

5. Tradicionalni i suvremeni pristupi u formantskoj analizi

5.1. Definicije formanata

Izvorišni zvuk koji nastaje radom glasnica u grkljanu oblikuje se prolaskom kroz vokalni trakt. Rezonantne šupljine vokalnog trakta²⁴ oblikuju početno stvoren govorni zvuk proizvodeći rezonantne i antirezonantne frekvencije. Na taj se način stvaraju spektralni vrhovi koji su rezultat filtrirajućih efekata vokalnog trakta (Harrison, 2013). Formantska analiza, odnosno analiza formantskih vrijednosti sastavni je dio svake forenzične fonetske analize i jedan je od osnovnih oblika akustičke fonetske analize govora. S obzirom na to da nam pružaju informacije o izgovoru vokala (prednjosti i stražnjosti, otvorenosti i zatvorenosti, (ne)zaokruženosti itd.) te o duljini vokalnog trakta govornika, ne začuđuje da Crystal (2010) formante predstavlja kao jedne od glavnih obilježja ljudskog govora.

Fant (1960: 20) definira formante kao „spektralne vrhove zvučnog spektra“. Autor naglašava da se pojmovi rezonantne frekvencije i spektralni vrhovi trebaju razlikovati, iako se u većini slučajeva mogu koristiti kao sinonimi. Malmberg (1974) za formante navodi da su rezultat pojačanih frekvencija koje uzrokuju akustička obilježja vokalnog trakta, tj. rezonantne karakteristike govornog prolaza. Iako Fry (1979) formante jednoznačno definira kao rezonantne frekvencije vokalnog trakta, napominje da termin formanta češće koristimo za frekvenciju na kojoj se pojavljuje spektralni vrh, s obzirom na to da formanti dovode do vrhova u spektru proizvedenog zvuka, s čime se slaže i Johnson (2012).

Formanti se u literaturi još definiraju kao rezonantne regije ili regije u kojima se koncentrira veća količina energije (Hollien, 1990). Baldwin i French (1990) opisuju ih kao vibracije na višim frekvencijama koje su uzrokovane trenjem zračne struje o površinu vokalnih organa i rezonancijama pojedinih dijelova govornog trakta. Clark i Yallop (1995: 246) upozoravaju na čestu tehničku nepreciznost u defini-

²⁴ Story, Titze i Hoffman (1996) ispitali su različite oblike vokalnog trakta tehnikom MRI. U radu su utvrdili inventar od 18 različitih oblika šupljina kod jednog govornika koje su odražavale izgovor vokala i konsonanata (12 vokala, 3 nazala i 3 okluziva). Rezultati su rada ukazali na opće sličnosti u izgledu rezonantnih šupljina govornika u odnosu na prethodne studije, ali i očite razlike koje upućuju na utjecaje tehnika snimanja/prikaza, metode obrade slika te, naravno, anatomske razlike kod govornika.

ranju formanata kao rezonantnih frekvencija, navodeći da su formanti: „posljedica rezonancije, a ne rezonancija sama.“ Spomenuti autori formante također opisuju kao „vrhove spektralne energije koji su nastali rezonancijom“ (Clark i Yallop, 1995: 246). Sličnu definiciju navodi i Crystal (2010), objašnjavajući da su formanti zapravo vrhovi akustičke energije čije su amplitude u usporedbi s drugim frekvencijama govornog spektra značajno veće. Škarić (1991), baš poput prethodno spomenutih inozemnih autora, formante definira kao pojačane dijelove spektra koji su rezultat rezonantnih frekvencija govornog prolaza. S obzirom na rezonanciju te formante kao posljedicu rezonancije, Harrison (2013) u svojoj disertaciji predlaže precizno razlikovanje rezonantne frekvencije kao akustičke karakteristike vokalnog trakta te spektralnih vrhova kao rezultata prijenosa zvuka.

5.2. Procjena formantskih frekvencija

Brojni autori navode kako se formanti *analiziraju*, međutim, zbog činjenice da različiti istraživači istog znanja i razine iskustva mogu odrediti različitu frekvencijsku vrijednost istog formanta, Harrison (2013: 30) predlaže da se umjesto izraza *mjerjenje* ili *analiza* formantskih frekvencija koristi izraz *procjena* formantskih frekvencija. Kendall i Vaughn (2015) pak govore o *najvjerojatnijoj* formantskoj vrijednosti, kojoj daju prvenstvo nad terminom *točna* formantska vrijednost. Spomenuti autori i brojni drugi stručnjaci upozoravaju da rezultati formantske analize ovise i o samokritičnosti, tj. strogosti istraživača.

U spektrogramskim prikazima formanti se mogu uočiti kao područja veće zvučne energije koja je odraz veće amplitude zvuka. Pojednostavljeno rečeno, rezonancija je selektivno pojačavanje određenih frekvencija u spektru. S obzirom na to da je osnovno obilježje vokala jasna i vrlo stabilna formantska struktura, možemo reći da su na spektrogramima najuočljiviji govorni elementi upravo vokali. Bakran (1996) napominje da ono što vidimo u zvučnom signalu i analiziramo kao formant nije samo produkt rezonantnih frekvencija izgovornog prolaza već rezultat triju sljedećih čimbenika: spektra pobude, karakteristika prijenosa zvuka od usta do mikrofona te rezonantnih pojačanja uvjetovanih prolaskom zraka kroz izgovorne šupljine različitih oblika i građe.

U fonetskoj znanosti formantska se analiza vrlo često provodi u programu za akustičku analizu i obradu zvuka Praat (Boersma i Weenink, 2022: verzija 6.2.14). U današnje vrijeme postoje brojne metode za procjenu formantskih frekvencija, a potreban alat dostupan je i u nekoliko besplatnih softvera za akustičku analizu govora (Bašić, 2018). Unutar Međunarodnog društva za forenzičnu fonetiku te akustiku govora najčešće se koristi Praat (Harrison, 2013), uz koji se istraživači često koriste i WaveSurferom (Sjölander i Beskow, 2017). Harrison (2013) navodi kako su tri najčešće korištene metode u mjerenju formantskih frekvencija: frekvencijski spektar, spektrogrami

te metoda LPC (engl. *linear predictive coding*). Svaka od spomenutih metoda ima svoje prednosti i nedostatke, a uloga je istraživača da s obzirom na cilj rada i snimke kojima raspolaže pronađe adekvatan alat čija će se primjena moći opravdati.

Procjena formantskih frekvencija može biti statička i/ili dinamička, ovisno o tome: a) u kojim se dijelovima formantskih kontura vokala procjenjuju vrijednosti, b) u koliko se točaka procjenjuju i c) na koliko se dugome govornom materijalu procjenjuju. S obzirom na mjesto možemo reći da se vokal dijeli na stabilni i promjenjivi dio. Stabilni dio vokala odnosi se (uglavnom) na središnji dio vokala, u kojem su konture formanata gotovo paralelne te su horizontalno položene. Taj se dio vokala u inozemnoj literaturi često pojavljuje pod terminima *vowel target* i *steady-state*. U akustičkim analizama vokalske kvalitete te u forenzičnim fonetskim radovima procjene formantskih frekvencija uglavnom se provode u stabilnom dijelu vokala (Lee i sur., 1999; Hayward, 2000; Becker, Jessen i Grigoras, 2009; Alotaibi i Hussain, 2010). Hayward (2000) također napominje da je stabilni dio vokala ono što govornik želi postići i ono što slušatelj nastoji identificirati u vokalu, dok su područja tranzijenata važnija za identifikaciju konsonanata.

Promjenjivi dio vokala predstavlja tzv. tranzijentska područja u kojima je vidljiv jasan utjecaj okolnih glasnika. Tako primjerice formantske frekvencije vokala [a] mogu značajno odstupati u riječima [čas], [mast], [pas], [jad], [sat], [rad] itd., ovisno o tome prethodi li mu ili nakon njega slijedi afrikata, nazal, okluziv, poluvokal, frikativ ili pak vibrant. Tranzijentska područja od posebnog su interesa za istraživače u području artikulacijske fonetike, koji nastoje dati odgovore na pitanje koja je temeljna govorna jedinica, koji artikulacijski procesi stoje u pozadini tipičnog, a koji u proizvodnji atipičnog govora i slično. Iako je u području forenzične fonetike fokus vještaka najčešće u stabilnom području formantskih frekvencija, ponekad je uslijed distorzija zvučnog signala ili pak maskiranja govornog signala korisno analizirati i tranzijentsko područje.

Statične metode procjene formanata provode se na kraćemu govornom materijalu (najčešće izoliranim riječima) i u jednoj do tri točke stabilnog dijela vokala. Pritom je najčešća procjena u naglašenom vokalu riječi kako bi se izbjegla česta glotalizacija na završnom slogu u riječi. Za razliku od toga, dugotrajna formantska distribucija (LTFD za engl. *long-term formant distribution*) daje nam brojnije informacije o prosječnom ponašanju, tj. kretanju svakog formanta kod pojedinog govornika na duljim snimkama govora. LTFD-om se nastoje zabilježiti opće (iz)govorne navike pojedinca koje se zadržavaju tijekom cjelokupnog trajanja govora, primjerice labijaliziranje ili palataliziranje (Bašić, 2018). Za potrebe primjene spomenute metode procjene formanata (ali i drugih) nužno je govorne snimke montirati (ukloniti govor ispitivača i/ili sugovornika, zvuk buke: cestovnog prometa, nakašljavanja, smijanja itd.) te potom segmentirati i anotirati vokalske segmente u kojima će biti provedena procjena formanata. Najčešće se za procjenu segmenti-

raju i anotiraju tipični²⁵ ostvaraji pojedinih vokala u naglašenim pozicijama u riječi, dok se atipični²⁶ ostvaraji izostavljaju iz analize i montiraju kao zaseban materijal za potrebe izvještaja sudskog vještaka forenzičnog fonetičara, u kojem će se zasebno navesti primjeri i specifičnosti izgovora vokala (i konsonanata) osumnjičenika/okrivljenika čiji se izgovor analizira.

Iako se dugotrajnim metodama mjerenja formanta dolazi do iscrpnijeg opisa izgovora govornika, prijašnja su istraživanja pokazala da nema statistički značajnih razlika u više mjerenih točaka istog vokala izuzmu li se tranzijentska područja (Tivadar, 2008; Alotaibi i Hussain, 2010; McDougal, 2014). Osim toga, statične su metode u usporedbi s dinamičnim procjenama formanta mnogo dugotrajnije, a jednako učinkovite.

Na temelju dostupne i konzultirane literature u potpunosti je usuglašeno brojanje označavanje formanta, prema kojemu je prvi i najviši vrh prvi formant, drugi niži vrh jest drugi formant itd. (Hayward, 2000; Crystal, 2010; Harrison, 2013). Frekvencijske se vrijednosti formanta skraćeno navode pod velikim slovom F, kojemu je pridružen i pripadajući broj formanta. Tako je primjerice oznaka za prvi formant F₁ ili F₁, za drugi formant F₂ ili F₂ i tako dalje.

U radu predstavljenom na Međunarodnom kongresu fonetskih znanosti (ICPhS) u Pragu 2023. god. Weenink je predstavio unaprijeden način analize i vizualizacije različitih mjerenja formantskih frekvencija. Pri procjeni formanta moguće su pogrešne procjene formanta, koje prije svega ovise o broju formanta koje smo unaprijed zadali u programskim postavkama. Ako primjerice odredimo premali broj formanta s obzirom na gornju graničnu frekvenciju u formantskim postavkama (engl. *formant ceiling*) te u odnosu na onaj broj formanta koji je zaista prisutan u određenome frekvencijskom rasponu, tada će program dva formanta vjerojatno prepoznati kao jedan, dok će u suprotnom slučaju, kada odredimo za analizu prevelik broj formanta u odnosu na formante koji se zaista ostvaruju, program umjesto jednog formanta pokušati (i često uspjeti) prepoznati u govornom signalu i formant koji zapravo nije formant. Upravo je iz spomenutih poteškoća pri procjeni formanta, kao i brojnih drugih, Weenink (2023) istaknuo da je za postizanje preciznije procjene formantskih frekvencija nužno prvo izračunati alternative izvođenjem više analiza formantskih frekvencija istog zvuka s različitim gornjim granicama formanta (engl. *formant ceiling*). Variranjem vrijednosti gornje granice moguće je prilagoditi frekvencijski interval spektra fiksnom broju formantskih vrhova koji se koriste za njegovo modeliranje. Na temelju toga početnog dijela analize prikupljeni se rezultati višestrukih analiza formanta prikupljaju u podatkovnu strukturu koju

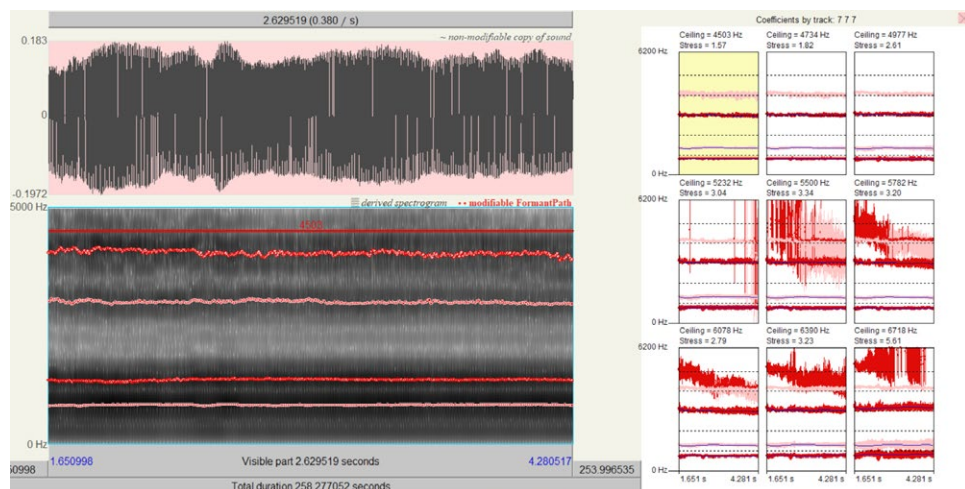
²⁵ Tipični vokali u hrvatskom jeziku u ovom bi kontekstu bili monofonški vokali /i/, /e/, /a/, /o/ i /u/.

²⁶ Atipičan izgovor vokala u kontekstu forenzične fonetike može primjerice biti neutralizacija vokala u finalnim položajima u riječi ili pak otvoren izgovor vokala u dugim naglašenim pozicijama (pri izgovoru naglašenog sloga s dugosilaznim ili pak dugouzlaznim naglaskom).

Weenink (2023) naziva *Formant Path*. Osim spomenutog noviteta u izračunavanju i prikazivanju vrijednosti formanata, u program Praat dodana je i opcija *Formant Path Editor*, koja korisnicima omogućava pregled i uređivanje dostupnih opcija formanata, koje se mogu i nadograditi anotacijskim zapisima.

Weenink (2023) naglašava da se primjenom te metode zapravo umjesto provođenja jedne analize s fiksnim parametrima na zvuku izvode višestruke analize, od kojih svaka ima različite postavke parametara za izračun formanata. Za svaku od provedenih analiza izračunava se i vrijednost *naprezanja* (engl. *stress value*), koja predstavlja ukupnu mjeru ujednačavanja formantskih kontura. Vrijednost naprezanja kvantificira koliko je formantska kontura dobro modelirana. S obzirom na to, što je spomenuta vrijednost manja, to je bolje modeliranje. Niži broj polinoma ukazuje na jednostavnost i dobro opisuje podatke, a što je red viši, to je krivulja fleksibilnija i bolje se može prilagoditi podacima²⁷. Parametar koji se primjenom te metode formantske analize varira jest gornja granična frekvencija za formante (engl. *formant ceiling*), jer nije moguće unaprijed odrediti optimalnu širinu frekventijskog intervala u kojem će biti najbolje prikazani fiksni brojevi formanata.

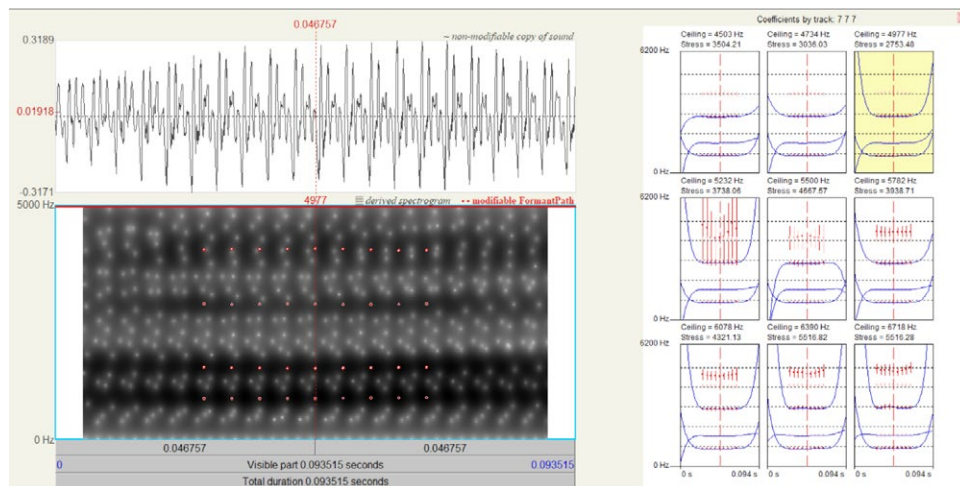
Na slikama koje slijede (slike 15 i 16) prikazane su procjene formantskih vrijednosti na primjeru ženskog glasa korištenjem prethodno spomenute metode *Formant Path* na uzorku fonacije vokala [a] te montiranog vokala [a] iz vezanog govora (iz riječi [navodi]). Na temelju vrijednosti *naprezanja* (engl. *stress value*) vidljivo je da vrijednosti variraju ovisno o gornjim graničnim frekvencijama: od najniže vrijednosti od 1,57 za gornju graničnu frekvenciju od 4.503 Hz, do najveće vrijednosti koja iznosi 5,61 za gornju graničnu frekvenciju od 6.718 Hz.



Slika 15. *Formant Path* – metoda izračuna i prikaza različitih gornjih frekvencijskih vrijednosti u procjeni formanata u fonaciji vokala [a] u ženskom glasu

²⁷ Više o metodi *Formant Path* na: <https://www.fon.hum.uva.nl/praat/manual/Formant-PathEditor.html>.

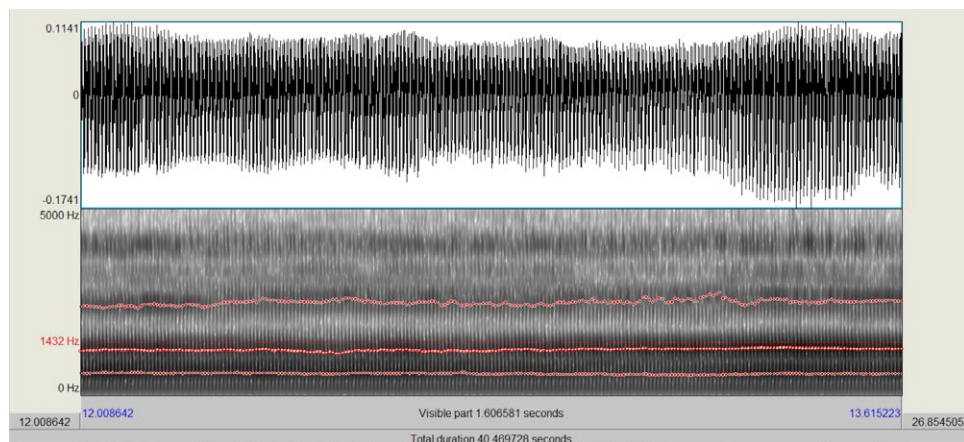
Kao što je i bilo očekivano, vrijednosti su naprezanja značajno veće u uzorcima iz vezanog govora (slika 16) zbog koartikulacijskih utjecaja i kraćeg trajanja vokalskog segmenta. Tako su primjerice u montiranom izgovoru vokala [a] dobivene vrijednosti 2.753,48 (najniža vrijednost) za gornju graničnu frekvenciju od 4.977 Hz i najviša 5.516,82 za vrijednost od 6.390 Hz.



Slika 16. Formant Path – metoda izračuna i prikaza različitih gornjih frekvencijskih vrijednosti u procjeni formanata u vokalu [a] montiranom iz vezanog govora ženske govornice

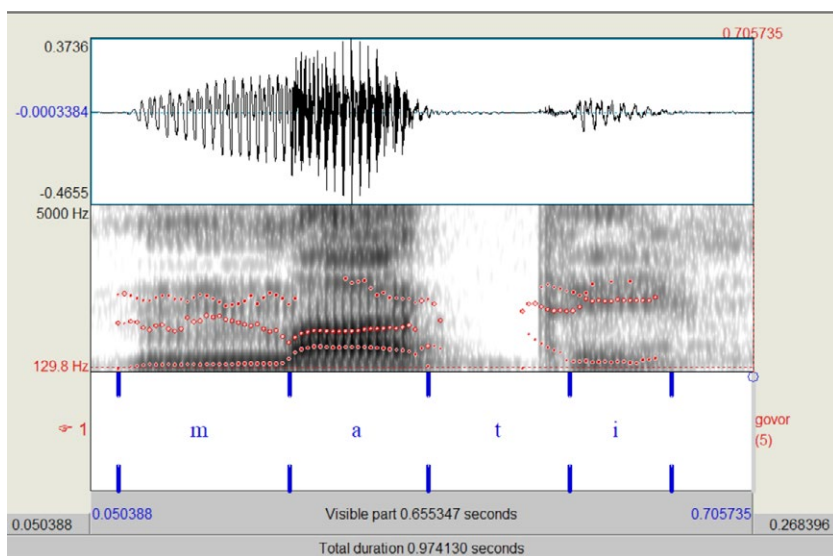
5.1.2. Metodološki izazovi u procjeni formantskih frekvencija

Za osnovno upoznavanje s programom za akustičku obradu zvuka (primjerice Praatom) potreban je manji udio vremena u odnosu na vrijeme koje je potrebno uložiti u upoznavanje svih nedostataka i ograničenja pojedinih analiza koje program nudi. Tako je npr. pri procjeni formantskih vrijednosti nužno nadzirati postavke Praata koje utječu na konačne vrijednosti: određivanje frekvencijskog pojasa unutar kojeg će program procijeniti formantske frekvencije, određivanje broja formanata i sl. Uzmimo za primjer vokal [a] u hrvatskom jeziku. Za muškog je govornika hrvatskog jezika opravdano ograničiti postavke formantske analize na raspon od 0 do 3.000 Hz za prva tri formanta (F1 do F3) ako je izgovor vokala alokalni. Na slici 17 prikazana je procjena formanata pri fonaciji vokala [a] kod muškog govornika hrvatskog jezika.



Slika 17. Procjena formanata u Praatu u fonaciji vokala [a] kod muškog govornika

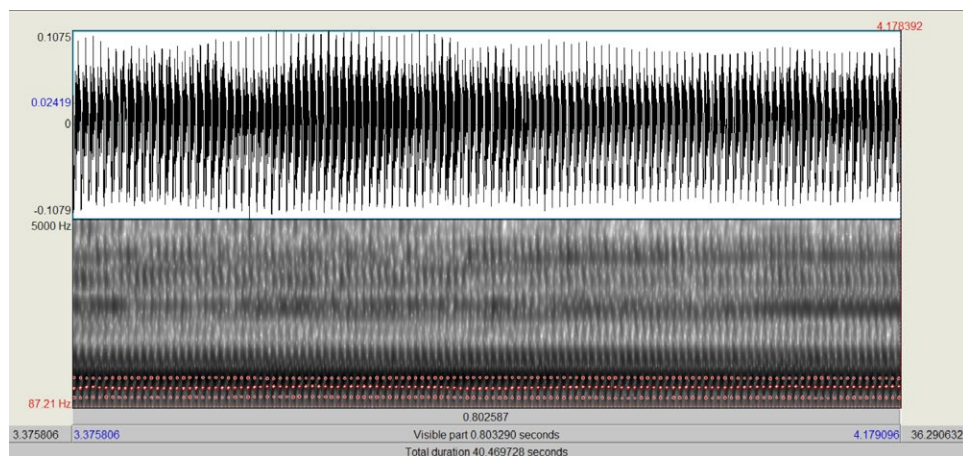
Zahtjevnost procjene formantskih vrijednosti u korelaciji je sa složenošću govornog materijala. Što je on kompleksniji, to je procjena zahtjevnija. Primjerice, pri fonaciji vokala značajno je lakše procijeniti vrijednosti formanata nego pri izgovoru riječi (jer je prisutna koartikulacija okolnih glasnika u odnosu na vokal čiji se formanti procjenjuju), što se može vidjeti na slici 18. Također, što je vokal kraći, to je i kraće trajanje stabilnog dijela vokalskih formantskih kontura, a pri čitanju teksta i/ili spontanom govoru govorni su parametri varijabilniji (formantske frekvencije, vremenske odrednice govora, fundamentalna frekvencija itd.) nego pri fonaciji.



Slika 18. Prikaz formantskih kontura (F1 – F3) kod muškog govornika u riječi [mati] s anotiranim govornim materijalom

U prethodno navedenom primjeru jasno se mogu primijetiti i druge otegotne okolnosti pri procjeni formanata u početnom, a posebice u završnom vokalu riječi. Naime, već se vizualnom inspekcijom može utvrditi da je pri izgovoru prvog vokala u riječi program neuspješno učitao kretanja trećeg formanta u početnom dijelu vokala [a] jer je spektralna energija u tom pojasu bila nešto slabija nego u drugom dijelu. Kod drugog je vokala u riječi [mati] došlo do glotalizacije, što je jedna od idiosinkratičnih osobitosti promatranoga muškog govornika. Naime, u završnim slogovima intonacijskih cjelina i rečenica u govornoj je snimci učestalo ostvarena glotalizacija, što je s jedne strane otežavalo procjenu formantskih frekvencija, a s druge strane onemogućavalo procjenu fundamentalne frekvencije u navedenom odsječku govora.

Kod dijalektalnog izgovora vrijednosti formanata razlikuju se od vrijednosti za standardni ili neutralni izgovor vokala, kao i kod osoba s izgovornim manama. Stoga je u spomenutim govornim materijalima ponekad nužno izmijeniti postavke programa. Broj formanata koji želimo procijeniti u snažnom je odnosu s frekvencijskim rasponom unutar kojeg ih program utvrđuje. Ako programu zadamo da procijeni vrijednost prvih triju formanata u rasponu od 0 do 1.000 Hz za vokal [a] u hrvatskom jeziku, on će to i učiniti, iako bi s obzirom na uobičajene vrijednosti formanata za isti vokal to bilo neopravdano (vidi sliku 19).



Slika 19. Rezultat neopravdanih postavki za očitavanje formanata vokala [a] u Praatu

Višestruka istraživanja provedena na brojnim jezicima pokazala su da je jedan od korisnijih koraka u formantskoj analizi temeljenoj na spektrogramskim prikazima određivanje i odvajanje specifičnih frekvencijskih pojaseva za svaki formant pojedinog vokala. Płonkowski (2015) razrađuje problematiku najbolje razdvojenosti frekvencijskih pojaseva za formante vokala u poljskom jeziku. Autor ističe da se optimalni formantski pojasevi postižu pronalaženjem najveće razlike između mak-

simalnih i minimalnih vrijednosti u istraživanom rasponu formantskih frekvencija. Za svaki je pojas određena prosječna frekvencija, SD te maksimalna i minimalna vrijednost (Płonkowski, 2015). Određivanje frekvencijskih pojaseva pojedinog formanta trebalo bi se temeljiti na studiji provedenoj na većem korpusu, a odvajanjem pojaseva postigla bi se veća točnost u daljnjim formantskim izračunima.

Kao rješenje za razlikovanje formanata čije se frekvencije formanata nalaze vrlo blizu (posebno za F1 i F2) često se predlažu modeli normalizacije vokala. Njihova svrha sastoji se u razlikovanju i odvajanju vokala čije su vrijednosti F1 i F2 vrlo slične, odnosno frekvencijski vrlo blizu jedna drugoj. Međutim, njihova mana leži u presnažnoj normalizaciji, koja rezultira umanjenim razlikama u vrijednostima formanata svakog vokala među različitim govornicima.

Pri formantskim procjenama korisno je voditi se i vizualnom inspekcijom, koja nam daje dodatne informacije o trajanju vokala, njegovim granicama, formantima itd. Međutim, s obzirom na to da je amplituda pojedinih formanata ponekad toliko snažna, dva se formanta mogu sjediniti u jedan te nam u tom slučaju vizualna inspekcija ne daje dobre pomoćne informacije. Formantska odnosno spektralna integracija ponekad je u govornim materijalima uvjetovana frekvencijskom blizinom dvaju susjednih formanata (npr. F1 i F2 kod stražnjeg vokala [u], F2 i F3 kod prednjeg vokala [i] u hrvatskom govoru) ili pak snažnom amplitudom susjednih formanata. Prije određivanja postavki za procjenu formanata nužno je odrediti granične frekvencije s obzirom na spol govornika²⁸, poznavati referentne vrijednosti formanata utvrđene u jeziku koji se analizira te broj formanata koji želimo procijeniti. Broj formanata može nam dati informaciju o tome kakav je izgovor vokala (F1 – F3) ili pak kakva je boja glasa, odnosno timbar govornika. Upravo se stoga često u literaturi govori o nižim, tj. vokalskim formantima (F1 – F3) i višim, vokalnim formantima (Bašić, 2018). Vokalski će nam formanti pomoći pri opisu vokala pojedinog govornika u kontekstu fonološkog sustava određenog jezika, dok će nam vokalni formanti pomoći pri opisu boje glasa govornika.

Osim spomenutih poteškoća pri procjeni formanata, valja napomenuti da je pri analizi ponekad moguća preslaba amplituda formanta, koja će programu onemogućiti očitavanje. Naime, ako je amplituda, odnosno zvučna energija primjerice drugog formanta preslaba, program je neće moći identificirati kao formantsku strukturu. U tom slučaju može se očekivati da će sljedeće izražene pojačanje spektralne energije (treći formant) pogrešno identificirati kao drugi. Upravo je iz tog, ali i svih prethodno navedenih razloga, nužna naknadna provjera procijenjenih vrijednosti formanata, posebice ako je procjena provedena na oskudnome govornom materijalu i malom broju primjera za pojedini vokal.

²⁸ Ženske govornice zbog uobičajeno kraćega vokalnog trakta, manjih rezonantnih šupljina te kraćih, napetijih i tanjih glasnica u odnosu na muškarce imaju više vrijednosti formanata i fundamentalne frekvencije.

O tome da je akustička analiza ženskog i dječjeg glasa zahtjevnija i kompleksnija pisali su brojni autori. Naime, kod žena se i djece zbog visokih frekvencija fundamentalnog tona prvi formant vrlo često ne može razlučiti od nultog formanta. Odnosno, kod osoba s višom fundamentalnom frekvencijom (višom od 150 Hz) teže je procijeniti vrijednosti formanta nego kod osoba s nižim vrijednostima F_0 (Diehl i sur., 1996). Također, neke su studije pokazale da postotak uspješno identificiranih vokala opada s porastom fundamentalne frekvencije (Ryalls i Lieberman, 1982).

Dosad navedeni faktori varijabilnosti formantskih frekvencija u forenzičnoj se fonetici ujedinjuju pod pojmom unutarnje varijabilnosti (*intravarijabilnosti*), odnosno varijabilnosti unutar pojedinog govornika. Razlike i varijabilnost formantskih frekvencija između različitih govornika obuhvaćene su pak pojmom vanjske varijabilnosti (*intervarijabilnosti*), tj. varijabilnosti među govornicima. Tako se primjerice mogu analizirati razlike u vrijednostima formanta kod jednog govornika ili između više različitih govornika, ovisno o postavljenim istraživačkim pitanjima. Varijabilnost je među govornicima uzrokovana i velikim razlikama u duljini, obliku i obujmu vokalnog trakta (posebice među skupinama muškaraca, žena i djece), što još 1952. godine ističu Peterson i Barney. Razlike u vrijednostima formanta kod identičnih blizanaca pokazuju nam da je varijabilnost moguća čak i kod gotovo identične fiziološke podloge (gotovo identičan vokalni trakt) zbog naučenoga idiosinkratičnog artikulacijskog ponašanja (Nolan i Oh, 1996; Loakes, 2006).

5.2.2. Mjera formantske disperzije

U brojnim akustičkim studijama u vezu su se dovodila različita akustička i somatska obilježja govornika. U nekim se istraživanjima razmatrao odnos fundamentalne frekvencije i/ili formantskih frekvencija s duljinom vokalnog trakta, fiziologijom glasnica i grkljanskih prostora, visinom, masom, dobi, spolom govornika itd. Zamjetan interes i odjek u znanstvenim krugovima zabilježen je 1997. godine, kada Fitch uvodi mjeru formantske disperzije (engl. *formant dispersion*), za koju se koristi kratica D_f . Spomenuta mjera raspršenja formanta (D_f) najavljena je u radu u kojem je predstavljeno istraživanje na rezus majmunima ($N = 23$). U studiji je Fitch (1997) ispitao korelaciju tjelesne mase i visine te duljine vokalnog trakta u odnosu na formantske vrijednosti ($F1 - F3$). Sve su akustičke mjere provedene na prethodno snimljenim prijetećim vokalizacijama rezus majmuna.

U radu je navedena i objašnjena formula prema kojoj se disperzija formanta izračunava zbrajanjem prosječnih razlika pojedinih parova formanta. Drugim riječima, zbroji se razlika u frekvenciji $F3$ i $F2$ s razlikom $F2$ i $F1$, što se na kraju podijeli s brojem parova razlika (u ovom slučaju s dva). Autor na kraju rada zaključuje da mjera D_f kod rezus majmuna izravno ovisi o duljini vokalnog trakta i tjelesnoj

veličini te pretpostavlja da vrlo vjerojatno može poslužiti kao robusna mjera kod većine sisavaca (Fitch, 1997).

Formula za izračunavanje formantske disperzije prilagođava se broju mjerenih i u analizi dostupnih formanata, a nije preporučljivo provođenje izračuna na samo dvama formantima. U slučaju da smo u formantskoj analizi mjerili frekvencijske vrijednosti prvih četiriju formanata, koristit ćemo prvu formulu, a u slučaju da smo mjerili prva tri formanta, koristimo drugu.

$$D_f = \frac{(F4-F3)+(F3-F2)+(F2-F1)}{3}$$

$$D_f = \frac{(F3-F2)+(F2-F1)}{2}$$

S obzirom na rezultate prethodno spomenute studije i činjenicu da se unatoč potvrđenoj razlici u vrijednostima formantskih frekvencija kod muškaraca i žena vrlo malo zna o odnosu formanata i tjelesne mase i/ili duljine vokalnoga trakta kod obaju spolova, u nadolazećem su razdoblju uslijedila brojna istraživanja u kojima se ispitivao odnos mjere D_f i duljine vokalnoga trakta kod ljudi. U nekim se studijama nastojalo utvrditi postoji li korelacija između mjere D_f i tjelesne veličine govornika (Puts, Hodges, Cárdenas, i Gaulin, 2007), dok su primjerice Knowles i Little (2016) ispitali utječu li različiti akustički faktori (među kojima i mjera D_f) na percepciju govornikove kooperativnosti. Drugi su pak autori ispitali korelaciju frekvencijskih vrijednosti formanata i tjelesne visine te težine govornika (Fitch, 1994; Fitch i Giedd, 1999). Odnos formantskih frekvencija i različitih somatskih obilježja kod ljudi bit će detaljnije obrađen u sljedećem poglavlju.

5.3. Odnosi formanata i artikulacijskih postavljanja

S obzirom na to da se produciraju velikom količinom energije, neki autori navode da su upravo vokali *ti* koji nose govor (Hollien, 1990). Iako su vokali natkategorija kojom su obuhvaćeni monofonzi, diftonzi, triftonzi, nazalni vokali i drugi, u forenzično-fonetskim radovima uglavnom se analiziraju monofonški oralni vokali. Monofonzi se po definiciji ostvaruju zadržavanjem izgovornih organa u određenom položaju (u istom obliku i istoj veličini) tijekom faze držanja (Horga i Liker, 2016), a ta se nepromijenjenost artikulacijskog prolaza tijekom izgovora na spektrogramu odražava relativno stabilnim formantskim konturama.

Istraživači Laboratorija Haskins u SAD-u 60-ih godina prošlog stoljeća utvrdili su da je u procesu prepoznavanja vokala moguće odrediti glavnu ulogu prvih dvaju formanata ($F1$ i $F2$) (Crystal, 2010). Metodom govornog sintetizatora, koja se

zasniva na ideji da elektronički uređaj kombiniranjem različitih frekvencija, intenziteta i vremena generira zvuk, potvrđena je važnost vokalskih formanata i važnost VUG-a²⁹ pri razlikovanju zvučnih i bezzvučnih konsonanata.

Brojni autori smatraju da najbolji način za opisivanje vokala nije artikulacijski opis, već opis njihovih akustičkih karakteristika (Hollien, 1990; Hayward, 2000; Titze, 2000; Ladefoged, 2003; Johnson, 2012). Artikulacijski opisi podliježu individualnim razlikama prilikom izgovora vokala te nekonzistentnosti između parova parametara: jezične (odnosno lingvističke) visine i prednjosti te izmjerene visine jezika i prednjosti tijekom produkcije samog vokala, na što se upozorava i u forenzičnoj fonetici (Varošanec-Škarić, 2019). Vokale, za razliku od konsonanata, koje je artikulacijski po mjestu tvorbe jednostavnije opisati, najčešće opisujemo prema sljedećim trima obilježjima:

- mjestu artikulacije (mjesto najvećeg suženja u vokalnom traktu)
- visini jezika (količina suženja vokalnog trakta) i
- zaokruženosti ili raširenosti usana (Hollien, 1990).

Osim po kriterijima prednjosti/stražnjosti, otvorenosti/zatvorenosti (odnosno visini jezika) te stupnju zaokruženosti (širenju usana) (Hollien, 1990; Hayward, 2000; Johnson, 2012), vokale možemo opisivati i prema sporednim vokalskim kriterijima koje dodaje Hayward (2000): stupnju nazalnosti i rotičnosti.

Akustički korelati vokalskih obilježja (prednjost i otvorenost) koristili su se stoljećima, a postali su očiti s pojavom zvučnog spektrografa (Joos, 1948; prema Johnson, 2012). Većina će se stručnjaka složiti da su najznačajnije akustičke karakteristike vokala njihovi formanti, čije se vrijednosti danas mogu relativno jednostavno procijeniti i opisati na temelju spektrogramskih prikaza. Iz tih je prikaza govorni zvuk, pa time i vokale, moguće opisati i analizirati prema trima temeljnim odrednicama: vremenu, frekvenciji i intenzitetu.

Autori koji se pretežno bave artikulacijskim aspektima ljudskog govora naglašavaju da odnos formanata i pomaka jezika nije toliko jednoznačan kao što se predstavlja u literaturi (Kent i Read, 2002; Harrington, 2013) te da je daljnjim istraživanjima potrebno precizirati odnos između artikulacijskih gesta i akustičkog signala (Horga i Liker, 2016). Rezultati nekih istraživanja pokazali su da je u slučajevima u kojima nije pronađena korelacija između F1 i vertikalnih pomaka jezika te između F2 i horizontalnih pomaka jezika utjecaj koartikulacije bio izrazit te da je došlo do uspostavljanja motoričke ekvivalencije u kojoj su bile uključene i druge strukture osim jezika (Dromey, Jang i Hollis, 2013).

Druga skupina autora napominje da se pomaci u formantskim frekvencijama te vokalski trapezi kod različitih govornika ili skupina govornika trebaju tumačiti u

²⁹ VUG je kratica za vrijeme uključivanja glasa, dok je engleska inačica VOT, odnosno *voice onset time*.

skupini te analizirati na razini referentnih vrijednosti za pojedini jezik, a ne pojedinačno (Brown i Watt, 2017). Odnosno, ističe se da nam pojedinačne vrijednosti formanata ne govore mnogo ni o jeziku kojim netko govori ni o govorniku samome ako se ne stave u širi kontekst na temelju kojeg se mogu tumačiti rezultati provedene studije.

Harrington (2013) navodi kako se odnos formanata i različitih oblika vokalnog trakta može opisati na temelju četiriju općih načela:

- svi dijelovi vokalnog trakta utječu na formantske frekvencije i formanti su ovisni o cjelokupnom obliku čitavoga govornog sustava
- najviše vrijednosti F1 postižu se pri izgovoru vokala kod kojeg je glavno suženje netom iznad larinksa, a usna je šupljina širom otvorena (povećano suženje u usnoj šupljini dovodi do sniženja F1)
- najviše vrijednosti F2 postižu se pri izgovoru vokala kod kojeg je suženje nastalo u prednjijem dijelu u prostoru između nepca i jezika (suženje u što prednjijem dijelu dovodi do povišenih vrijednosti F3 i F4 zbog skraćivanja izgovornog prolaza, dok se vrijednosti F2, F3 i F4 snižavaju što je mjesto suženja više pomaknuto prema straga; najniže su vrijednosti F2 kad je sužen gornji dio farinksa) i
- svako produžavanje vokalnog trakta dovodi do sniženja vrijednosti formanata, najznačajnije vrijednosti F3 nalaze se kod prednjih vokala, a F2 kod stražnjih vokala.

5.3.1. *Prvi formant*

Većina autora smatra da se vokali mogu vrlo precizno opisati pomoću vrijednosti prvih triju formanata: F1, F2 i F3 (Hollien, 1990; Hayward, 2000; Ladefoged i Ferrari Disner, 2012; Varošanec-Škarić, 2019). Opće je poznato da vrijednosti prvog formanta povezujemo s obilježjima otvorenosti/zatvorenosti, a na njih utječu:

- mandibularni pokreti (otvaranje i zatvaranje čeljusti)
- položaj jezika (položaj jezika prema prednjem ili stražnjem dijelu usne šupljine) te
- faringalno stezanje (sužavanje ili širenje farinksa).

Niže vrijednosti prvog formanta javljaju se kod zatvorenijih vokala, prednji je položenog jezika te kod širenja farinksa (Lindblom i Sundberg, 1971; Laver, 1980; Hollien, 1990; Hayward, 2000; Ladefoged, 2003; Erickson, 2004; Ladefoged i Ferrari Disner, 2012; Klug, 2014). Spuštanjem larinksa te protruzijom i zaokruživanjem usana dolazi do snižavanja vrijednosti svih formanata, pa tako i F1 (Klug, 2014), zbog toga što dolazi do produljivanja vokalnog trakta za prostor isturenih i zaokruženih usana. Moosmüller (2007) je na korpusu od 750 vokala proizvedenih

modalnim i škripavim glasom utvrdila da škripavi glas podrazumijeva dodatnu modifikaciju vokalnog trakta. Naime, u skupini ženskih govornika Moosmüller (2007) je potvrdila da je pri škripavoj fonaciji vokala niža vrijednost F2 u usporedbi s istim samoglasnicima kod istih govornica, ali proizvedenih modalnim glasom, dok kod muških govornika nisu utvrđene dosljedne razlike u vrijednostima formanta kod vokala proizvedenih modalnim i škripavim glasom.

Osim u govoru, vrijednosti formanta u pojedinim su radovima analizirane i u pjevanju. Naime, praćenjem kretanja vrijednosti formanta u pjevanju pokazalo se da povišena pozicija larinksa dovodi do povišenja frekvencija F1 kod otvorenih vokala (Sundberg i Nordström, 1976). U drugoj provedenoj studiji (Chan i Nolan, 2016) ispitan je odnos tonske F_0 i vokalskih formanta u kantonskom i mandarin-skom kineskom. Rezultati su pokazali da je korelacija između ispitivanih parametara vrlo mala ili je nema. Drugim riječima, autori su zaključili da su skraćivanjem (ili produžavanjem) vokalnog trakta (ponajviše faringalne šupljine) zbog povišene (ili snižene) pozicije larinksa frekvencijske vrijednosti formanta ostale nepromijenjene ili su se minimalno promijenile, što upućuje na to da su konture F_0 i formantske frekvencije u velikoj mjeri neovisne te da ton i vokali podliježu različitom koartikulacijskom utjecaju (Chan i Nolan, 2016). U studiji u kojoj su analizirani koartikulacijski utjecaji na različite formante vokala (Paillereau, 2016) rezultati su pokazali da je koartikulacija imala najslabiji učinak na prvom formantu, odnosno F1 se pokazao kao najotporniji i najstabilniji formant.

5.3.2. Drugi formant

S obzirom na to da na kretanje drugog formanta utječe izgovorno obilježje prednosti, odnosno stražnjosti, F2 je viših vrijednosti što je vokal predniji, a nižih što je mjesto tvorbe vokala pomaknutije prema straga (Gordon i Heath, 1998; Erickson, 2004; Varošaneć-Škarić, 2005; Johnson, 2012). Niže će vrijednosti drugi formant imati i kod protruzije i zaokruživanja usana (Lindblom i Sundberg, 1971; Alotaibi i Hussain, 2010), dok će se širenjem usana vrijednosti povesiti. Iako spuštanje larinksa u govoru i pjevanju dovodi do snižavanja vrijednosti svih formanta (Sundberg i Nordström, 1976), Klug (2014) naglašava da je F2 najosjetljiviji od svih formanta. U literaturi se vrlo često prednost/stražnjost vokala procjenjuje isključivo na temelju vrijednosti drugog formanta, no Stevens (1998) upozorava da na stupanj prednosti/stražnjosti određenog vokala utječe i relativan odnos drugog i trećeg formanta.

Osim odnosa F2 i F3, u studijama se razmatrao i odnos prvih dvaju formanta. S obzirom na to da su prosječne vrijednosti formanta uobičajeno više kod žena nego kod muškaraca, očekivano je i veća razlika između vrijednosti procijenjenih kod F1 i F2 u skupini ženskih govornika. Gordon i Heath (1998) tu su razliku

analizirali u engleskom jeziku, a rezultati studije ukazali su na spolne razlike kod svih vokala. Naime, kod prednjih je i središnjih vokala utvrđena značajna razlika između F1 i F2 u skupinama muškaraca i žena, osim kod stražnjih vokala, gdje je razlika zanemariva.

5.3.3. Treći formant

Opće je poznato da na frekvencijske vrijednosti trećeg formanta utječe larinjalna napetost i sam položaj larinksa. Što je veća napetost i viši položaj larinksa, to su njegove vrijednosti više, upravo zato što se skraćuje duljina vokalnog trakta i glasnice postaju napetije. Izražen utjecaj na F3 imaju i protruzija te zaokruženost usana, prilikom kojih dolazi do snižavanja vrijednosti trećeg formanta (Lindblom i Sundberg, 1971; Ladefoged, 2003; Gold, French, i Harrison, 2013; Klug, 2014). Iako su istraživanja pokazala da F3 istovremeno nosi dvije vrste informacija: one o vokalskoj kvaliteti i one o timbralnim karakteristikama samog govornika, neki autori F3 svrstavaju u više formante (uz F4 i F5). Viši se formanti uglavnom tumače s obzirom na timbar, fiziološke karakteristike govornika i sl. (Barreda, 2016). Neki autori smatraju da su za opis vokala engleskog jezika nužne samo vrijednosti prvih dvaju formanta, dok je za distinkciju rotičnog vokala, obojenog glasnikom [r], te prednjih vokala u analizu potrebno uključiti i vrijednosti F3 (Lindau, 1978; Ladefoged, 2003; 2013).

U forenzičnoj fonetici prilikom procesa prepoznavanja govornika najveću diskriminatornu ulogu ima treći formant jer nam daje najviše informacija o govorniku (Gold i sur., 2013). Može se pretpostaviti da se F3 istaknuo kao presudan u usporedbi s nižim formantima (F1 i F2) jer su upravo prva tri formanta najčešće akustički dostupna za analizu u dokaznome govornom materijalu.

Osim što je (u studiji provedenoj na muškim i ženskim glasovima) potvrđena tendencija povišenja frekvencijskih vrijednosti formanta vokala izgovorenih šaptom u odnosu na fonirane vokale, iznjedrena je i nova uloga trećeg formanta (Halberstam i Raphael, 2004). Naime, rezultati su pokazali da je F3 važniji za prepoznavanje vokala izgovorenih u šaptu nego za prepoznavanje foniranih vokala. Na temelju navedenog autori zaključuju da je F3 ključan u govornim situacijama u kojima nemamo informaciju o fundamentalnoj frekvenciji, o čemu su ranije pisali Nusbaum i Morin (1992).

5.4. Referentne vrijednosti formanta za hrvatski jezik

Formantske se vrijednosti, kao i neki drugi akustički parametri, mogu predstaviti kroz numeričke i/ili grafičke prikaze u obliku *plot* dijagrama. Izolirane vrijednosti formantskih frekvencija ne govore nam mnogo te isključivo na temelju njih ne

možemo donositi neke zaključke o izgovoru kod pojedinog govornika. Stoga ih je uobičajeno tumačiti u odnosu na referentne vrijednosti za određeni jezik (npr. standardni/neutralni/alokalni izgovor) ili pak u odnosu na određeni vremenski odmak (npr. prije i nakon korekcije izgovora vokala). Također, formantske se frekvencije mogu tumačiti u odnosu na različite uvjete snimanja (u prostoriji sa sniženom razinom buke, u studijskim uvjetima, u različitim kanalima za transmisiju zvuka – telefonski prijenos različitih mreža (3G, 4G, 5G), od kojih mreže novijih generacija imaju veći raspon frekvencija za prijenos informacija itd.), pri različitim govornim stilovima (čitanje teksta, fonacija, slobodni govor u formi intervjua) i sl. Formantskom se analizom možemo koristiti pri ispitivanju spolnih, dobnihi ili regionalnih razlika među govornicima.

U dosadašnjim se fonetskim istraživanjima hrvatskog govora opisao hrvatski alokalni, standardni, opći hrvatski izgovor kod muškaraca i žena, uglavnom na izolirano izgovorenim riječima, te su rezultati u pravilu iskazani u numeričkoj formi (Bakran, 1990; Bakran i Stamenković, 1990; Škarić, 1991; Bakran, 1996; Varošaneć-Škarić, 2005; Pletikos-Olof, 2008; Bašić, 2018), a sve češće u novijim radovima preteže prikaz rezultata formantske analize u numeričkim vrijednostima, ali i na grafičkim prikazima. *Plot* dijagrami kojima se uobičajeno u odnos stavljaju vrijednosti prvog i drugog formanta (za jedan ili više: vokala, jezika, govornika) mogu biti pomoćna metoda prikazivanja rezultata kojom fonetski opis izgovora vokala postaje holistički jer pruža sveobuhvatnu predodžbu izgovora vokala, međuodnosa formanta te cjelokupnoga vokalskog sustava govornika/jezika (Varošaneć-Škarić, 2005; Bašić, 2018; Varošaneć-Škarić i Bašić, 2015; Biočina i Bašić, 2021).

Za potrebe uspostavljanja referentnoga formantskog okvira hrvatskog jezika za alokalni opći izgovor muških i ženskih govornika hrvatskog jezika u doktorskoj disertaciji Bašić (2018) snimljeno je 100 govornika. S obzirom na to da je izgovor pojedinih govornika naknadnom verifikacijom procijenjen kao lokalni ili atipični u nekom od aspekata govorno-jezične proizvodnje, konačni je korpus za analizu činio 81 govornik ($N = 81$; $\bar{Z} = 35$ i $M = 46$)³⁰. Za procjenu je formanta korištena

³⁰ Govornici su snimljeni u Studiju za akustička snimanja na Filozofskom fakultetu u Zagrebu, na Odsjeku za fonetiku, uz pomoć studijskog tehničara Jordana Bičanića te u prostorijama sa sniženom razinom buke, kvalitetnom opremom za snimanje (u razdoblju od 20. travnja 2015. god. do 13. veljače 2017. god.).

U studiju je korišten visoko kvalitetan mikrofoni AKG (model C 414 B/ULS: Large Diaphragm Condenser Microphone), frekvencijskog raspona od 20 do 20.000 Hz. U prostorijama snižene razine buke korišteni su visokokvalitetni snimači Marantz (model PMD 660) i ZOOM (model H4n) te mikrofoni Røde (model NT5 – Compact Cardioid Condenser Microphone) i AKG (model AKGC 391). Prilikom izbora prostorije za snimanje vodilo se računa da se uklone mogući izvori buke (klimatizacijski uređaji, sva računalna oprema, mobilni aparati, telefoni, prolaznici, cestovna buka, stiskanje ke-

Praatova metoda *formant tracking*, koja nam daje najtočniju procjenu formantskih frekvencija (Kendall i Vaughn, 2015). U disertaciji je kroz kontrastivnu analizu predstavljen referentni okvir formantskih frekvencija u hrvatskom jeziku³¹ kod muških i ženskih govornika na dotad najvećem broju govornika za hrvatski jezik. U tablicama 5 i 6 prikazane su referentne vrijednosti za hrvatski jezik kod govornika obaju spolova čiji je izgovor verificiran kao alokalni (neutralni).

Tablica 5. Referentne vrijednosti za hrvatski jezik na temelju 35 ženskih govornika (preuzeto iz Bašić, 2018)

	Hz	F1	F2	F3
[a]	M	816	1 370	2 521
	C	817	1 370	2 579
	Min.	671	1 217	2 061
	Maks.	929	1 719	2 880
	SD	17,9	14,3	41,4
[e]	M	578	2 107	2 706
	C	572	2 097	2 704
	Min.	475	1 905	2 310
	Maks.	756	2 446	3 027
	SD	11,9	23,6	48,2
[i]	M	377	2 399	2 875
	C	372	2 393	2 887
	Min.	296	2 144	2 548
	Maks.	505	2 607	3 054
	SD	10,5	29	63,2

mijske olovke, šuškanje papirima itd.). Za metodološke pojedinosti vidjeti doktorsku disertaciju Bašić (2018) pod naslovom *Akustička analiza općeprihvaćenoga hrvatskog i srpskog govora – formantska analiza i mjere fundamentalne frekvencije*.

³¹ Osim govornika hrvatskog jezika za potrebe doktorske disertacije snimljeni su i analizirani izvorni govornici srpskog jezika (Bašić, 2018). U pilot-istraživanju Varošaneć-Skarić i Bašić (2015) provedena je formantska analiza prvih triju formanata na manjem broju govornika hrvatskog jezika i opsežnijemu govornom materijalu. Snimljeno je po pet ženskih i muških govornika koji su čitali listu od ukupno 125 riječi. Cilj je tog snimanja bilo provjeravanje uvjeta snimanja te kriterija za sastavljanje govornog materijala. Svaki je vokal bio predstavljen s 25 riječi kojima su obuhvaćeni utjecaji svih hrvatskih konsonanata na vrijednosti formanata medijalnog dijela naglašenog vokala.

[o]	M	566	1 040	2 468
	C	558	1 016	2 476
	Min.	446	868	2 201
	Maks.	680	1 257	2 685
	SD	12,8	20,1	25,2
[u]	M	395	844	2 575
	C	393	845	2 570
	Min.	319	639	2 308
	Maks.	486	962	2 766
	SD	10,4	18,7	27

Tablica 6. Referentne vrijednosti za hrvatski jezik na temelju 46 muških govornika
(preuzeto iz Bašić, 2018)

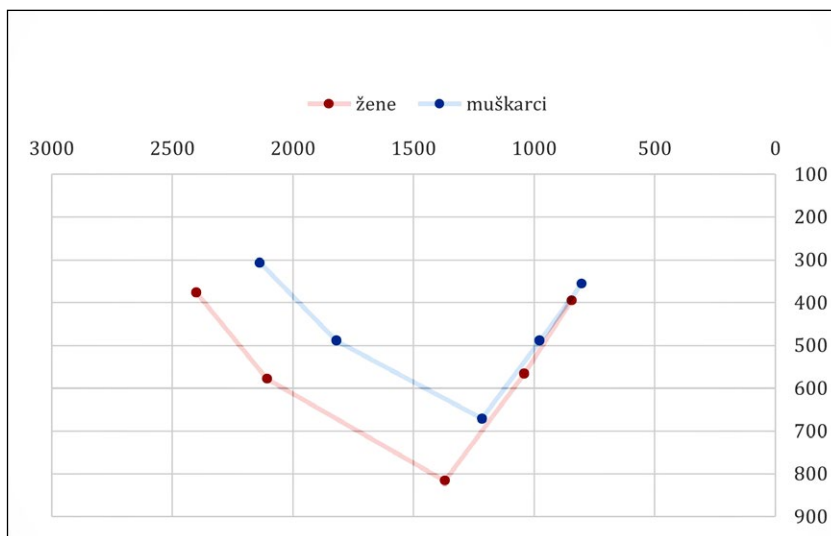
[a]	M	671	1 216	2 438
	C	673	1 213	2 429
	Min.	549	1 068	2 171
	Maks.	792	1 374	2 662
	SD	14,7	16,4	24,9
[e]	M	489	1 819	2 449
	C	493	1 820	2 433
	Min.	377	1 718	2 306
	Maks.	577	1 966	2 679
	SD	9,7	17,4	24,6
[i]	M	307	2 137	2 751
	C	307	2 125	2 757
	Min.	239	2 010	2 537
	Maks.	369	2 309	2 931
	SD	6,9	20,6	27,8
[o]	M	489	977	2 402
	C	498	669	2 427
	Min.	334	718	2 023
	Maks.	592	1 194	2 683
	SD	12,8	22,2	31,3

[u]	M	356	803	2 428
	C	353	792	2 420
	Min.	258	652	2 068
	Maks.	486	948	2 741
	SD	13	18,7	42,9

U prethodnim tablicama iznesene su vrijednosti deskriptivne statistike za formantske vrijednosti vokala u hrvatskom jeziku: medijan, prosječna, minimalna i maksimalna vrijednost te standardna devijacija. Sve su navedene mjere procijenjene za prva tri formanta (F1 – F3). Valja napomenuti da su prosječne vrijednosti i medijan (centralna vrijednost) uobičajene mjere u opisu formanta vokala, od kojih je potonja mjera statistički pouzdanija i stabilnija mjera na koju manje utječu ekstremne vrijednosti. Minimalne i maksimalne vrijednosti te standardna devijacija daju nam uvid u formantski raspon, tj. u veličinu distribucije rezultata, iako su snažno motivirane prethodno ručno određenim vrijednostima graničnih frekvencija formanta u postavkama programa za akustičku analizu zvuka Praat.

Referentni formantski okvir u hrvatskom jeziku s jedne je strane pridonio dosadašnjemu fonetskom opisu hrvatskog govora, dok je s druge strane omogućio usporedbu izgovora pojedinog govornika u odnosu na opći hrvatski izgovor vokala. Potonji se opis pokazao korisnim u opisu dijalektalnih, regionalnih govora (Biočina i Bašić, 2021), kontrastivnih analiza različitih jezika (Varošaneć-Škarić i Bašić, 2015), ali i u stvarnim forenzičnim fonetskim vještačenjima. Naime, u forenzici se formantska analiza smatra jednom od temeljnih fonetskih procedura koja je sastavni dio svakog izvještaja fonetskog vještaka. Govor osumnjičenika/okrivljenika opisuje se u odnosu na postojeći referentni formantski okvir hrvatskog jezika, a analize se provode na snimkama govora poznatog i nepoznatog govornika (engl. *known te questioned recordings*).

U već ranije spomenutim radovima (Varošaneć-Škarić i Bašić, 2015; Biočina i Bašić, 2021 itd.) korišteni su *plot* dijagrami koji omogućavaju jednostavniju interpretaciju rezultata. Na takvim je prikazima izražen odnos prvog i drugog formanta (F1 i F2). Na slici 20 prikazan je *plot* dijagram za muške i ženske govornike hrvatskog jezika na temelju vrijednosti medijana navedenih u doktorskoj disertaciji Bašić (2018: 85 i 92). Vrijednosti su navedene za sve vokale hrvatskog jezika ([i], [e], [a], [o] i [u]). Međuformantski i međuvokalski odnosi tumače se s obzirom na položaj u dijagramu te na izgovorna obilježja otvorenosti i zatvorenosti te prednosti i stražnjosti.



Slika 20. Plot dijagram prvih dvaju formanata (F1 i F2) za referentni formantski okvir hrvatskog jezika

5.5. Normalizacija formanata

Analiza vokala predmet je brojnih sociolingvističkih i sociofonetskih istraživanja. Fonetska, fonološka i druga obilježja vokala analiziraju se između govornika različitih jezika, dijalekata, različitog spola, dobi, podrijetla, stupnja obrazovanja itd. U radu Bašić i Varošaneć-Škarić³² (2023) predstavljena je kontrastivna analiza vokalskih prostora hrvatskog i srpskog jezika normalizacijskom formantskom metodom.

Normalizacija je vokala statistički postupak kojim možemo opravdano uspoređivati akustičke podatke poput formantskih vrijednosti između različitih skupina govornika; primjerice, između govornika različitog spola, jezika, dijalekta itd. Idiosinkratične fiziološke karakteristike govornika (veličina glave, duljina i oblik vokalnog trakta) uklanjaju se kao nepotrebna akustička varijabilnost iz podataka. Drugim riječima, ako u korpusu koji analiziramo imamo muške i ženske govornike, u interpretaciji rezultata nije opravdano uspoređivati vrijednosti formanata žena s

³² Za potrebe tog istraživanja korišten je korpus prikupljen u okviru projekta Ministarstva znanosti, obrazovanja i športa *Forenzička fonetika: slušno prepoznavanje i zvučna analiza govora* (br. 130-0000000-0786) te dviju kratkoročnih potpora Sveučilišta u Zagrebu: *Forenzička fonetika 2: standardiziranje zvučnih postupaka* (2014) te *Forenzička fonetika 3: Mjere fundamentalne frekvencije (f0) muških govornika* (2015), pod voditeljstvom prof. dr. sc. Gordane Varošaneć-Škarić.

vrijednostima formanata muškaraca, zbog toga što su formantske vrijednosti kod muškaraca niže zbog duljega i većega vokalnog trakta u odnosu na žene. Normalizacijom vokala uklanjaju se fiziološke razlike među govornicima, a pritom se ne utječe na dijalektalne, međujezične i sociolingvističke razlike među govornicima. Nenormalizirani rezultati formantskih frekvencija ne daju nam informaciju o tome koliko su razlike među analiziranim skupinama govornika uzrokovane jezičnim i sociolingvističkim faktorima, a koliko fiziološkim. Na službenoj stranici na kojoj su dostupni neki od poznatijih normalizacijskih paketa, NORM, ističe se da je „vokalska normalizacija presudna pri usporedbi vokalskih realizacija između različitih govornika na lingvistički i sociolingvistički smislen način“ (<http://lingtools.uoregon.edu/norm/index.php>, posjet 2. travnja 2019.).

5.5.1. *Vrste normalizacijskih metoda*

Temeljna je podjela normalizacijskih metoda na intrinzične i ekstrinzične metode. Kada su sve informacije za normalizaciju u jednom tokenu, riječ je o vokalski intrinzičnoj normalizacijskoj metodi, u kojoj se kombiniraju formantske vrijednosti (F1, F2, F3 i ponekad F4), vrijednosti fundamentalne frekvencije, a nerijetko i širine formantskih pojaseva. U vokalski ekstrinzičnoj normalizaciji uspoređuju se pak formantske vrijednosti različitih vokala između različitih govornika. Parametri govornika, vokala i formanta mogu se mijenjati u metodama, pa su sukladno tomu moguće višestruke kombinacije: intrinzično/ekstrinzično po pitanju govornika, intrinzično/ekstrinzično po pitanju vokala, intrinzično/ekstrinzično po pitanju formanata. Ukoliko koristimo informacije od samo jednog govornika, utoliko je prikladno korištenje govornički intrinzične metode. U slučaju da imamo više govornika predlaže se korištenje ekstrinzične normalizacije (primjerice normalizacija Labov, Ash i Boberg (2006)).

Adank (2003) je na velikom uzorku vokala standardnoga nizozemskog ispitala 12 normalizacijskih metoda. Rezultati su pokazali da se sociolingvističke informacije (o regionalnoj raznolikosti nizozemskog) najuspješnije mogu sačuvati metodama koje su formantski i govornički intrinzične, a vokalski ekstrinzične. Autorica posebno naglašava prednosti normalizacijskih metoda Lobanova (1971) i Neareya (1977). Isti su rezultati dobiveni u Flynnovoj (2011) studiji, u kojoj autor ispituje učinkovitost neutraliziranja varijabilnosti formantskih podataka s obzirom na fiziološke i anatomske razlike između govornika korištenjem 20 različitih metoda normalizacije vokalskih formanata.

Brojni autori naglašavaju da prije vokalske normalizacije valja preispitati svrhu normalizacijskog postupka te odlučiti koja je metoda prikladna za korpus kojim raspolazemo. Nužno je pritom odrediti je li nam uopće potrebna vokalska normalizacija, i ako jest, koji tip. Nasumičnim odabirom pojedinoga normalizacijskog pa-

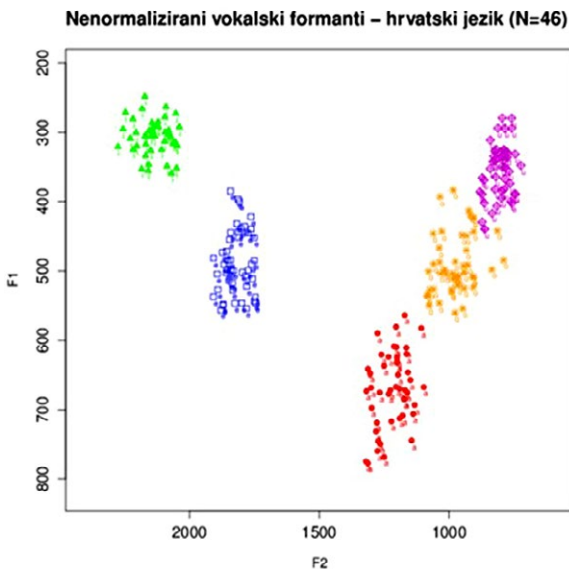
keta manipulira se podacima i u konačnici rezultatima, bez jasnih saznanja koje su se akustičke informacije eliminirale. Disner (1980) i Thomas (2002) stoga navode četiri opća cilja normalizacije:

- a) eliminirati varijabilnosti uzrokovane fiziološkim razlikama između različitih govornika
- b) sačuvati sociolingvističke, dijalektalne i međujezične razlike u vokalskoj kvaliteti
- c) sačuvati fonološke razlike među vokalima i
- d) oblikovati kognitivne procese koji nam omogućavaju da normaliziramo vokale koje izgovaraju različiti govornici.

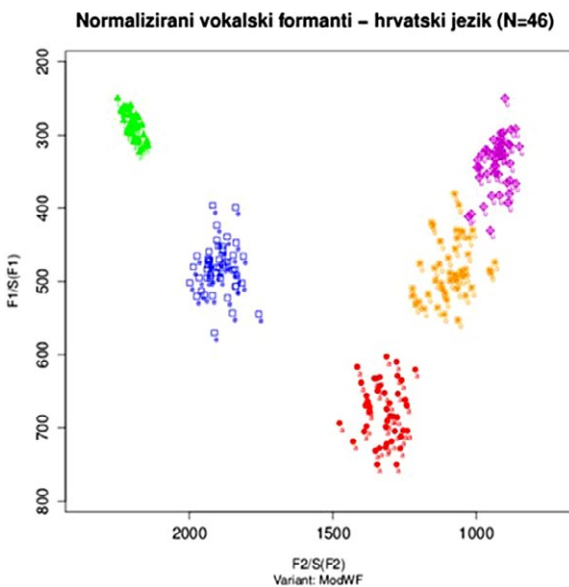
U fonetskim se radovima cilj normalizacije razlikuje u odnosu na sociolingvističke i dijalektološke radove i najčešće se odnosi na oblikovanje kognitivnih procesa uz pomoć kojih kao slušatelji normaliziramo vokale svojih sugovornika. Suprotno tomu, u sociolingvističkim i dijalektološkim radovima upravo je taj potonji cilj najmanje važan, a prva su dva cilja, (a) i (b), najvažnija.

Radi provođenja normalizacije formantskih frekvencija nakon akustičke obrade i analize govornog materijala dobivene je podatke nužno prilagoditi za obradu u normalizacijskim paketima koji su dostupni na mrežnoj poveznici *NORM: The Vowel Normalization and Plotting Suite* (<http://lingtools.uoregon.edu/norm/norm1.php>). Dostupno je nekoliko normalizacijskih metoda: određivanje razlika u barkovima, Labovljeva metoda, Lobanovljeva metoda, Neareyjeva metoda te metoda Watta i Fabricius. Neke od metoda ponuđene su u više varijanti (intrinzičnim i ekstrinzičnim), što istraživačima omogućava primjenu odgovarajućega normalizacijskog paketa na različite vrste korpusa.

Za analizu podataka (formantskih frekvencija) u radu Bašić i Varošaneć-Škarić (2023) korištene su tri normalizacijske metode: modificirana metoda Watta i Fabricius, Lobanovljeva i Neareyjeva metoda. Sve tri navedene metode pripadaju istoj klasifikacijskoj skupini (vokalski ekstrinzične te formantski i govornički intrinzične), koja, kako smatraju Adank (2003) i Flynn (2011), postiže najbolje rezultate u ujednačavanju i poravnavanju/centriranju vokalskih sustava različitih govornika. U radu su uspoređene nenormalizirane (slika 21) i normalizirane (slika 22) formantske vrijednosti za govornike obaju jezika. Za provedbu normalizacijskih metoda pojedini su tokeni kategorizirani u kategorije *vokala* ([i], [e], [a], [o] i [u]), *govornika*, *ciljane riječi*.



Slika 21. Nenormalizirane vrijednosti formanta u hrvatskom jeziku za muške govornike (Bašić i Varošaneć-Škarić, 2023)



Slika 22. Normalizirane vrijednosti formanta u hrvatskom jeziku za muške govornike (Bašić i Varošaneć-Škarić, 2023)

Vokalska normalizacija pridonosi snazi tumačenja jezičnih, sociolingvističkih i dijalektoloških razlika među govornicima, izostavljajući pritom individualne fiziološke razlike analiziranih govornika. Prilikom izbora vokalske normalizacijske metode valja razmotriti i uskladiti postavljena istraživačka pitanja s obilježjima korpusa i uvidjeti da normalizacija vokala nije jednodimenzionalna procedura, već složena mreža suodnosa između više čimbenika. Na primjeru hrvatskog i srpskog jezika u radu Bašić i Varošaneć-Škarić (2023) pokazane su prednosti normalizacijskog sustava u interpretaciji akustičkih rezultata, najčešće provedenih na vrijednostima vokalskih formanata. Rezultati su ukazali na očekivane razlike u varijabilnosti između govornika te (izgovorne) jezične razlike u obilježjima otvorenosti/zatvorenosti i prednjosti/stražnjosti (s obzirom na plan F1 i F2).

5.6. Formanti vokala u govoru različitog intenziteta

Pri uobičajenoj govornoj aktivnosti u ulozi govornika uglavnom smo nesvjesni svih varijacija u govoru: od promjena glasnoće, visine tona, stanki u govoru do boje glasa, fluentnosti govora itd. Kompleksnost govorne produkcije posebno je izražena u njezinoj varijabilnosti tijekom duljeg trajanja, što predstavlja posebne izazove u percepciji i analizi. Promjene u vokalnom naporu (engl. *vocal effort*, *vocal mode*) uglavnom su istraživane s obzirom na promjene osnovnog tona glasa, dok su malobrojna istraživanja ispitala promjene vrijednosti formanata pri govoru u šaptu i glasnom govoru u odnosu na govor prosječnog, neutralnog intenziteta.

Poznato je da se intenzitetske vrijednosti govora u šaptu kreću u rasponu od 20 do 30 dB, u prosječno glasnom govoru od 50 do 60 dB, dok u glasnom govoru one ostvaruju iznose za otprilike 20 dB više od intenziteta u normalnom, umjereno glasnom govoru (oko 80 dB). Zhang i Hansen (2007) govorni signal dijele na pet kategorija ovisno o vrsti govorne produkcije: šapat, meki govor, neutralni govor, glasni govor i vikanje. Šapat naravno podrazumijeva najnižu intenzitetsku vrijednost koja se postiže radom ograničenog dijela glasnica, iako je pritom vokalni napor u području glasnica veći nego pri neutralnom govoru. Upravo je iz tog razloga štetno govoriti u šaptu kada smo promukli ili imamo glasovnih poteškoća. Šapat možemo definirati i kao onu vrstu govorne aktivnosti koja je niskog intenziteta i koja ne sadrži fundamentalnu frekvenciju. Drugim riječima, pri produkciji šapta glasnice nisu aktivne.

Najveći intenzitet možemo očekivati pri vikanju, dok se meki govor intenzitet-ski nalazi između šapta i neutralnog govora, a intenzitet glasnog govora između neutralnog govora i vikanja. S obzirom na to da u podlozi različite govorne produkcije spomenutih pet kategorija leže i različiti mehanizmi proizvodnje glasa, njihova se varijabilnost ogleda i u različitim vrijednostima fundamentalne frekvencije i formantskih frekvencija. Pri analizi šapta učestalo možemo naići na poteškoću

utvrđivanja formantskih frekvencija zbog smanjenog intenziteta, dok se pri analizi govora u modalitetu vikanja ili glasnog govora može pojaviti tzv. fenomen *clippinga* (kada intenzitet ulaznog govora koji snimamo prelazi postavke maksimalnog raspona programa za snimanje glasa).

Hansen, Stauffer i Xia (2021) ističu da fenomen *clippinga* snižava kvalitetu govora na snimci te narušava uspješnost automatskog prepoznavanja govornika. Taj se fenomen može pronaći u iznimno kratkim dijelovima govorne snimke (rijetke govorne situacije na razini govorne riječi ili čak sloga koji je govornik preglasno izgovorio) ili pak u potpunosti može narušiti kvalitetu cijele snimke, ako npr. govornik tijekom cijelog snimanja zauzima poziciju koja je preblizu mikrofona ili ako postavke snimanja nisu prilagođene habitualnoj amplitudi govora pojedine osobe čiji se govor snima.

Autori objašnjavaju da gubitak govornih uzoraka u visokim amplitudama (uslijed *clippinga*) dovodi do nelinearnih distorzija koje se očituju kroz izobličene harmonike u visokim frekvencijama, što rezultira pojavljivanjem zvučnih artefakata. *Clipping* je unutar govorne snimke razmjerno lako uočiti jer se u valnom obliku ne pojavljuje šiljati oblik, već svojevrsan oblik platoa, odnosno „odrezanih” spektralnih vrhova govornog signala, koji su u biti izgubljeni jer im je amplituda bila prevelika u odnosu na postavke snimanja. S obzirom na akustičke značajke vokala, diftonga, likvida i poluvokala, koji imaju značajnu energiju, očekuje se da će se fenomen *clippinga* češće pojaviti u tim skupinama zvučnih glasnika nego kod drugih, primjerice bezzvučnih. Isti je fenomen zamjetan i pri izgovoru okluziva (u fazi otpuštanja okluzije, kada zračna struja naglo izlazi kroz suženi otvor govornog prolaza), posebice ako su govornici preblizu mikrofona (Hansen, Stauffer i Xia, 2021). S obzirom na spomenuto, autori su ispitali i utvrdili da je *clipping* u većoj mjeri utjecao na više formante, odnosno na formantske vrijednosti koje se nalaze u višim područjima spektra, dok su niži formanti ostali nepromijenjeni.

O vrijednostima formanata u izolirano foniranim vokalima u šaptu i neutralnom intenzitetu pisao je Jović (1998) te je utvrdio da su vrijednosti prvih dvaju formanata statistički značajno više kod vokala [i], [e] i [a] u usporedbi s neutralno glasno izgovorenim vokalima, dok kod trećega i četvrtoga formanta nisu zamijećene slične tendencije porasta frekvencija formanata. Zanimljivo, Jović (1998) također utvrđuje da je širina formanata veća kod vokala u šaptu te da je gotovo konstantna za vrijeme cijelog trajanja vokala. U istraživanju formanata vokala u šaptu na kineskom jeziku utvrđeno je da se, osim veće širine formanata te viših vrijednosti formanata pri šaptu, vrijednosti prvih i drugih formanata približavaju, odnosno da je frekvencijska udaljenost prvih i drugih formanata u šaptu manja nego u neutralno glasnom govoru (Ding, Li i Xu, 2004; Li, Ding i Xu, 2005).

Wilson (1985) je ispitao kretanja formantskih frekvencija u vezanom govoru pri promjeni iz neutralno glasnog govora u šapat. Rezultati tadašnje studije provedene

u sustavu LPC-10 vocoder pokazali su znatno povišenje frekvencija formanata i širine formanata kod govornika obaju spolova.

Novija istraživanja pokazala su da pri usporedbi govora u šaptu i u neutralno glasnom govoru možemo primijetiti razlike u govornom tempu te u vrijednostima formanata (Houle i Levi, 2020), a neka su istraživanja pokazala da su slušatelji manje precizni u određivanju jezičnih informacija (npr. određivanju sadržaja izgovorenog materijala) i informacija o govorniku (npr. određivanju spola govornika) u šaptu nego u neutralno glasnom govoru. To ni ne čudi jer smo u šaptu lišeni fundamentalne frekvencije, koja nosi brojne informacije o govorniku (spol, dob itd.), a zbog niskog intenziteta razabirljivost govora značajno je teža.

U studiju za akustička snimanja pri Odsjeku za fonetiku snimljen je govorni materijal za svaki vokal hrvatskog jezika kod jedne ženske govornice hrvatskog jezika. Formantske frekvencije (F1 – F3) analizirane su u naglašenim vokalima jednosložnih riječi: za vokal [a]: /pas/, /rat/, /mast/, /šah/, /rast/, za vokal [e]: /peh/, /šef/, /red/, /med/, /led/, za vokal [i]: /mir/, /pir/, /vir/, /sir/, /žir/, za vokal [o]: /dom/, /hod/, /pod/, /rob/, /nož/, za vokal [u]: /sud/, /mulj/, /puh/, /duh/, /žulj/, zasebno za govorni modalitet šapta, neutralno glasnog govora te glasnog govora. Intenzitetske vrijednosti govora u šaptu kretale su se u rasponu od 30 do 35 dB, u neutralno glasnom govoru oko 60 dB, a u glasnom govoru u rasponu od 75 do 80 dB. Sve su riječi izgovorene intonacijski neutralno, što je verificirao jedan fonetičar. Nakon montaže i anotacije govornog signala na svakom je govornom modalitetu primijenjena skripta za formantsku analizu u Praatu (Boersma i Weenink, 2023) te su utvrđene formantske vrijednosti prvih triju formanata. Na temelju uprosječenih vrijednosti prvih dvaju formanata za svaki vokal hrvatskog jezika razlike u šaptu, neutralnom i glasnom govoru prikazane su u dijagramu vokalskih trokuta.

U tablici 7 prikazane su vrijednosti prvih triju formanata vokala hrvatskog jezika u šaptu te u glasnom i neutralno glasnom govoru. Vrijednosti pokazuju povišenje prvih dvaju formanata svih vokala u usporedbi šapta i neutralno glasnog govora (osim F2 kod vokala [i]), od kojih su statistički značajne razlike u vrijednostima F1 kod svih vokala te F2 kod vokala [a], [o] i [u], dok su razlike kod prednjih vokala statistički beznačajne. Vrijednosti trećeg formanta ne pokazuju iste tendencije, što je potvrđeno i u ranijim studijama. Naime, kod šapta su vrijednosti F3 kod gotovo svih vokala povišene u usporedbi s neutralno glasnim govorom (osim kod stražnjeg vokala [o]), dok pri usporedbi vrijednosti F3 kod glasnog govora i neutralno glasnog govora uopće ne možemo govoriti o tendencijama: pri izgovoru vokala [i] i [a] F3 je nižih vrijednosti u glasnom govoru, pri izgovoru vokala [e] približnih vrijednosti, dok je kod stražnjih vokala F3 statistički značajno viših vrijednosti u usporedbi s neutralno glasnim govorom.

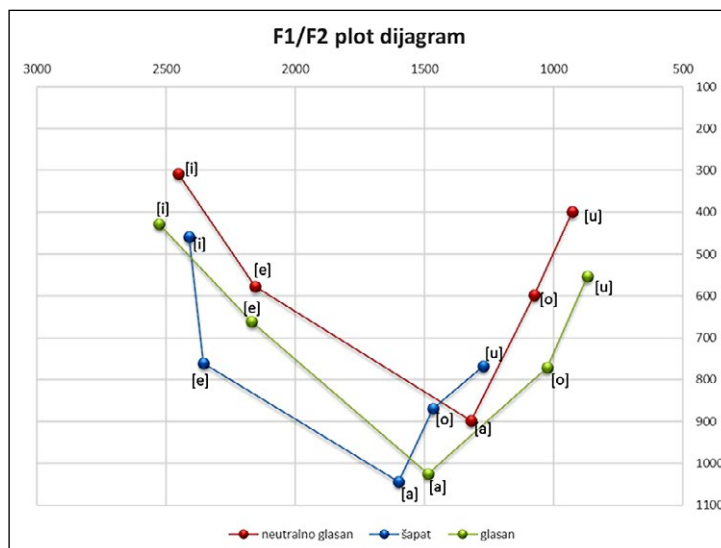
Tablica 7. Vrijednosti prvog i drugog formanta (Hz) vokala hrvatskog jezika u neutralno glasnom govoru, u šaptu te u glasnom govoru

neutralno glasni govor				šapat				glasni govor			
vokal	F1	F2	F3	vokal	F1	F2	F3	vokal	F1	F2	F3
i	307	2452	3005	I	458	2411	3038	i	428	2527	2875
e	578	2156	2785	E	761	2355	2926	e	662	2169	2782
a	898	1319	2970	A	1044	1601	3295	a	1026	1486	2876
o	598	1076	2795	O	869	1466	2789	o	772	1026	3066
u	399	929	2766	U	768	1273	2858	u	553	870	3262

Tablica 8. Statistička značajnost razlike frekvencija prvih triju formanta u usporedbi šapta, neutralno glasnog i glasnog govora kod svih vokala hrvatskog jezika (izražena kroz vrijednost p t -testa)

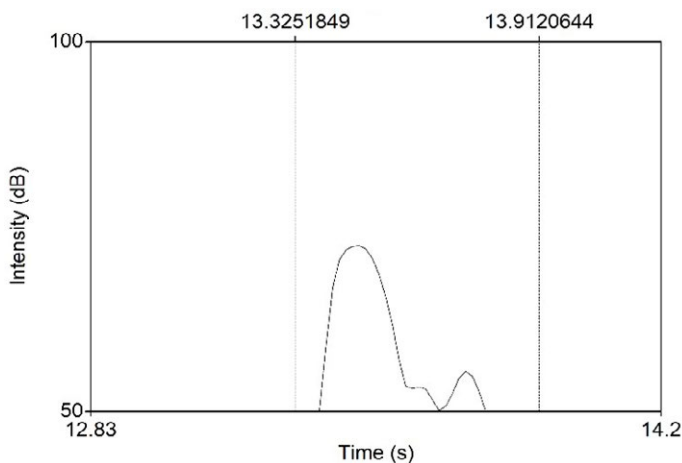
neutralno glasni govor : šapat				neutralno glasni govor : glasni govor			
vokal	F1	F2	F3	vokal	F1	F2	F3
i	0,009	0,002	0,26	I	0,00830	0,06939	0,25568
e	0,00000003	0,063	0,12	E	0,00062	0,45388	0,49127
a	0,000054	0,065	0,0264	A	0,00355	0,27016	0,11079
o	0,00085	0,002717	0,4141	O	0,00523	0,28048	0,03843
u	0,00000047	0,000437	0,3757	U	0,00187	0,22967	0,01879

Na dijagramskom prikazu vokalskih trokuta na temelju vrijednosti prvih dvaju formanta (F1 – F2) vokala [i], [e], [a], [o] i [u] pri šaptu, glasnom govoru i neutralno glasnom govoru možemo zamijetiti značajna odstupanja. Razlike u izgovoru vokala među govornicima, ali i unutar istog govornika učestalo se prikazuju dijagramskim prikazima jer oni mogu zorno prikazati odstupanja u izgovoru s obzirom na vrijednosti F1 i F2, odnosno s obzirom na os otvorenosti i zatvorenosti te os prednosti i stražnosti. Slika 23 prikazuje nam da su odstupanja od vrijednosti formanta veća pri šaptu nego pri glasnom govoru u usporedbi s neutralno glasnim govorom.

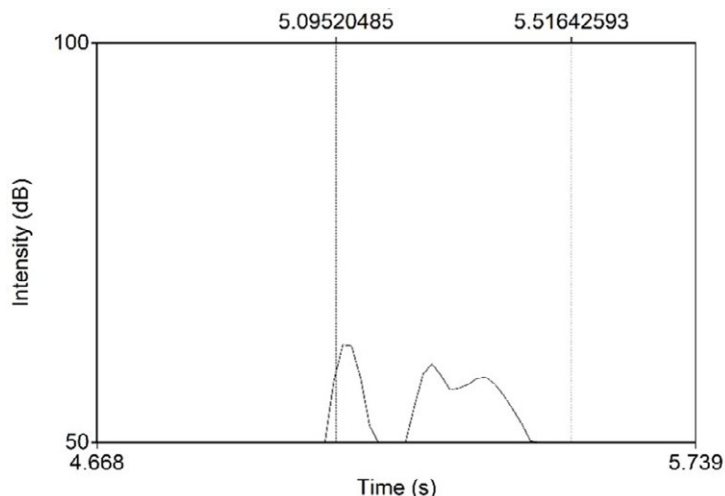


Slika 23. Dijagramski prikaz vokalskog trokuta na temelju vrijednosti prvih dvaju formanta (F1 – F2) vokala [i], [e], [a], [o] i [u] u modalnom i šaptavom govoru

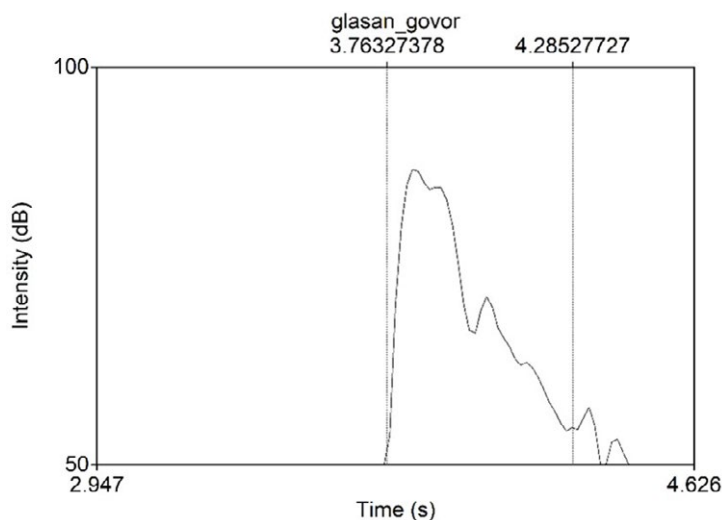
Intenzitetske krivulje šapta, neutralno glasnog govora i glasnog govora prikazane su na slikama 24, 25 i 26 (Praat: Boersma i Weenink, 2023).



Slika 24. Intenzitetska krivulja na primjeru neutralno glasnog govora



Slika 25. Intenzitetska krivulja na primjeru šapta



Slika 26. Intenzitetska krivulja na primjeru glasnog govora

5.7. Formantska analiza atipičnoga govornog materijala

Kao što je u prethodnim poglavljima opisano, formantska se analiza provodi u različitim kontekstima govorno-jezične produkcije te u najvećoj mjeri ovisi o tome što, zašto i s kojim ciljem analiziramo. Formanti se osim u vokalima kao zvučnim realizacijama fonema pojedinog jezika mogu analizirati i u atipičnome govornom materijalu, primjerice u govornim zvučnim stankama. Naime, u forenzičnim je fonetskim istraživanjima upravo govorna disfluentnost ponekad idiosinkratičan znak

pojednog govornika te ga može istaknuti u odnosu na druge govornike istog jezika. Tako primjerice govornik može biti govorno disfluentan upotrebom čestih poštapalica, tzv. rafalnog ponavljanja određenog dijela riječi i slično.

Tipologija govornih disfluentnosti uvelike se razlikuje kod različitih autora te je stoga u njihovoj analizi nužno istaknuti koje su sve vrste disfluentnosti uvrštene u analizu te navesti primjere iz korpusa. One se mogu analizirati u kontekstu forenzične fonetike, govorno-jezične proizvodnje i produkcije, atipičnog i tipičnog govora i slično. Jedno od ranijih istraživanja proveli su Maclay i Osgood (1959: 19) donoseći kvantitativnu studiju ispunjenih i praznih stanki, ponavljanja i pogrešnih započinjanja. Shriberg (1994) se bavila analizom disfluentnosti u spontanom govoru odraslih govornika američkoga engleskog te je na temelju 5.000 zabilježenih disfluentnosti u trima različitim govornim stilovima iznijela svoju klasifikaciju koja uključuje ispunjene stanke, ponavljanja, umetanja, zamjene, izostavljanja i artikulacijske pogreške.

Rodríguez, Torres i Varona (2001), Clark i Fox Tree (2002), Fox Tree (1995: 709) te Lickley (2017) koriste sličnu tipologiju disfluentnosti analiziranu na korpusima spontanog govora. U području forenzične fonetike disfluentnosti su se u izvještajima forenzičnih fonetskih vještaka uglavnom opisivale impresionistički, primjerice rečenicom: „U govoru osumnjičenika i počinitelja kaznenoga djela prisutna su vrlo slična obilježja disfluentnosti govora. U snimkama su utvrđena zamuckivanja koja se očituju u čestim ponavljanjima slogova i riječi.“ Uvođenjem profila govornika TOFFA (engl. *Taxonomy of Fluency Features for Forensic Analysis*, u slobodnom prijevodu *Taksonomija obilježja fluentnosti u forenzičnim analizama*) disfluentnostima se pristupa na objektivniji način s jasnom metodologijom. Njegova je svrha praktična primjena u stvarnim forenzičnim fonetskim slučajevima, najčešće pri prepoznavanju govornika. TOFFA je verificirani profil koji razlikuje pet kategorija disfluentnosti: prazne stanke, pune stanke, ponavljanja, duljenja te samoprekidanja (McDougall i Duckworth, 2017: 18).

Primjeni profila TOFFA prethodi ortografska prilagodba i anotiranje govorenog materijala (*text grid* u Praatu), što uključuje popisivanje svih disfluentnosti iz tipologije profila TOFFA te raspodjelu govora na slogove (riječ „mito“ anotirat će se kao „CV CV“). Na taj način utvrdit će se ukupan broj slogova i vrsta disfluentnosti u anotaciji govornih snimki, a udio disfluentnosti uobičajeno se iskazuje mjerom broja disfluentnosti u jedinici vremena, odnosno na 100 slogova. U radu McDougall i sur. (2019) pojednostavljaju se dvije kategorije unutar profila TOFFA. Naime, u kontekstu forenzične fonetike u fokusu su interesa samo one prazne stanke koje se ne pojavljuju na gramatičkom mjestu (na granicama govornih riječi, sintagmi, na kraju rečenice i slično), već one prazne stanke koje ne bismo očekivali i koje su stoga specifične za samog govornika i njegov iskaz (u profilu TOFFA označavaju se kao [po], engl. skraćeno za *pause other* u usporedbi s [pg] za *pause grammatical*).

Drugo se pojednostavljenje sastoji u sjedinjenju kategorija duljenja frikativa ([prof] za engl. *prolongation of fricative*) te kategorije produljenja zatvorenoga dijela pri izgovoru okluziva i afrikate ili trajanja otpuštanja ([prop] za engl. *prolongation of plosive*) u jedinstvenu kategoriju duljenja [proc] koja predstavlja sva konsonantska produljenja. Na taj način duljenja glasnika dijelimo na vokalska [prov] (primjenjiva na vokale, nazale i laterale) i konsonantska [proc] (primjenjiva na preostale glasnike). O disfluentnostima u forenzično-fonetskom smislu pisala je Varošaneć-Škarić (2019) navodeći pritom tipologiju blisku onoj koju iznosi Shriberg (1994) te primjere iz stvarnih forenzično-fonetskih slučajeva.

Hughes i sur. (2016) istraživali su dokaznu vrijednost punih stanki (ponajprije onih koje su u anglosaksonskom govornom području vrlo učestale: /um/ ili /uh/) kao varijabli u forenzičnoj usporedbi glasova. Disfluentnosti se u kontekstu forenzične fonetike istražuju imajući na umu govornika i njegove specifičnosti, odnosno govorne posebnosti. S tim na umu, u govoru osumnjičenika zanima nas sve što ga razlikuje od uobičajenog i prosječnog, univerzalnog i očekivanog kod prosječnoga izvornog govornika istog jezika. U većini drugih istraživačkih područja disfluentnosti interes je istraživača usmjeren na ono opće i univerzalno, a ne nužno specifično. Pritom valja napomenuti da uredno fluentni govornici (koji imaju uredan govorno-jezični status bez poremećaja u fluentnosti govora) nisu savršeno fluentni, već savršeno fluentnom govoru mogu samo težiti. S obzirom na to da je govorne disfluentnosti teško kontrolirati (maskirati) te da odražavaju način na koji jezično planiramo svoj iskaz, interes za njih u forenzici ne jenjava. Pridamo li tomu i podatak da je varijabilnost među govornicima velika te da ovisi o brojnim čimbenicima (spol, dob, regionalnost, stupanj obrazovanja, poznavanje stranih jezika, govorni stil i zadatak, emocionalna angažiranost govornika, vremensko ograničenje, poznavanje teme o kojoj se govori itd.), postaje jasnim i da je metodologija proučavanja disfluentnosti složena.

Među disfluentnostima u forenzičnoj fonetici posebno su važne prazne zvučne stanke (tzv. punjači). Naime, o njihovoj koristi možemo saznati iz prirode njihove upotrebe:

- one su zvučne, zbog čega su podatne za akustičku analizu
- ne nose jezično značenje i govornici ih vrlo često nisu svjesni
- dugo traju, zbog čega su na njih primjenjive razne akustičke mjere koje mogu biti analizirane u više vremenskih uzoraka
- nose biološke anatomske informacije jer su mjerljive različitim pouzdanim akustičkim parametrima (fundamentalna frekvencija, formanti i slično)
- javljaju se neovisno o mjestu u rečenici i česta su pojava u diskursu
- uglavnom su omeđene stankama, zbog čega nema koartikulacijskih utjecaja okolnih glasnika.

Osim spomenutih karakteristika Hughes i suradnici (2016) potvrdili su provedenim istraživanjem da su upravo zvučne stanke vrlo malo varijabilne unutar govornika (mala intravarijabilnost) te da smo kao govornici vrlo dosljedni u njihovoj upotrebi. Istraživanje iste vrste stanki u kontekstu tempa govora proveo je Künzel (1997), dok su učinak intoksikacije alkoholom na pojavu disfluentnosti u govornika istražili Schiel i Heinrich (2015: 19).

Profil TOFFA vrlo je nalik podjeli koju ističu Clark i Wasow (1998: 201), prema kojoj je spontani govor ispunjen neželjenim stankama, produljenim segmentima, punjačima (kao što su *uh* ili *um*), uređivačkim izrazima (kao što su *I mean* i *you know*, koje češće nazivamo poštapalice, npr. *znači* u hrvatskom jeziku), zatim fragmentima riječi, samoispravljanjima i ponavljanjima. Rodríguez, Torres i Varona (2001: 2) pak, uza sve što razlikuju Clark i Wasow (1998), disfluentnostima smatraju još i umetanja, zamjene, govorne pogreške te, zanimljivo, šumove proizvedene tijekom snimanja, kojima pripadaju negovorni elementi poput udaha, izdaha, kašlja ili pozadinskih zvukova. Iako su svjesni da to nisu disfluentnosti u strogom smislu, autori ih ubrajaju u klasifikaciju jer su vrlo rijetki u čitanom, a svojstveni su spontanom govoru (Rodríguez, Torres i Varona, 2001: 2).

Neki hrvatski autori također smatraju da su udasi i izdasi disfluentnosti (Horga i Požgaj Hadži, 2012; Horga i Vidović Zorić, 2022) jer prekidaju govor i narušavaju njegovu tečnost. Stanke potrebne za udah i izdah povezane su sa sintaktičko-semantičkim granicama, zbog čega je očekivano da će se pojaviti na granicama rečenica i fraza (Lickley, 2017: 373). Tu valja napomenuti da je stoga u analizi fluentnosti govora nužno odrediti jesu li u fokusu istraživanja disfluentnosti u širem smislu (svi prekidi govora koji negativno utječu na njegovu fluentnost) ili pak disfluentnosti u užem smislu (svi prekidi govora koji su ujedno govorni i nisu izvangovorne naravi).

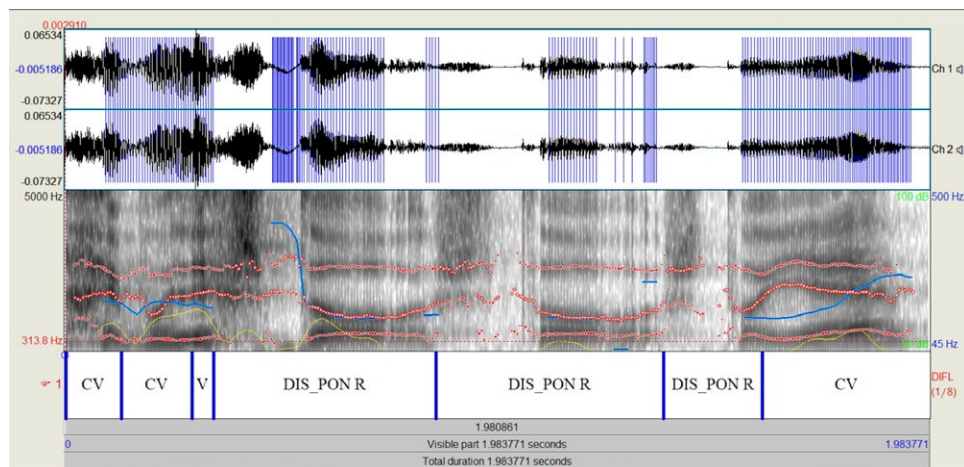
U nekim su istraživanjima ispitane dobne razlike u disfluentnosti među govornicima (Caruso i sur., 1997; Horga i Vidović Zorić, 2022), dok su drugima u fokusu bile razlike u distribuciji disfluentnosti ovisno o govornom stilu, govornim zadacima i različitim televizijskim žanrovima (Horga i Požgaj Hadži, 2012; Golub i Vidović Zorić, 2022, Žagmešter i Bašić, 2023; Varošaneć-Škarić, Bašić i Šegvić, 2023).

U nastavku će biti navedeni neki primjeri iz spontanog govora u formi intervjua govornika preuzetih iz baze govora hrvatskog i srodnih jezika Bazvuka dostupnoj na poveznici <https://bazvuka.ffzg.unizg.hr/> (Varošaneć-Škarić, Bašić i Biočina, 2022).

S obzirom na to da brojna istraživanja u hrvatskom, ali i u drugim jezicima ukazuju na najveću čestotnost zvučnih praznih stanki (tzv. punjača) i njihovih različitih varijanti, valja započeti s njihovim ovjerenim primjerima u korpusu hrvatskog govora (Bazvuka: Varošaneć-Škarić, Bašić i Biočina, 2022).

Među brojnim govornicima hrvatskog jezika uočene su različite vrste punih stanki: izolirana duljenja glasnika šva [ə] u različitim trajanjima te u različitim varijantama s fonemskim i nefonemskim sastavnicama hrvatskog jezika.

Na slici 27 prikazano je tzv. rafalno ponavljanje riječi /što/ (DIS_PON R), koje se inače odnosi na uzastopno trostruko ili višestruko ponavljanje slogova, dijelova riječi, riječi ili većih govornih jedinica. Ženska govornica mlađe dobi učestalo u spontanom govoru koristi spomenutu vrstu govorne disfluentnosti, koja je na slici 25 prikazana u govornom isječku [najbo^{le} fto je] (/najbolje što što što je/).

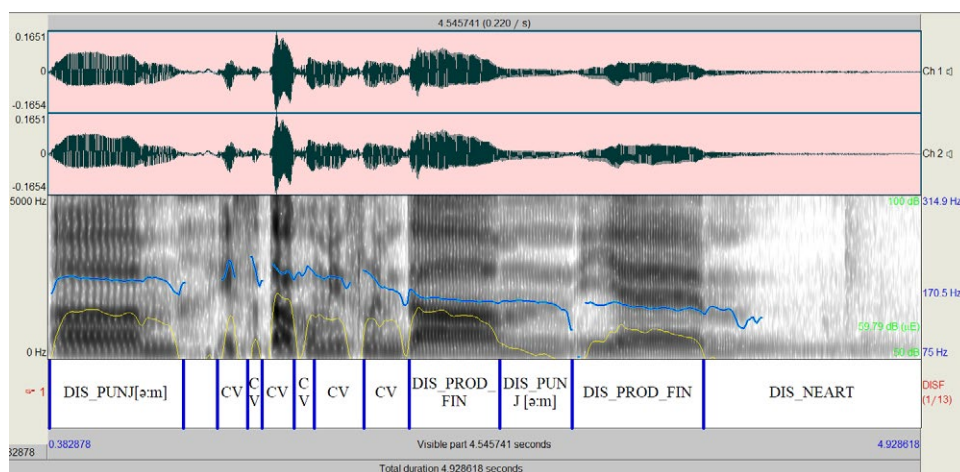


Slika 27. Rafalno ponavljanje riječi /što/ u dijelu iskaza spontanog govora mlađe govornice hrvatskog jezika (spektrogramski prikaz uz anotaciju)

Spektrogramski prikaz s iscrtanim vrijednostima tona (fundamentalne frekvencije plavom linijom) i intenziteta (žutom linijom) ukazuje na nekoliko posebnosti promatrane vrste disfluentnosti. Naime, pri prvom izgovoru riječi /što/ intenzitet je bio najveći, a potom je pri preostalim ponavljanjima slabio, kao što je i fundamentalna frekvencija bila niža u nepočetnim izgovorima (F_0 [što]₁: 208 Hz; F_0 [što]₂: 154 Hz; F_0 [što]₃: 141 Hz). Također, slušnom je analizom utvrđen i slabiji, tj. manje dostatan izgovor u naknadnim ponavljanjima, a zanimljivo je sagledati i vremensku odrednicu uzastopnih ponavljanja. Naime, prvi i drugi izgovor riječi /što/ trajali su gotovo podjednako (0,51 i 0,52 s), dok je posljednji, treći izgovor riječi /što/ bio iznimno kratak i trajao je svega 0,2 s. Može se pretpostaviti da je upravo ta vrsta disfluentnosti bila svojevrсна stanka oklijevanja u kojoj je govornica procesirala kako će nastaviti rečenicu te je u trenutku završetka planiranja ostatka govornog iskaza okončala oklijevanje i ubrzano nastavila kreirati svoj daljnji govorno-jezični iskaz.

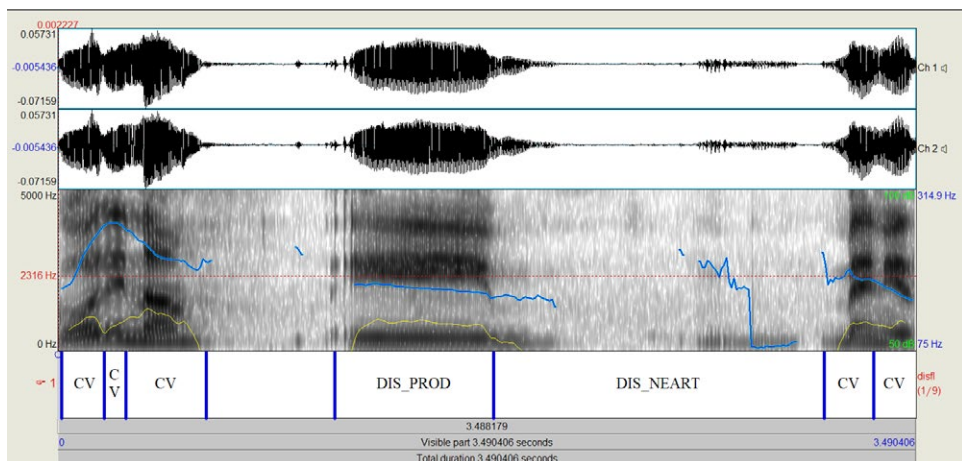
Na slici 28 prikazane su tri vrste disfluentnosti (ukupno pet disfluentnosti) izgovorene u vrlo kratkome govornom isječku (4,5 s) mlađe ženske osobe ([ə:m isto tako mislim da: ə:m je:]). Anotacijom su označene dvije disfluentnosti u obliku pune stanke (DIS_PUNJ [ə:m]) u slijedu produljenog vokala šva i bilabijalnog nazala, dvije disfluentnosti finalnog duljenja vokala (DIF_PROD_FIN) te jed-

na neartikulirana disfluentnost koja slušno podsjeća na duljenje okluzivnog dijela okluziva ili afrikate. S obzirom na to da iza nje nije uslijedio izgovor jednog od spomenutih glasnika, ona je svrstana u neartikulirane stanke, a ne u duljenja konsonanata (kao primjerice u tipologiji profila TOFFA). Na slici je također jasno da su disfluentnosti znatno duljeg trajanja od govornih segmenata (CV struktura) te da primjerice finalna produljenja poput [da:] i [je:] znatno premašuju prosječno trajanje vokala čak i pod dugim naglaskom ili pak zanaglasnom duljinom (što i nije očekivano paradigmatско obilježje u navedenim primjerima). Od disfluentnosti prikazanih kod ove mlađe ženske osobe najdulje je trajanje neartikulirane stanke (DIS_NEART), koje iznosi 1,13 s, zatim trajanje punjača (0,69 s) i finalnih duljenja (0,68 i 0,47 s).



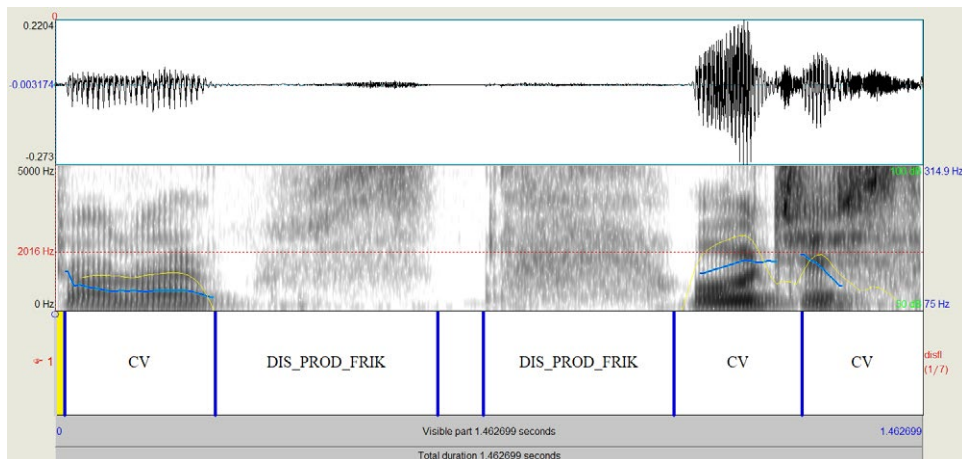
Slika 28. Primjer razmjerno nefluentnog govora mlađe ženske osobe (spektrogramski prikaz uz anotaciju)

Na slici 29 u nastavku teksta prikazan je drugi primjer disfluentnosti kod iste govornice, duljenje vokala (DIS_PROD) u vezniku /i/ te još jedan primjer značajno duge neartikulirane stanke, od kojih potonja nije vrlo česta vrsta disfluentnosti. U spomenutom primjeru govornog isječka u kojem je izgovoreno [godina i: meni], veznik je [i] trajao ukupno 0,64 s, dok je neartikulirana stanka koja je nakon njega uslijedila iznosila 1,34 s.



Slika 29. Primjer duljenja vokala i neartikulirane stanke (spektrogramski prikaz uz anotaciju)

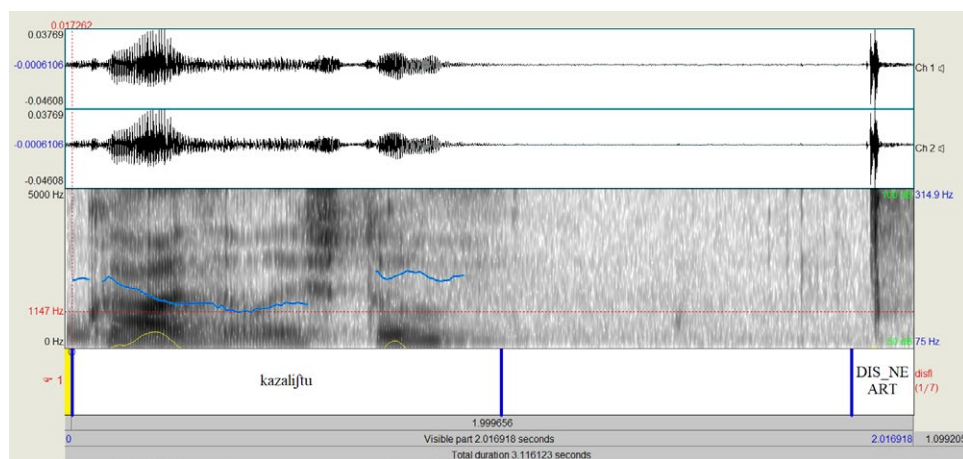
Kao što je već ranije spomenuto, među disfluentnosti ubrajamo i različite vrste produljenog izgovora. Osim vokala, mogu se duljiti nazali i laterali (u profilu TOFFA u kategoriji [prov]), ali i drugi glasnici, npr. frikativi. Na slici 30 u nastavku teksta vidljiv je spektrogramski prikaz uz anotaciju mlađega muškog glasa koji je u svojem govoru koristio produljen izgovor frikativa ([DIS_PROD_FRIK]). U kratkome govornom isječku u trajanju od svega 1,46 s utvrđena su dva primjera spomenutih disfluentnosti u trajanju od 0,37 s i 0,32 s.



Slika 30. Primjer produljenog izgovora frikativa (spektrogramski prikaz uz anotaciju)

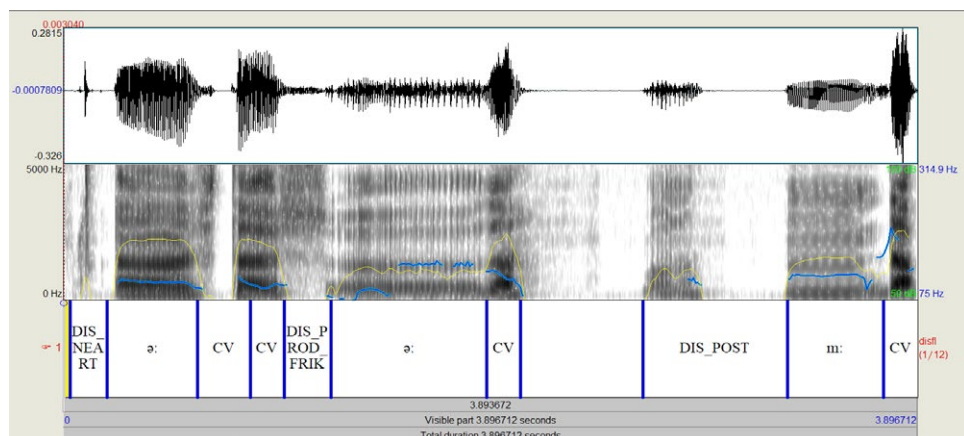
Od neartikuliranih je zvučnih stanki kod nekih govornika prisutan i tzv. alveolarni cektaj ([nts]), koji može biti vrlo učestao, posebice na granicama govornih riječi ili u inicijalnoj rečeničnoj poziciji. Na slici 31 prikazan je spektrogramski prikaz

uz anotaciju izgovorenog isječka [kazaliftu] iza kojeg je uslijedila spomenuta neartikulirana zvučna stanka (alveolarni coktaj [nts]). Spomenuta se atipična govorna realizacija može pokazati vrlo korisnim idiosinkratičnim govornim markerom pri prepoznavanju govornika u forenzičnoj fonetici, s jedne strane zbog svoje posebnosti i relativne rijetkosti među općom populacijom, a s druge strane zbog toga što je govornici uglavnom nisu svjesni. Također, na primjeru ove govornice utvrđena je vrlo visoka razina čestotnosti svih vrsta disfluentnosti (23 disfluentnosti u 100 riječi; 22,4/min), od kojih se alveolarni coktaj pojavio 12 puta u 3,2 min (196 s), što bi značilo u prosjeku svake 16,3 s.



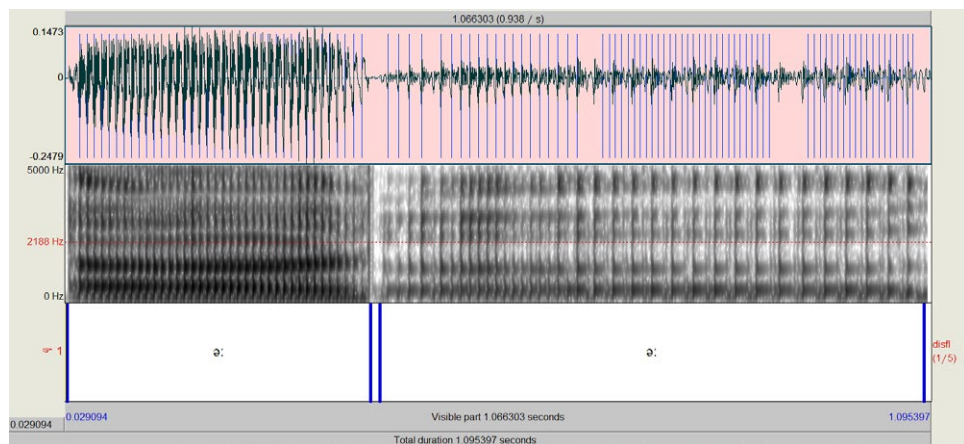
Slika 31. Spektrogramski prikaz dentalnog coktaja kao posebne vrste neartikulirane disfluentnosti uz anotaciju

Usporedbom kvalitativnih (vrsta disfluentnosti) i kvantitativnih (broj disfluentnosti) razlika među govornicima primijećeno je da će određeni govornici koristiti vrlo malo vrsta disfluentnosti u duljem govornom isječku, dok će drugi pak u vrlo kratkom isječku upotrijebiti više vrsta disfluentnosti, koje osim toga mogu biti i učestale (slika 32). Na kraćem govornom isječku od 3,89 s kod jednoga je mlađeg muškog govornika utvrđeno šest disfluentnosti: alveolarni coktaj (DIS_NEART), puna stanka ([ə:]), stanka duljenja frikativa u finalnoj poziciji u riječi /još/ (DIS_PROD_FRIK), druga puna stanka ([ə:]), poštapalica (DIS_POST) te treća puna stanka ([m:]).



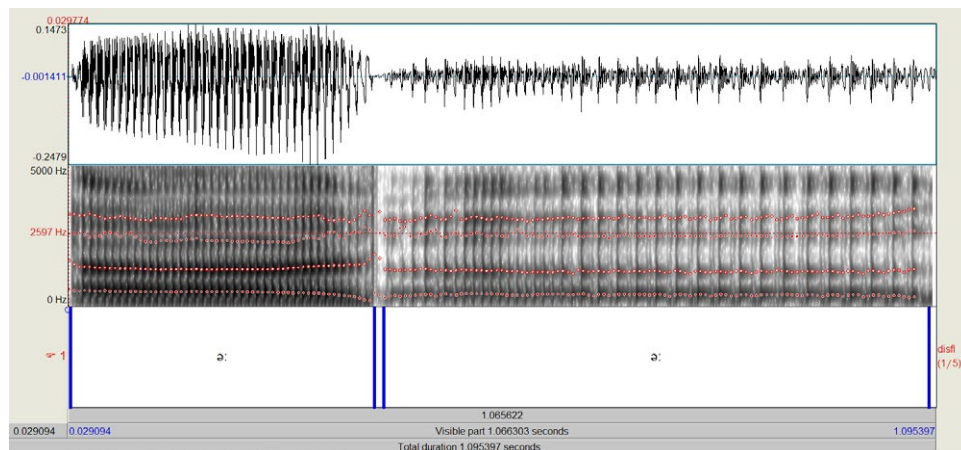
Slika 32. Primjer učestalih disfluentnosti u kratkome govornom isječku kod mlađeg govornika hrvatskog jezika

Osim velike čestotnosti govornih disfluentnosti, kod ovog je govornika zanimljiva i razlika u timbru punih stanke. Usporedimo li primjerice prve dvije pune stanke u obliku duljenja glasnika šva, možemo zamijetiti da je druga puna stanka izgovorena pod utjecajem glotalizacije (slika 33). Prosječna vrijednost fundamentalne frekvencije prve pune stanke u trajanju od 0,37 s iznosi 106,5 Hz, dok prosječna vrijednost F_0 druge pune stanke u nešto duljem trajanju od 0,67 s iznosi 94,36 Hz. Akustička je analiza glasa pokazala da je broj pulsacija glasiljki veći u prvoj stanci (38 pulsova i 37 perioda u 0,36 s prve stanke te 31 puls u 28 perioda u izdvojenih 0,36 s druge stanke), na što upućuje i vizualni prikaz valnog oblika s gušće raspoređenim popratnim plavim okomitim linijama (pulsacije glasiljki). Spektrogramski prikaz također ukazuje na glotalizaciju, koju ponekad susrećemo na kraju govornih iskaza.



Slika 33. Spektralni prikaz uz anotaciju dviju punih stanke izgovorene različitom kvalitetom glasa (muški mlađi glas)

Kao što je ranije spomenuto, u forenzičnoj se fonetici uvelike ističe vrijednost punih stanki pri prepoznavanju govornika. Osim što su vrlo često omeđene stanka- ma te nisu kao ostali glasnici pod koartikulacijskim utjecajem okolnih glasnika, tra- ju dulje od govornih segmenata (glasnika) te odražavaju anatomske karakteristike govornika (mjerljive fundamentalnom frekvencijom i formantima). Formantska je analiza (slika 34 i tablica 9) tih dvaju punjača pokazala vrlo slične vrijednosti prvih četiriju formanata (F1 – F4) kao odraza rezonantnih šupljina vokalnog trakta, iako je fundamentalna frekvencija u drugoj stanci bila niža.



Slika 34. Spektrogramski prikaz s formantskim konturama (F1 – F4) uz anotaciju dviju punih stanki kod mlađega muškog govornika

Tablica 9. Vrijednosti trajanja, fundamentalne frekvencije te formantskih frekvencija (F1 – F4) dviju punih stanki ([ə:]₁) kod mlađega muškog govornika

	trajanje	F ₀	F1	F2	F3	F4
[ə:] ₁	0,36 s	106,5 Hz	483 Hz	1337 Hz	2392 Hz	3132 Hz
[ə:] ₂	0,64 s	94,36 Hz	414 Hz	1204 Hz	2478 Hz	3089 Hz

5.8. Dugotrajni prosječni spektar govora govornika (LTASS)

Osim spektrogramskih analiza i prikaza u programu Praat, u fonetskim se istra- živanjima učestalo koriste i analize i prikazi LTASS. LTASS se odnosi na dugo- trajni prosječni spektar govora (prema engl. *long-term average speech spectrum*), koji nam daje informaciju o uprosječnoj zvučnoj energiji u govornom spektru koja je akustički korelat boje glasa. Osim dugotrajnoga prosječnog spektra, za procjenu i analizu kvalitete glasa koristi se i kepsralni vokalni vrhunac (CPP – engl. *Cepstral Peak Prominence*), mjera uobičajeno korištena za procjenu atipičnih glasova.

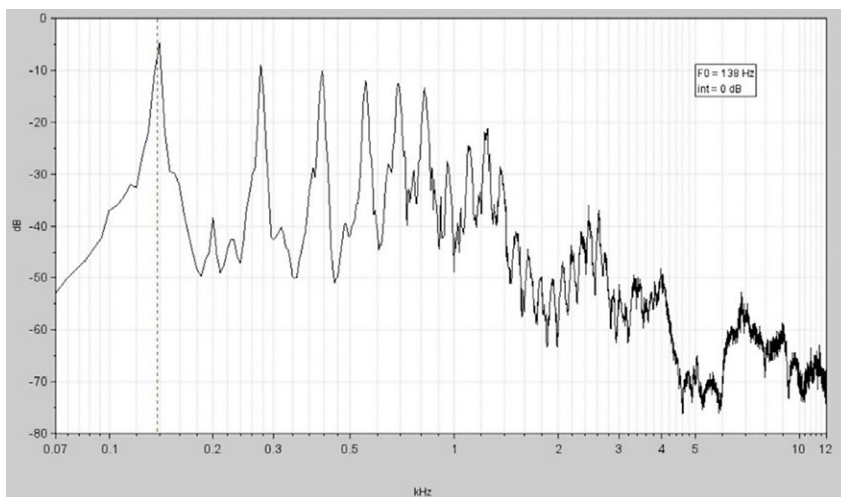
Za valjanu procjenu dugotrajnoga prosječnog spektra govora, tj. LTASS-a, nužna je snimka vezanoga, kontinuiranoga spontanog govora u trajanju od 70-ak sekundi ili pak govora pri čitanju kraćeg teksta. Pritom je važno napomenuti da je pri odabiru govornog materijala za čitanje optimalno izabrati nefrikativni tekst (Varošaneć-Škarić, Stevanović i Bašić, 2021). Osim procjene razine zvučne energije u govornom spektru, pri analizi je LTASS-a nužno odrediti i u kakvu su međusobnom odnosu energije pojedinih frekvencijskih područja. Tako primjerice možemo odrediti da su glasovi čija spektralna energija naglo pada i niskog je intenziteta u području voluminoznosti slabi glasovi, dok je primjerice u LTASS-ima disfoničnih glasova prisutna velika zvučna energija u višim dijelovima spektra (Tanner, Roy, Ash, Buder, 2005). U tumačenju LTASS-a često se analizira spektralni nagib/pad, prosječna razina zvučne energije, standardna devijacija energije te spektralni vrhunac.

Osim spomenutoga, pri analizi LTASS-a možemo se koristiti podjelom zvučnoga govornog spektra na sljedeća područja: voluminoznost (do 400 Hz), punoća (400 – 800 Hz), zvonkost (800 – 2.000 Hz), okruglost (2.000 – 2.500 Hz), blistavost³³ (koja se dijeli na nižu, svojstvenu za muškarce: 2.500 – 3.500 Hz, te višu, svojstvenu za žene i djecu: 3.500 – 5.000 Hz) te pucketavost (iznad 5.000 Hz) (Škarić, 1991).

Valja napomenuti da se češće primjenjuje Hussonova podjela na sljedeća glavna područja spektra: područje A (od F0 do prvoga formanta, F1), područje koje zahvaća prvi formant (F1), područje B (od F1 do F2; tzv. međufornantsko područje), područje koje zahvaća drugi formant (F2), područje C (od drugog formanta do frekvencije prekida (2.300 – 2.500 Hz)), područje D (od frekvencije prekida do 3.500 Hz) te područje E (od 3.500 Hz do kraja spektra) (1962; Varošaneć-Škarić, 2005).

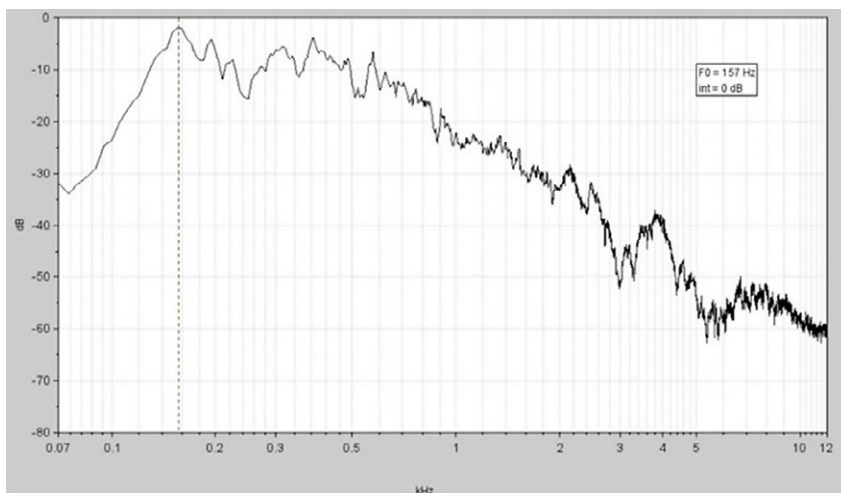
Na slici 35 vidljiv je prikaz LTASS pri fonaciji vokala [a] kod mlađega muškog izvornog govornika hrvatskog jezika. Prvi vrh na krivulji uprosječenog govora predstavlja fundamentalnu frekvenciju (138 Hz), dok preostali istaknuti vrhovi predstavljaju formante.

³³ Blistavost glasa ili blještavost (Škarić, 1991: 292) određena je snagom spektra u području od 2,5 kHz do 5 kHz te se dijeli na nižu (2,5 – 3,5 kHz), svojstvenu muškim glasovima, i višu (3,5 – 5 kHz), svojstvenu dječjim i ženskim glasovima.



Slika 35. Prikaz LTASS fonacije vokala [a] kod mlađega muškog govornika

U vezanom je govoru (vidi sliku 36), očekivano, uprosječni spektar govora znatno drugačiji zbog jezičnog materijala. Naime, pri fonaciji vokala jezični je sadržaj vrlo jednostavan, dok je pri npr. čitanju nefrikativnog (NFT) i frikativnog teksta (FT)³⁴ (Škarić, 1993) on znatno složeniji. Nefrikativni tekst sadrži sve glasnike osim frikativa, dok frikativni sadrži sve glasnike hrvatskog jezika (Varošaneć-Škarić, Stevanović, Bašić, 2021).



Slika 36. Prikaz LTASS nefrikativnog govora kod mlađega muškog govornika

³⁴ Prilagođeni su tekstovi objavljeni na hrvatskom i engleskom jeziku u Varošaneć-Škarić, Stevanović i Bašić (2021: 54), a uobičajeno se koriste kao govorni materijal za procjenu kvalitete glasa.