranica upitnika sa šaljivom porukom.

Slika C5. Kraj upitnika sa šaljivom porukom.

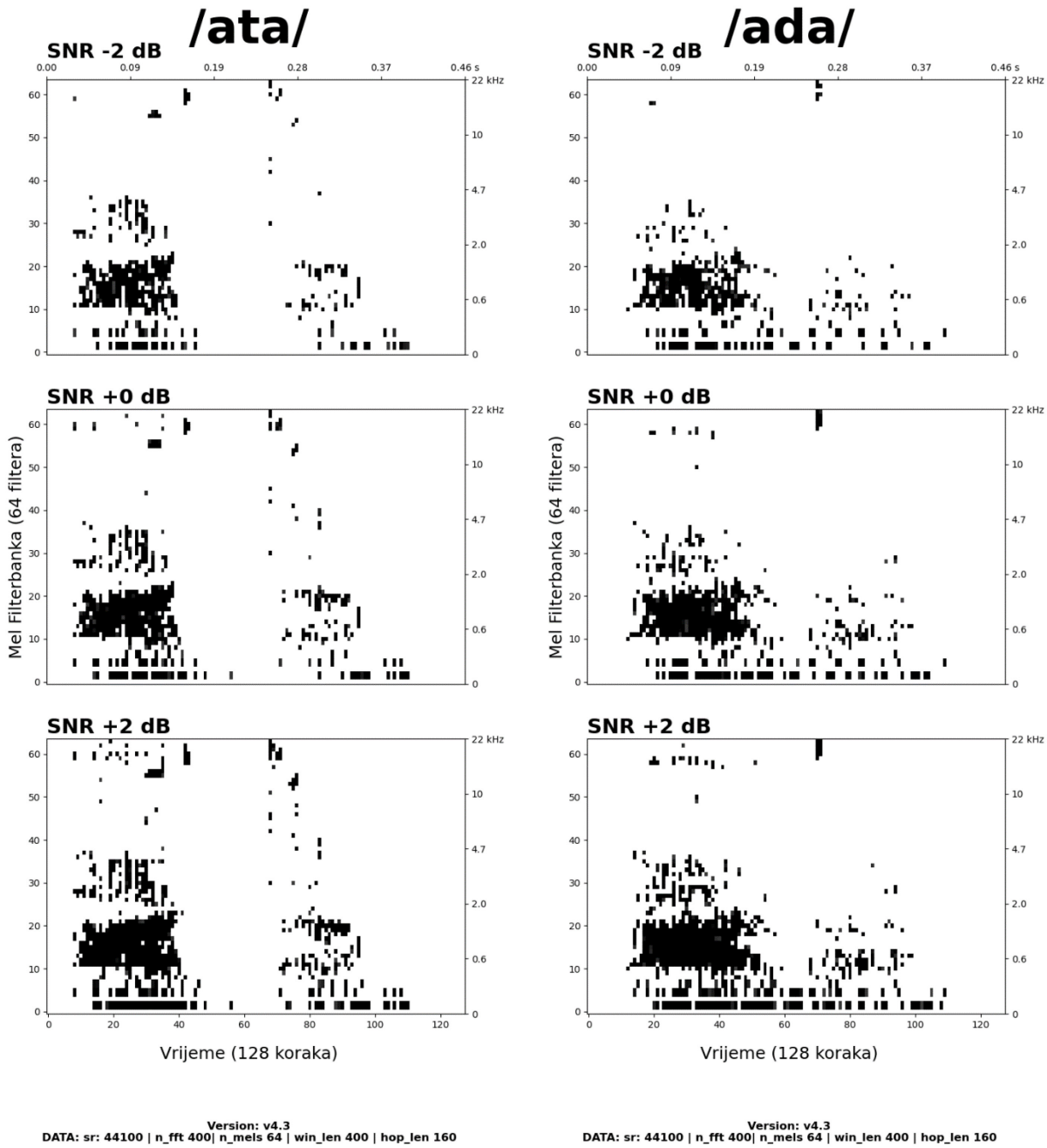
Prilog D – MFCC prikazi

/apa/, /aba/

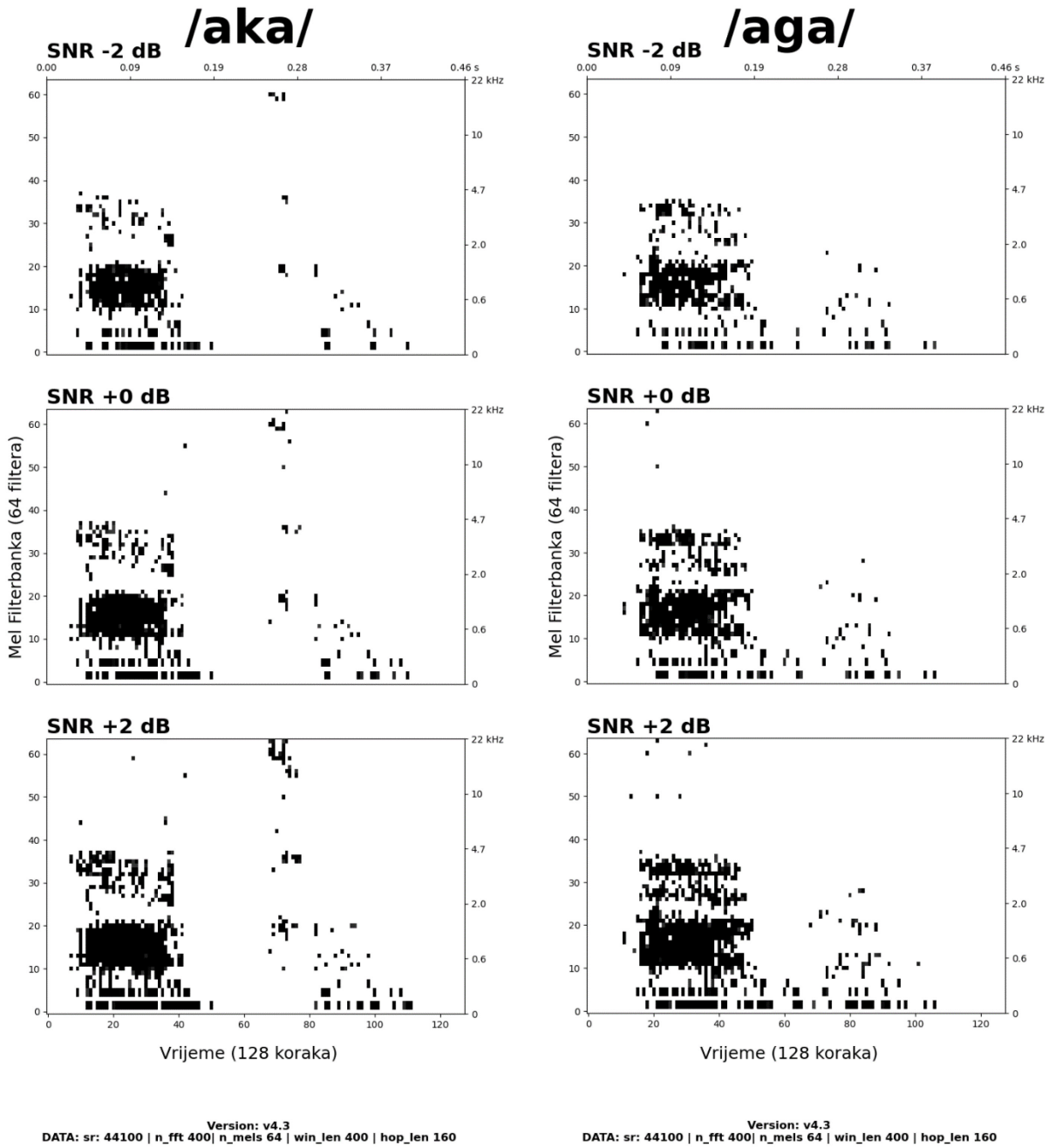




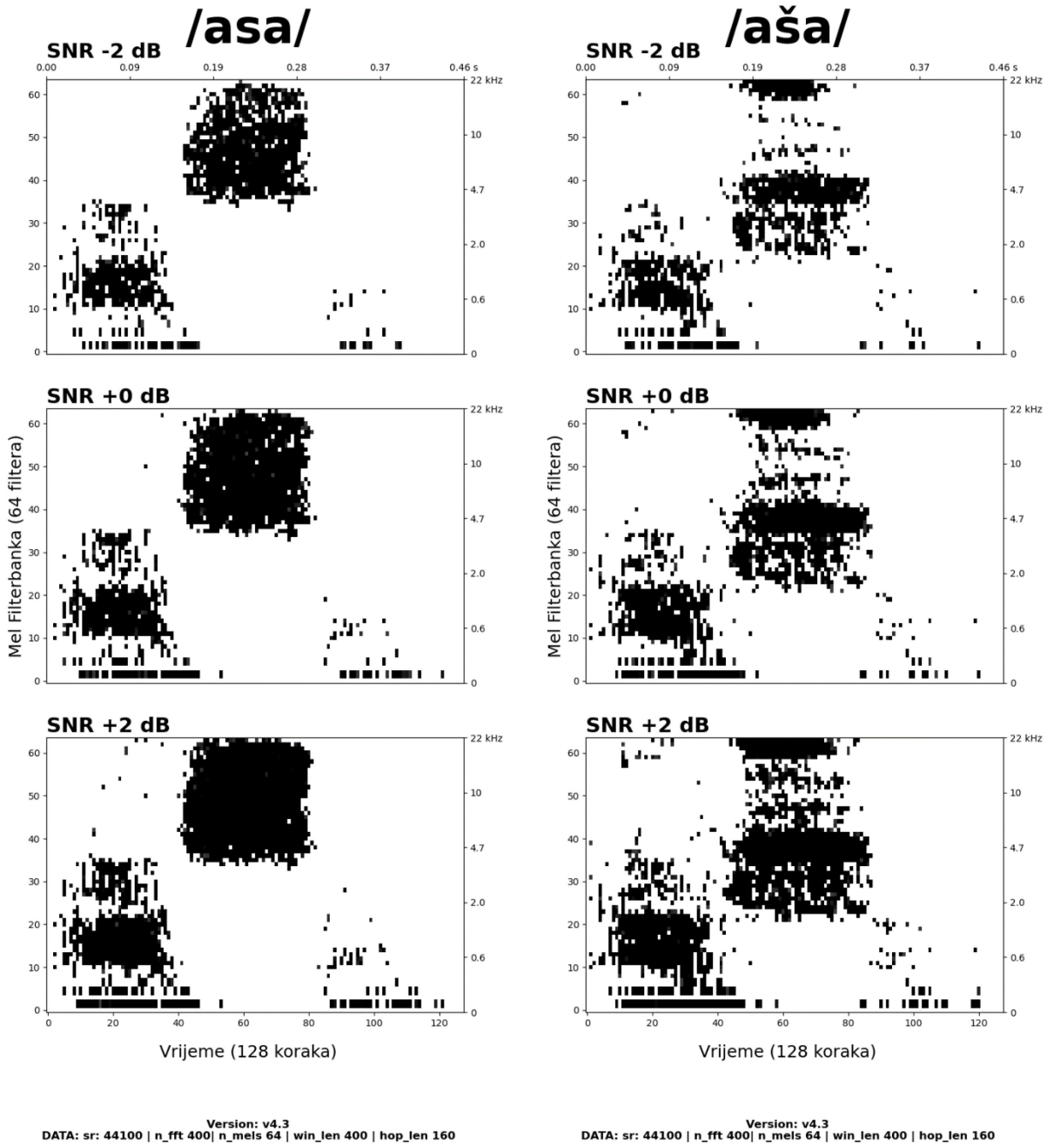
/ata/, /ada/


/ata/
/ada/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
22 kHz
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

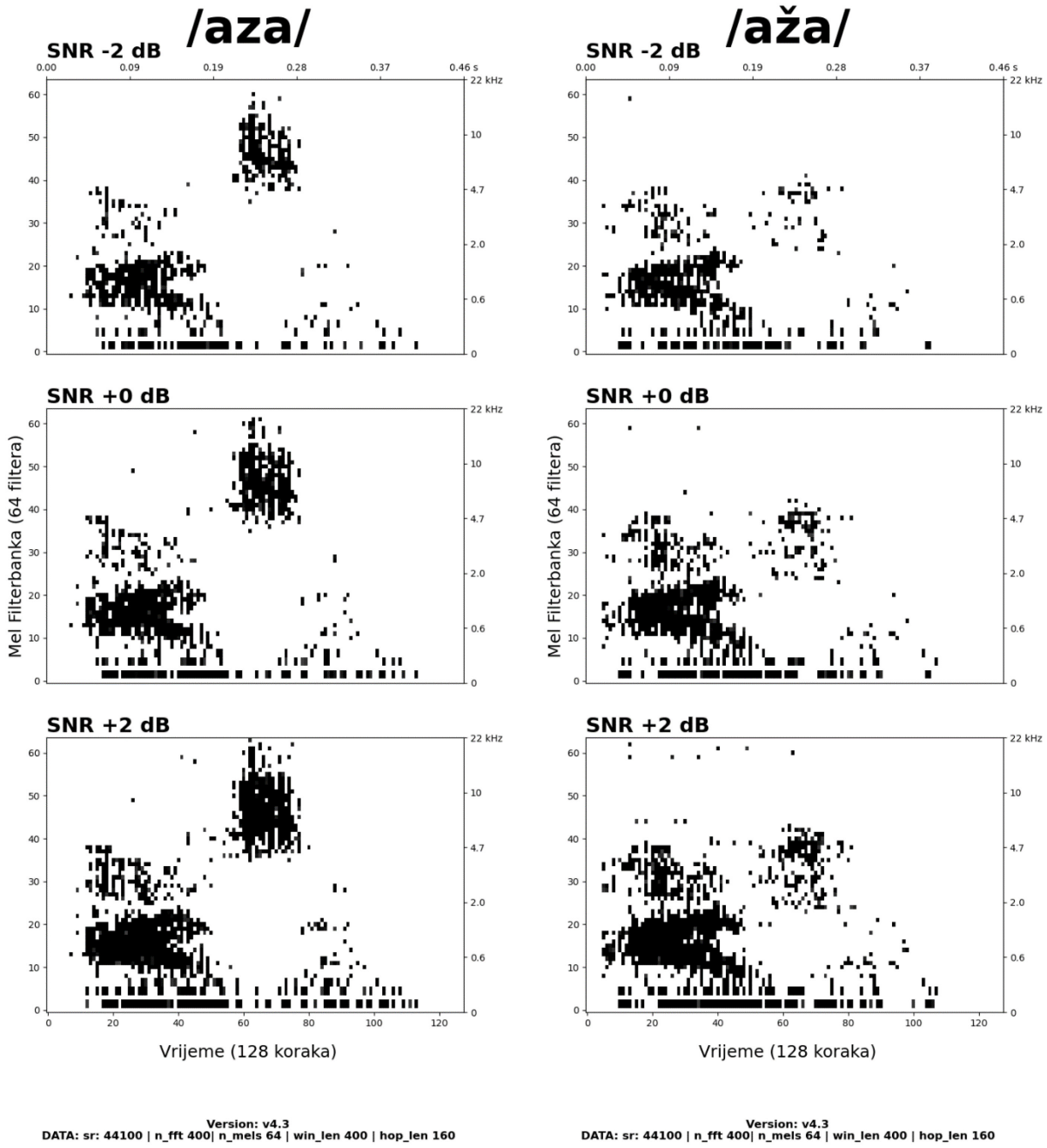
/aka/, /aga/


/aka/
/aga/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

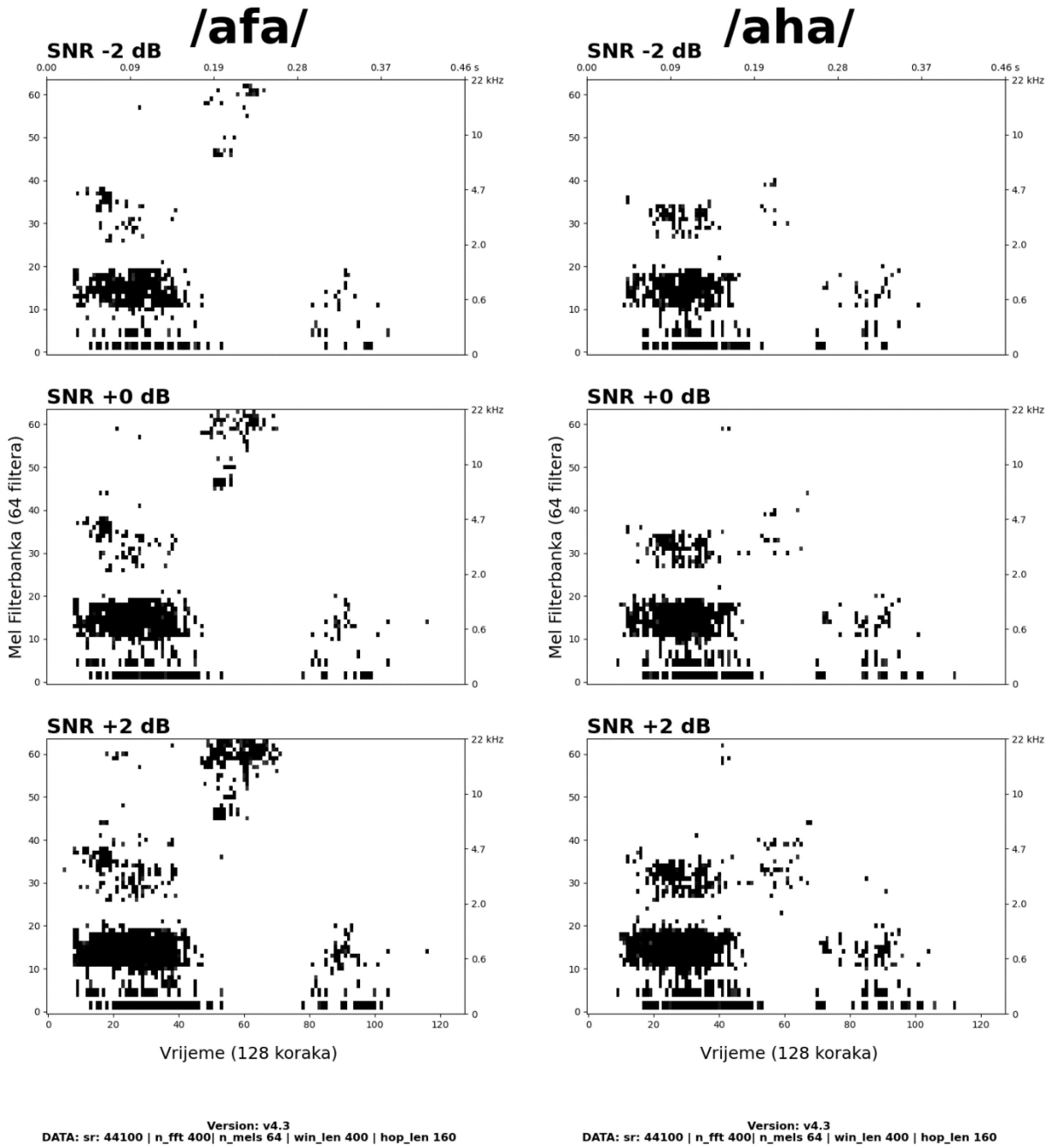
/asa/, /aša/


/asa/
/aša/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

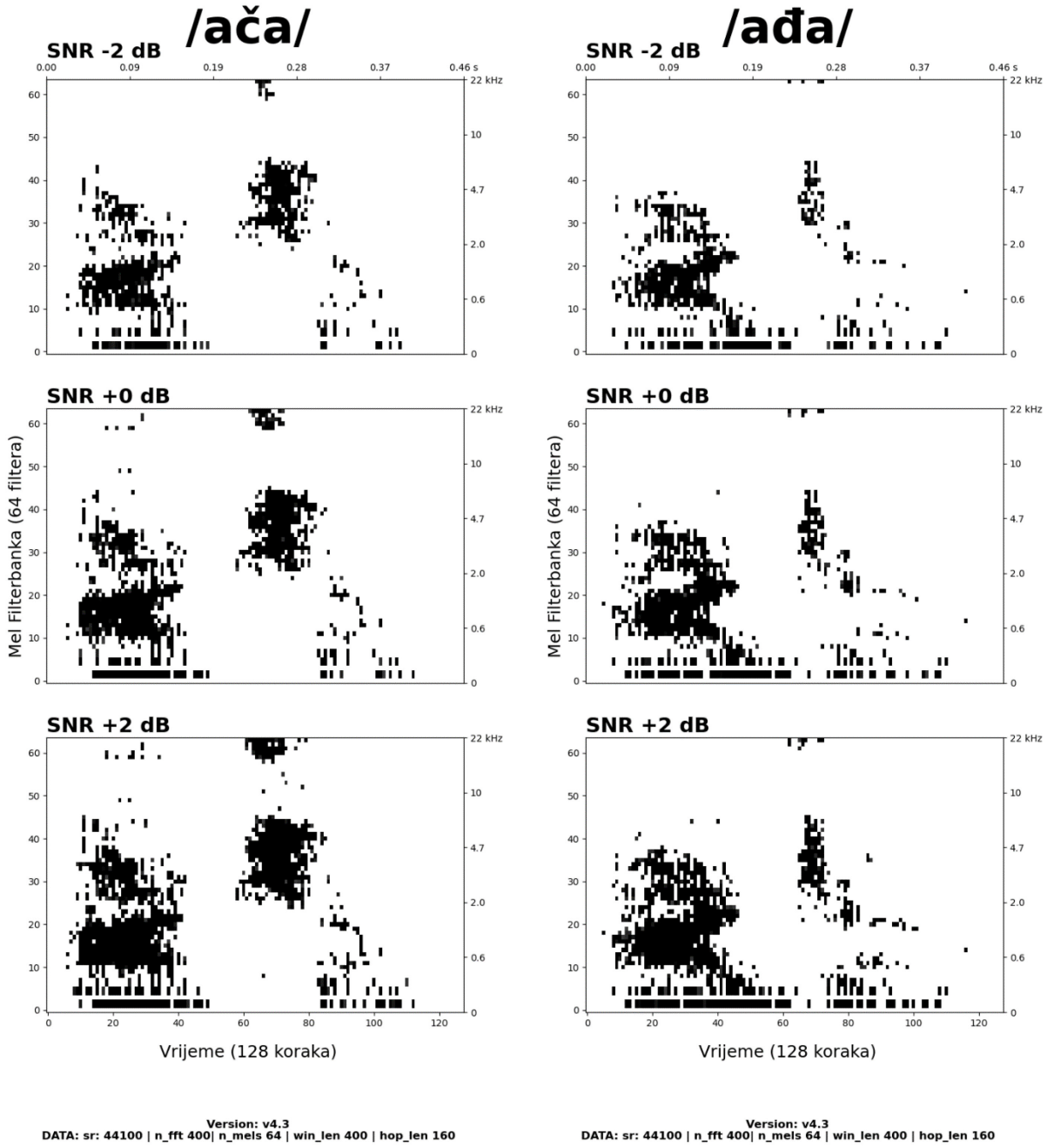
/aza/, /aža/


/aza/
/aža/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

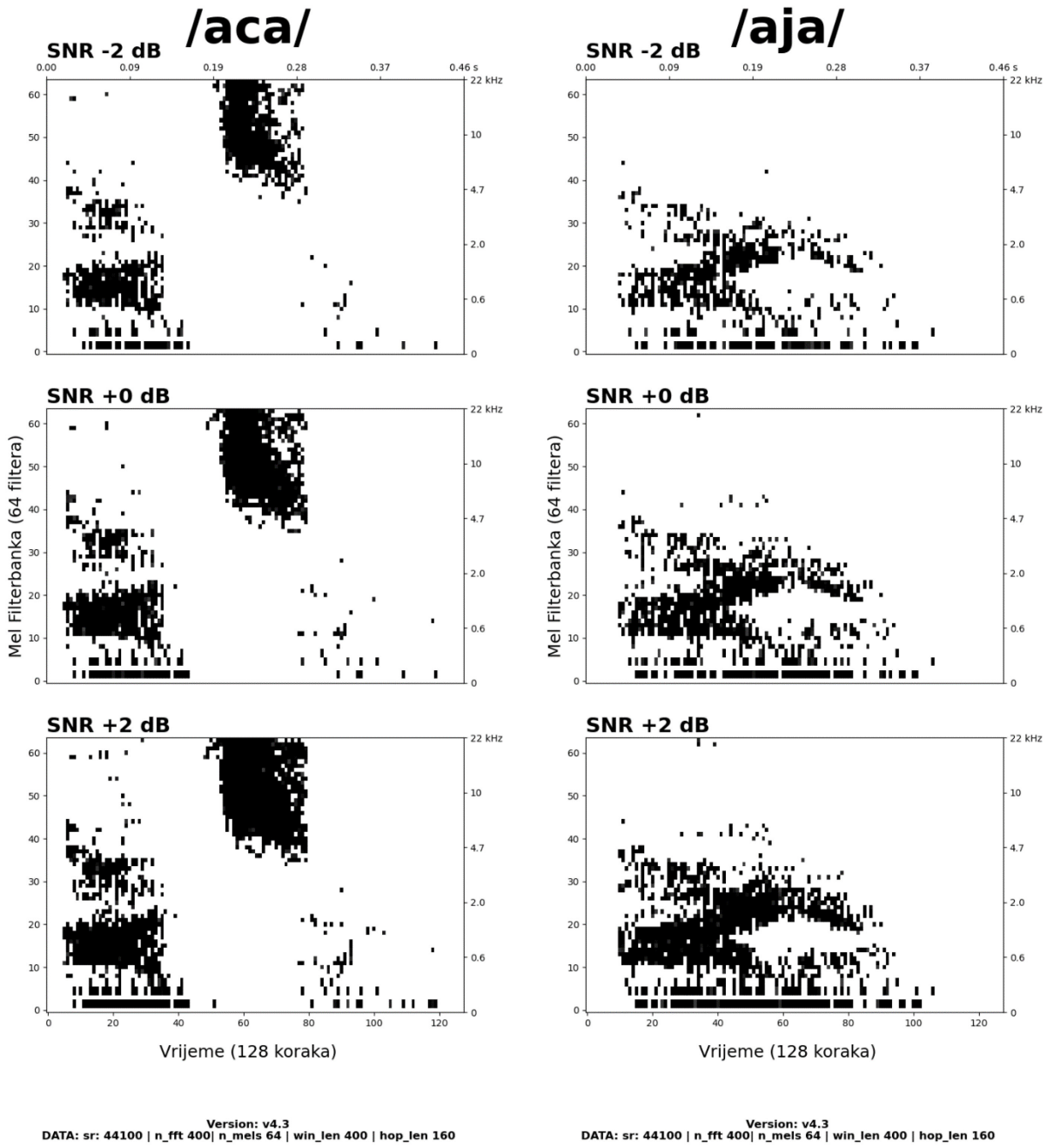
/afa/, /aha/


/afa/
/aha/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

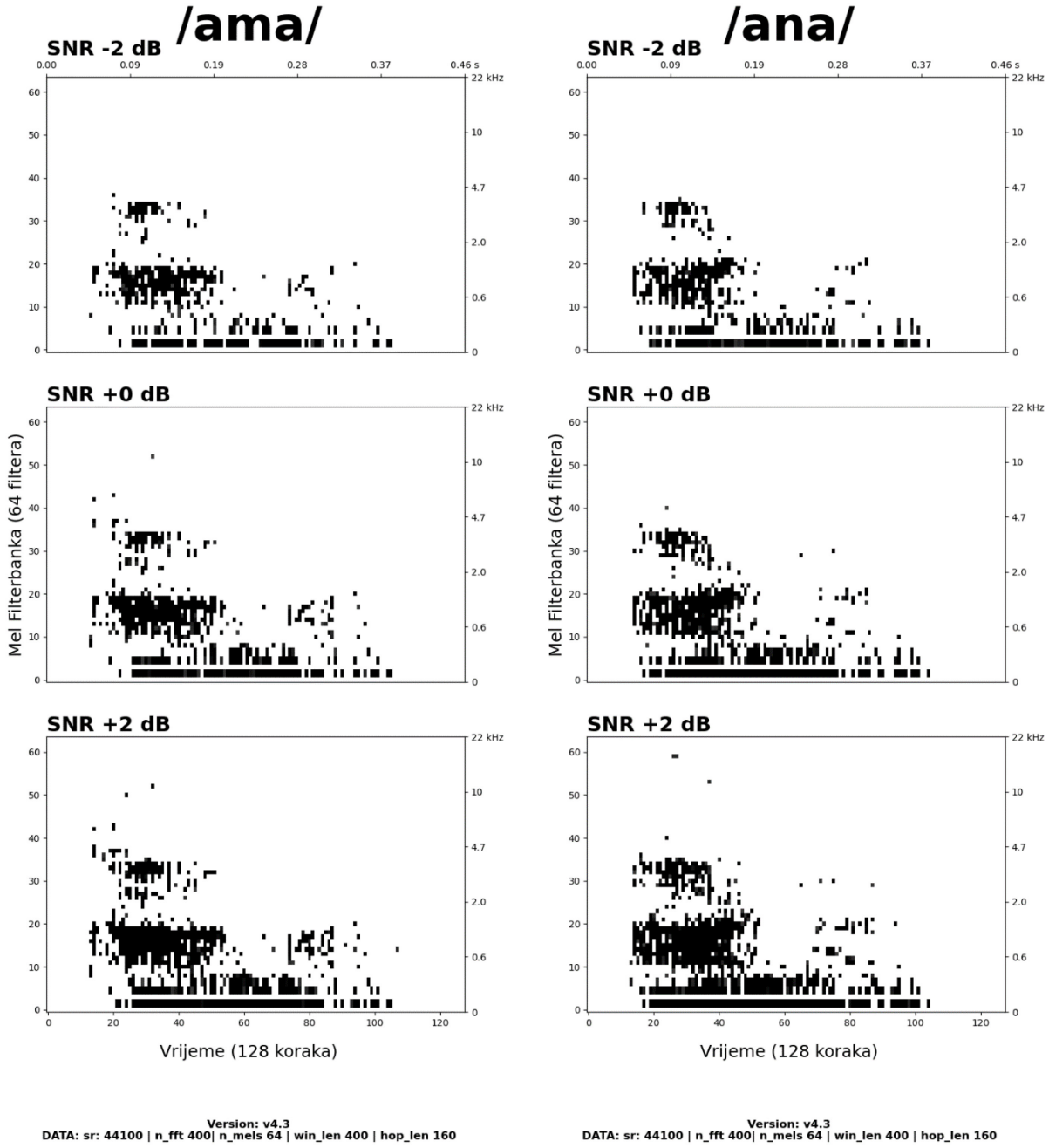
/ača/, /ađa/


/ača/
/ađa/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

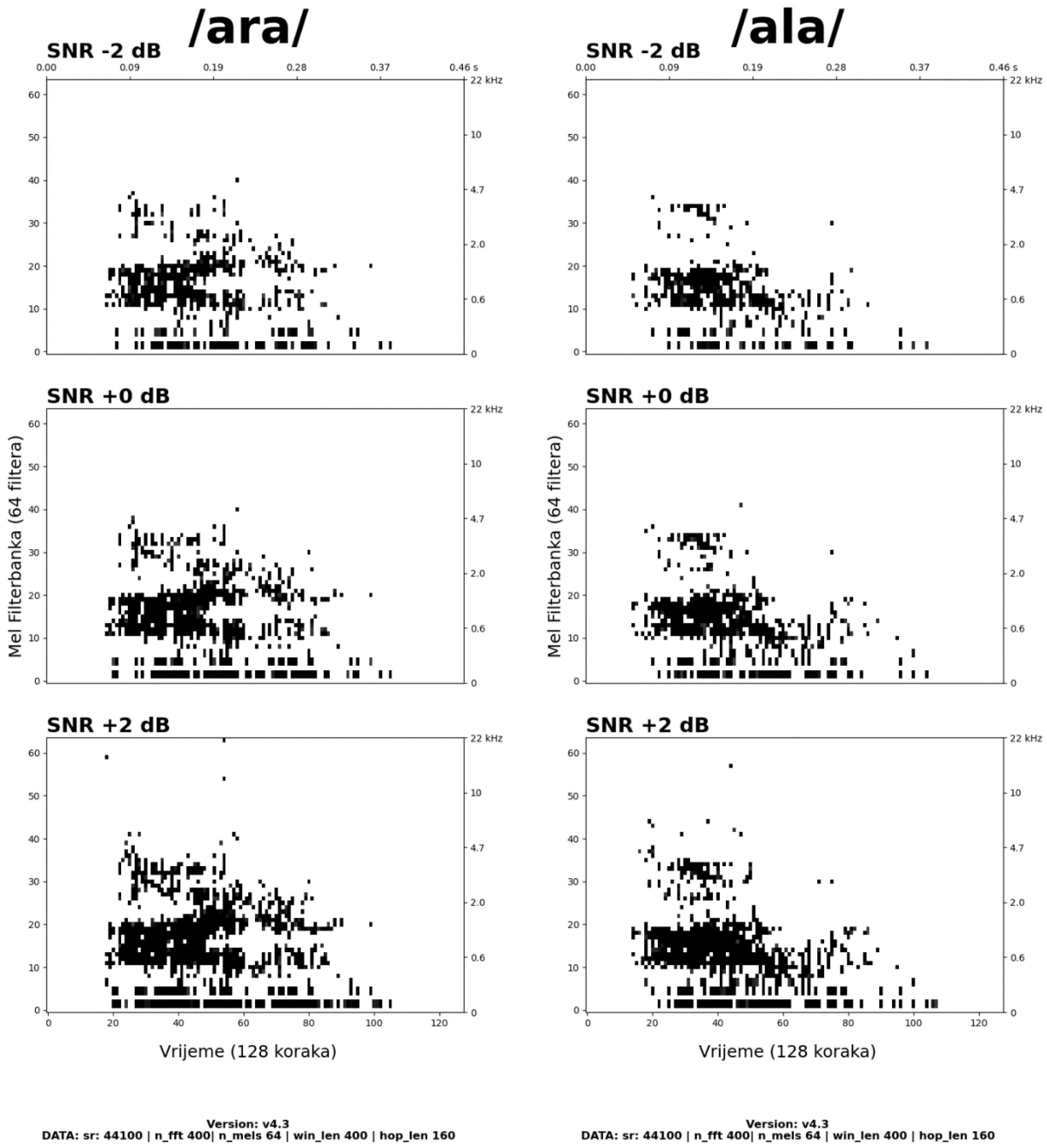
/aca/, /aja/


/aca/
/aja/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

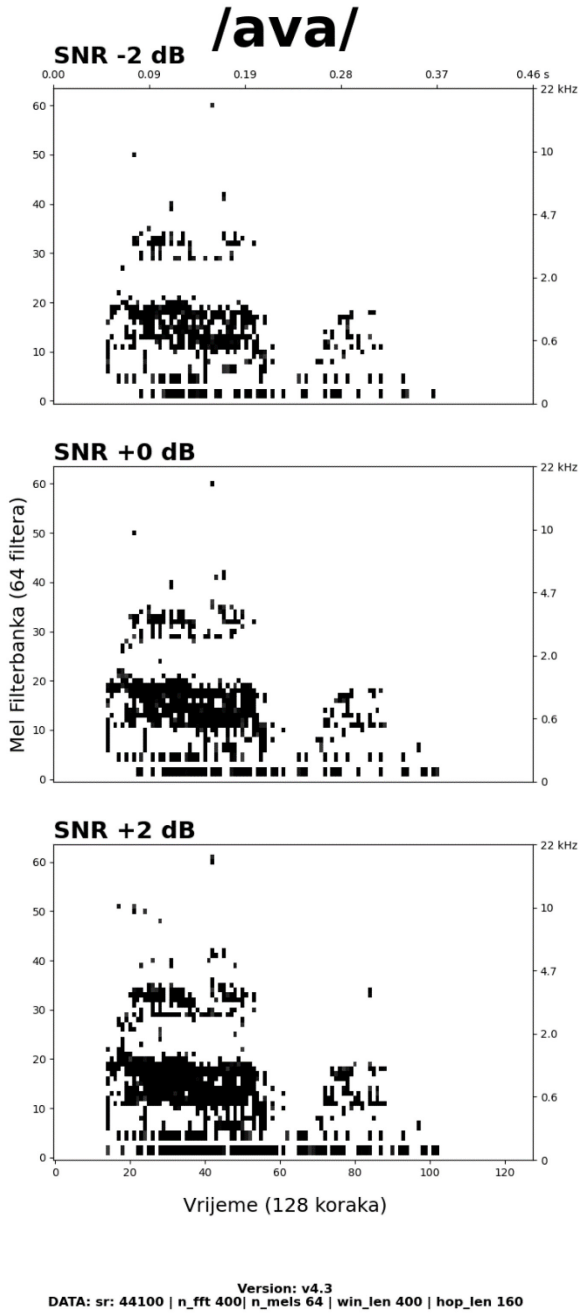
/ama/, /ana/


/ama/
/ana/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
22 kHz
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

/ara/, /ala/


/ara/
/ala/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160

/ava/


/ava/
SNR -2 dB
SNR +0 dB
SNR +2 dB
Mel Filterbanka (64 filtera)
Vrijeme (128 koraka)
Version: v4.3
DATA: sr: 44100 | n_fft 400| n_mels 64 | win_len 400 | hop_len 160